ESSAYS ON ECONOMIC BEHAVIOR UNDER UNCERTAINTY

Edited by

M. BALCH

University of Iowa

D. McFADDEN

University of California at Berkeley

and

S. WU

University of Iowa

1974

NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM · OXFORD

AMERICAN ELSEVIER PUBLISHING COMPANY, INC. – NEW YORK

Library of Congress Catalog Card Number: 73-92557

ISBN North-Holland for this series 0 7204 3100 X
ISBN North-Holland for this volume 0 7204 3191 3
ISBN American Elsevier for this volume 0 444 10618 9

Publishers:

NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM
NORTH-HOLLAND PUBLISHING COMPANY, LTD – OXFORD

Sole distributors for the USA and Canada:

AMERICAN ELSEVIER PUBLISHING COMPANY, INC.
52 VANDERBILT AVENUE, NEW YORK, NY 10017

Printed in the Netherlands

INTRODUCTION TO THE SERIES

This series consists of a number of hitherto unpublished studies, which are introduced by the editors in the belief that they represent fresh contributions to economic science.

The term economic analysis as used in the title of the series has been adopted because it covers both the activities of the theoretical economist and the research worker.

Although the analytical methods used by the various contributors are not the same, they are nevertheless conditioned by the common origin of their studies, namely theoretical problems encountered in practical research. Since for this reason, business cycle research and national accounting, research work on behalf of economic policy, and problems of planning are the main sources of the subjects dealt with, they necessarily determine the manner of approach adopted by the authors. Their methods tend to be 'practical' in the sense of not being too far remote from application to actual economic conditions. In addition they are quantitative rather than qualitative.

It is the hope of the editors that the publication of these studies will help to stimulate the exchange of scientific information and to reinforce international cooperation in the field of economics.

THE EDITORS

PREFACE

In recent years there has been a steady increase in the investigation of both theoretical and applied problems of decision making under uncertainty. To facilitate communication and the exchange of ideas as they are developing, the National Science Foundation has sponsored the NSF–NBER Conference on Decision Rules and Uncertainty. Four conferences have been held to date: the first two at the Massachusetts Institute of Technology in May 1971 and January 1972, the third at the University of Iowa in May 1972, and the last at Princeton University in March 1973. There were a number of papers presented at each of these conferences and their titles are given in an appendix following the bibliography. After the Iowa conference it was decided to publish a volume. Because of prior commitments many of the papers presented at these conferences do not appear in this volume while, on the other hand, several additional papers have resulted from post-conference interaction and are here included. Although this volume is not strictly a proceedings, we felt that it would be instructive to include not only papers but comments as well.

In addition to an introductory essay, this volume is roughly divided into three parts. Part 1 comprises two papers that deal with the conceptual development of the conditional expected utility framework. Part 2 includes five papers on various micro-aspects of behavior under uncertainty. The five papers in part 3 are concerned with welfare economics and general equilibrium. The last paper, as indicated by its title, was originally a comment on the Kesten–Stigum paper. Since it also provides an excellent discussion on uncertainty and on the problem of modeling an appropriate equilibrium concept, we felt that by making it the last paper, it would also serve as a concluding remark for the volume.

We wish to thank the National Science Foundation, the National Bureau of Economic Research and the Murray Fund at the University of Iowa for their financial support of the conferences; Miss Diane Grottola, Mr. Kent Currie and Mr. An-Sik Min for research assistance; and Mrs. Marguerite Knoedel for typing the manuscript.

MICHAEL BALCH
DANIEL MCFADDEN
S. WU

TABLE OF CONTENTS

Part 3

LIST OF CONTRIBUTORS

M. Balch
University of Iowa

W. Bœ
University of Iowa

E. Bolker
University of Massachusetts at Boston

P. Diamond
Massachusetts Institute of Technology

P. Fishburn
Pennsylvania State University

J. Green
Harvard University

J. Hadar
Southern Methodist University

K. Hamada
University of Tokyo

J. Hickman
University of Wisconsin

C. Hildreth
University of Minnesota

R. Jeffery
University of Pennsylvania

H. Kesten
Cornell University

D. Krantz
University of Michigan

M. Kurz
Stanford University

J. Ledyard
Northwestern University

H. Leland
Stanford University

D. Levhari
Hebrew University of Jerusalem

R. Luce
University of California at Irvine

D. McFadden
University of California at Berkeley

L. Mirman
University of Massachusetts

J. Mirrlees
Oxford University

R. Mukherjee
University of Rochester

G. Nordquist
University of Iowa

J. Pratt
Harvard University

R. Radner
University of California at Berkeley

S. Ross
University of Pennsylvania

W. Russell
Southern Methodist University

E. Sheshinski
Hebrew University of Jerusalem

B. Stigum
Northwestern University

S. Wu
University of Iowa

E. Zabel
University of Rochester

CHAPTER 1

SOME INTRODUCTORY REMARKS ON BEHAVIOR UNDER UNCERTAINTY

M. Balch and S. Wu

1.1. Introduction

In the past few decades, economic theorists have become progressively engaged in developing models to account for the presence and impact of real world uncertainties. *A priori* reason for this interest is obvious enough: the assumptions of 'deterministic' theory sometimes appear to be rather far removed from reality[1]. Indeed, the 'usual' microcontext is in many cases more faithfully described by recognizing uncertainties that are real world typical for that context, and which contribute to its operational definition in a significant way. An entrepreneur, for example, may have to decide *ex ante* whether to commit resources that could turn out to be inappropriate *ex post*; thus, as events would have it, an unproductive and unmarketable sunk cost. On the other hand, the decision not to commit could result in an opportunity forgone, since the ability to seize such opportunities may depend upon having some sort of resource structure already in place[2].

Our purpose in this essay is to offer an informal discussion of how uncertainties affect economic behavior, and how behavior in turn influences the nature and incidence of uncertainty. These microfoundational questions have attracted attention not only because they are of interest in their own right, but, more broadly, for their explanatory power concerning the very structure of economic systems and the evolution of economic processes.

---

[1] Of course models compete with one another on the basis of their relative power to explain and predict; this relative power, in turn, depends upon the primitive assumptions on which the competing models rest.

[2] In a competitive world in which adjustment costs increase sharply with rate of adjustment.

An economic actor is concerned with the course of his future, which course is shaped jointly by his own actions and by circumstances beyond his knowledge and control – and so in this way from one moment to the next. Of course some moments are more significant than others since, by and large, decision makers seem to respond to a *status quo* (in the broad experiential sense) by maintaining it. Time is irreversible in the true economic sense: opportunities may sometimes arrive but then vanish soon after, and disasters, once visited, leave their effect for some time to come. What makes this interesting is that a decision maker can often contribute essential 'karma' to his environment by putting himself in the way of such contingencies, or by avoiding their incidence. His actual choice in any context depends upon how he believes action may resolve and, when there is uncertainty about this for at least some options, upon his attitudes with respect to the bearing of risk. Thus one option may appear 'quite "interesting", but "risky"' while, by comparison, another appears '"less exciting", but "reasonable" and "secure"'.

In section 1.2 we discuss these perceptive and attitudinal determinants as they bear in general upon economic behavior, and as they derive from environmental context. In particular, we note the intrinsic presence of endogenously created uncertainties, and suggest a *contextual* explicant for the hypothesis of risk avoiding behavior[3]. In section 1.3 we look at the mechanisms by which actors in a decentralized economy reduce, generate and redistribute uncertainties. These not only encompass the markets, of course, but a variety of topics in industrial and social organization as well.

1.2. Behavioral Determinants

The characteristic feature of a decision problem under uncertainty is that the generic option *must* be implemented before its precise outcome can be known (though some options may appear less vague in this respect than others). A rival's reaction to an oligopolist's price reduction

[3] The theory of risk-avoiding behavior has emerged as perhaps the most significant analytic contribution of the uncertainty view, and its application to a considerable spectrum of economic settings will be apparent throughout. This hypothesis is 'testable' when consequences can be represented in a linear space; otherwise we shall be using the term 'risk-aversion' in a heuristic sense.

cannot be known until price has been reduced; the resolution of a coin toss cannot be known until the coin is tossed; and similarly, the profitability of a capital expansion cannot be known before the decision to expand has been executed (in so far as its resolution depends on future market conditions).

The determinants of individual behavior that operate in a world of uncertainty subsume (in the 'proper' sense) those that operate when action stands in a one-to-one causal relationship with its outcome[4]. In such latter circumstances a decision maker's perception of causality is 'perfect', and his attitudes in respect to the bearing of uncertainty simply do not come into play. The micro-models of conventional theory are by and large characterized on this 'perfect everything' view: they are conceived in a way that collapses future time to the present, and in such fashion that current environment is perfectly and commonly[5] perceived by all. Thus it is usually assumed that all decision-relevant 'truth' may be perfectly acquired in principle, and that this is costless to effect in practice. When this necessarily involves the world as it will be 'tomorrow' – as, for example, when production is recognized to be a process that takes place over time and requires commitments that must be made *ex ante* – then an appropriate deterministic variant is called for: *viz.* that all actors enjoy perfect (and perfectly validated) foresight. Whether in this way or by simply assuming that production is instantaneous and that every production period is (somehow) perfectly isolated from all others, it follows that firms produce so as to maximize profit:

---

[4] There is a complementary sense in which the behavioral determinants that operate under uncertainty may be understood to 'subsume' those which operate under conditions of certainty. For example, consider a von Neumann–Morgenstern expected utility maximizer whose preference order $\succsim$ over a lottery set $\mathscr{L}(X)$ (where X is some convex subset of an underlying space of commodity bundles in R^n, say) is represented by a cardinal utility indicator $u: X \to Re$. The assumption of risk aversion for $\succsim$ on $\mathscr{L}(X)$ is represented by concavity of u. Convexity of the 'certainty world' preference sets $P_y = \{x \in X | x \succsim y\} = \{x \in X | u(x) \geq u(y)\}$ for every $y \in X$ (where $\succsim$ on X is just the restriction of $\succsim$ on $\mathscr{L}(X)$) is an *automatic* consequence of the assumption of risk-aversion for $\succsim$ on $\mathscr{L}(X)$. This serves to indicate the power of model-theoretic robustification: one is hard-put to rationalize behavioral assumption α for certainty model A (which assumption is usually made for purposes of analytic convenience), but this appears as an automatic consequence of a 'more palatable' behavioral assumption β for uncertainty model B, where B subsumes A.

[5] This is germane to the logical construction of the conventional view; thus, while data set D may not be relevant to the choice problem that faces actor α, he *could* know this information if indeed it were relevant.

a well defined univalent function of known parameters. In logical complement to this view of production, individuals are assumed to choose market baskets and make investment decisions under conditions that leave no room for doubt, anxiety, hope or regret. Traders meet, not bilaterally, but with impersonal commodity and factor markets that are perfectly perceived by all who may be concerned. These markets convene costlessly and clear on the instant, by virtue of a *tatonnement* process that informs economic actors as surely as if they were Newtonian mass particles in a gravitational field[6].

In the real world, however, the actual consequence of an action may depend upon uncertain circumstances that lie beyond the decision maker's control[7]. That is, notwithstanding the decision maker's ability to exercise some conditionalizing influence over his environment, preferred consequences can no longer be guaranteed by action alone. The choice of an option, then, must depend upon the decision maker's relative evaluation of 'pure' consequences as well as upon his judgments concerning their relative likelihoods (given implementation of the alternatives for which they are conditionally relevant).

When relative likelihoods can be described according to known statistical distributions, then the decision context has 'full information' spirit: actors are universally privy to the same *characterizing* data set of probability laws. In such cases choice is founded on risk attitudes alone[8]. When relative likelihoods cannot *all* be so described – and this is quite the common real world situation – then the basis for 'universal perceptive agreement' erodes, and an individual's personal appreciation of context becomes relevant to choice. Indeed, two actors may look at

[6] This is not to say that economists have not for some time been quite aware of the presence and impact of market imperfections, but rather that the formal models that attempt to deal with these imperfections still rest, by and large, upon an essentially static *tatonnement* framework, and thus necessarily bypass those questions that are associated with the bilateral (multilateral) phenomenon. Cournot equilibrium among competing oligopolists, for example, conceives what is in fact an 'extensive' process (one that evolves over time in stepwise fashion) as one that may be characterized in 'normal' form (in which strategic options are simultaneously played, once and for all). While this sort of time-collapse is perhaps a not unreasonable place to begin analysis, it does bypass questions of judgment, learning, and reaction to unexpected changes in circumstance. Economic intercourse to come is surely no less vague than the evolution of a chess game between two human beings. Cf. also footnote 11.

[7] Beyond the strategic (or contingency) possibilities for that action.

[8] These subsume, of course, those behavioral determinants that would operate under conditions of certainty.

what ostensibly appears to be the same context but still have different judgments as to how uncertainty is likely to resolve.

We may consider two codeterminants by which context-perception can differ among actors.

1.2.1. Information

Ignorance about one's (non-statistically characterizable) environment and/or its inherently unclear future plays a fundamental perception-shaping role. It is sometimes possible to reduce one's ignorance by acquiring 'meaningful' information, although this is not always the case[9]. A decision maker may find it most difficult, for example, to collect and assess information regarding uncertainties that are generated by the interrelated nature of human behavior[10]. This is perhaps most evident in any bilateral context (or more generally, in a multilateral context with a small number of participants) where the action of one has a direct bearing on the welfare of the other and where both have a (non-singleton) set of such options from which to choose. Since the typical situation is one of conflicting self-interest, each may attempt *ex ante* to bluff or otherwise hide useful information from the other, or to rest upon what he considers to be better staying power; *ex post*, there may be possibilities for reneging or for other externality-producing forms of morally hazardous behavior. The point is that uncertainty generates from the fact that neither player can completely know the preferences, judgments and options of the other, nor how these will change with a change in circumstances, nor therefore the precise influence that his action will have upon the behavior of the other[11].

---

[9] Even when this may be feasible – as when an unalterable state of nature obtains in truth but has yet to be discovered – the (uncertain) benefits of additional information may be prejudged not worth the costs of acquisition.

[10] We concur with Professor Kurz (chapter 13) that uncertainties of endogenously created origin (which of course include all future market prices) appear to be more significant for economic theory than those of the 'natural disaster' variety.

[11] This subject area is formally addressed by the models and solution concepts of game theory, of which three have achieved some preeminence.

The notion of the core focuses on the question of where economic blocking power resides within an actor-set in which coalition formation is permissible. However, this model employs the notion of a 'value' for every coalition that does not depend upon the behavior of the complementary coalition.

The notion of a Nash equilibrium suffers from this shortfall in another way. It deals with a decentralized actor-set whose participants have agreed to act simultaneously, but with no other form of contracting allowed. Moreover, the model is 'full information'

1.2.2. The interpretive filter

An actor's perception of current truth and future likelihoods is colored not only by his (imperfect) information about the present, but also by his experiences and understandings of the past. Heuristically speaking, these constitute the interpretive filter through which 'raw' data pass on their way to becoming subjective judgment. Thus, two actors may look at the same 'horse race' and prejudge its outcome quite differently, each according to what he 'knows' at this moment in time. In particular, a professional speculator will specialize in obtaining relevant information as a primary input to his decision process. But he acts finally because he believes that his judgments are more accurate than others that may currently prevail, and this wisdom is based upon his cumulative past experiences with, and 'savvy' of, the information-gathering activity. Average long-run return to 'betting' or speculative activities is not so much a matter of luck as it is one of better perception, and thus depends upon the accuracy of both data collection and its interpretation.

1.2.3. The bankruptcy endpoint and risk-avoiding behavior

With respect to attitudes on the bearing of uncertainty, it would appear from the existence of a wide variety of insurance markets and from a host of nonmarket 'insurance-surrogate' activities that we shall explore below that choices are *biased*, by and large, toward a preference for 'security', whenever the (opportunity) cost of achieving this is not too high. This tendency toward risk-avoiding behavior (as characterized, in the 'simplest' case, by a concave utility-of-wealth indicator) may be understood in the following way. Suppose an act may result either in a favorable outcome or, as events might have it, in an unfavorable one. This unfavorable contingency, if realized, would bring the decision

---

in spirit since the option set for each is known to all. The solution concept is then developed in terms of (statistically) mixed strategies which, because of the full information and simultaneity assumptions, do not depend upon *ex ante* anticipations of *ex post* reactions. This is the game-theoretic analogue of the instantaneous *tatonnement* concept of a market.

The notion of a Nash bargaining game does attempt to account for the presence of threat and counterthreat possibilities in the sense that this is tacit backdrop for the formal description of the model. Because of such mutually sub-optimal possibilities the solution concept revolves on finding a 'fairly' bargained division with respect to a distinguished starting point (the status quo). Again, however, the model is full information in spirit.

maker closer to the 'bankruptcy endpoint' of the 'endowment (or wealth) half-line' on which all economic actors necessarily live. This endpoint is 'fuzzy', of course, because the meaning of bankruptcy is itself a matter of context, but its presence and import for the decision maker is clear enough in a world that does not support the debt of (demonstrable) paupers without limit[12]. In such a world, bankruptcy is (and is seen by all as) an absorbing barrier: the closer one comes to it, the more difficult (in a stochastic sense) it is to escape; once there, the supply of credit to the actor in question, and therefore his *effective* opportunity for escape, vanishes altogether. Given this sort of environment, we might expect an individual to be more protective of his current wealth (as measured, say, by how much he will pay for any 'test' gamble) the closer is this endowment to his bankruptcy endpoint[13].

To illustrate this net of ideas, it may be useful here to have an impressionistic look at how imperfect perception and risk-averse attitudes affect some aspects of economic life. The general theme is that economic actors are often perceptively bound to the local circumstances in which they find themselves and that aversion to unknown circumstances has a tendency to promote 'middle-of-the-road' policies that remain stable over time. We shall have more illustrations in section 1.3.

1.2.4. Entry

The nature of the capital decision is that liquid resources must be embodied in those specific non-liquid forms that are called for by the process and, once this precommitment takes place, the return to real productive capital is thereafter inextricably bound to the market fate of the process at hand. In neoclassical language, we may say that uncertainty associates to long-run equilibrium price. (For simplicity; more precisely, uncertainty associates to the profit stream.) If this price should realize substantially smaller than would be necessary to justify the capital decision, then the firm may be forced into bankruptcy. Of course a decision to expand is taken in view of this possibility and rests upon entrepreneurial judgment concerning prospects for success.

The entry decision is something more than just a garden variety limiting case of the generic expansion question because of characteristic

[12] This would be seen as an inferior risk by the market that faces our Principal Actor.

[13] Cf. the remarks of Professor Ross (chapter 6) for a complementary view of this same question.

differences in the nature and magnitude of associated uncertainties. To begin with there is the quantum character of the real-productive capital requirement itself, which is typically bounded from below because of set-up indivisibilities or for other reasons having to do with economies of scale. This induces an uncertainty atom of corresponding magnitude that contributes to the fundamental entry barrier; i.e. an entrepreneur must be prepared to accept this atom of uncertainty or find others to share its incidence. In addition, however, a new entrant must survive the rigors of the birth process itself, and thus faces a spectrum of problematic questions that established firms have already weathered. These have to do with initial bugs in the production/marketing process, and with establishing market position in the company of less vulnerable rivals who *may* act to frustrate this purpose. To meet these and other such contingencies an entering firm must maintain sufficient internal flexibility in the form of liquid and semi-liquid reserves: external credit is typically limited for a firm that has not yet demonstrated its earning power (we shall return to this credit aspect just below). Thus, while prospects for a successful passage through the birth canal enhance with scale of flexible reserves, the barrier to entry increases as well.

The entry decision in respect to human capital has similar features, of course. In this connection we may emphasize the role of risk-aversion, which has obvious significance for the question of occupational choice and, thus, interesting implications for the theory of profit and distribution.

1.2.5. The credit constraint

Bankruptcy is a natural Darwinian feature of the economic process; firms do fail for their inability to survive random shocks in market environment. Thus, the random profit flow of any given firm may sometimes be negative. This could happen whenever production is interrupted[14], for example, or because of random shifts in demand[15]. If such conditions should continue of sufficient magnitude over a sufficiently long period of time, and if survival prospects should appear to be thus

[14] Say due to the nonavailability of essential inputs, or to a sudden prohibitive rise in their costs.

[15] Perhaps because of a shift in tastes, or aggregate income (when income elasticity is greater than 1), or due to the advent of new substitutes that render the firm's current operations uncompetitive.

dimmed, then the firm may be forced into bankruptcy. Indeed, as we have remarked more generally above, this possibility is self-aggravating: risk-averse investors and creditors will not support a firm that appears to be headed for economic ruin. Thus the firm faces an externally imposed credit constraint which assumes binding force at precisely the worst moment so far as the question of the firm's survival is concerned[16].

Of course the result of bankruptcy is that owners would lose their equity and managers would lose their employment (as well as jeopardize their possibilities for similar future employment). Given this, and in view of the credit dynamic just noted, firm management is moved to take internal arbitrage measures in precaution of debilitative market shocks. In respect to the profit *stream*, for example, firm management has operational control over its intertemporal distribution through choice of dividend/investment policy. Realized profits are divided between dividends (discretionary current period return to owners) and investment, where investment may be undertaken so as to promote the vitality of the future profit stream[17]. The point is that the uncertainty characteristics of this stream are not independent of investment path, and the tradeoff between current dividend *versus* (uncertain) potential for continued dividend-generating vitality is a reflection of managerial risk preferences on this matter[18].

---

[16] Of course this contrasts with the 'perfect everything' view which holds that firms do not face a budget constraint. Rather, inputs may always be purchased – through borrowing, if necessary – so long as present discounted value of the profit stream is positive.

In the real world, however, the presence of a potential externally imposed credit constraint leads the firm to self-impose a limitation on both the amount and the rate of borrowing so as not to jeopardize its borrowing power at times when credit is critical for its survival. This policy decision – which depends upon characteristics of the debt-equity market – sets the firm's short-run budget for operating capital purposes.

[17] We may have demand uncertainties in mind, for example, where advertising flow (as viewed in this investment role) has some positive effect on (random) sales flow, and where the ability of the firm to withstand random market shocks depends in a positive way on its relative market position. We shall discuss other conditionalizing mechanisms in the section that follows.

[18] This replaces the simpler perfect foresight notion that firms act so as to maximize the present discounted value of their profit streams. Of course this rule is predicated on the idea that all actors (and especially those who supply capital) perceive the *same* stream, and that there is no uncertainty regarding its 'premature' truncation due to bankruptcy.

1.2.6. Price formation

We observe stable prices in a wide variety of real world sectors. While traditional theory must regard this pricing phenomenon as something of a theoretical anomaly, such policies make clear economic sense once it is recognized that they are implemented under and have influence upon intertemporal uncertainty.

When a firm produces consumer goods, for example, it may prefer a stable price policy on the basis of the behavioral characteristics of its generic buyer. Consider a population of 'information myopic' individuals whose natural perceptive capacities – relative to the complexity of the world in which they live – are limited. Quite heuristically, we may have in mind that what an individual 'knows' (is aware of, however 'vaguely') is a ' "diminishing function" of generalized experiential and psychological "distances" '. This intrinsic form of uncertainty is reducible to some extent through the acquisition of data but, even so, costs of search are positive and personal resource endowments are limited. Since transactions costs are also positive, individuals 'come to market' at their own discrete time epochs and do so by *first* visiting those firms which, according to their subjective preconceptions, offer greatest 'promise for satisfaction'. If these confrontations between preconception and reality (now immediate, and therefore well-perceived) are not sufficiently 'disturbing' to warrant further search, then planned purchases will be (more or less) carried through. These experiences, in any case, contribute to the preconceptions that will operate at future (individual) purchasing epochs.

Now suppose that a firm sells consumer goods or services on an essentially repetitive basis (for example, a restaurant, or a grocery store) and has decided upon a stable price policy. This firm may well acquire goodwill capital in the form of a 'clientele' market; i.e. its random demand flow is drawn from a sub-population that contains some significant 'core' of firm-loyal purchasers. The explanation is simple enough: in the absence of reasons for searching out other firms, many consumers may prefer to continue purchasing under relatively familiar conditions. More preferable conditions may well exist, but not so far as our 'myopic' consumer is aware (or is disposed to search out). The

[19] In general, when discrete individual purchase plans are (randomly) aggregated over a population stock, resultant demand activity has the character of a random flow.

habit effect, in other words, may thus be understood as a generalized form of risk-aversion in the context of a heterogeneous and differentially perceived world. By ignoring random short run signals to raise price (so long as these are thought to be 'transitory'), the firm does not chance a long-run contraction in its clientele market; for once a consumer is driven to search elsewhere, he may never return (the habit effect again, this time under more preferable conditions, newly discovered). On the other hand, because demand visits our model firm as a flow, a randomly[20] timed short-term reduction in price may not have anything like the immediate and dramatic impact that follows under perfect information assumptions, especially if it is true that event-specific advertising and word-of-mouth diffusion effects amount to little more than random noise.

A stable price policy thus appears to reduce uncertainty for both buyer and seller: what the buyer has experienced before he expects to experience again; and the firm hopes to secure a relatively stable future profit stream through the isolation of its market.

When a firm produces intermediate goods, on the other hand, preference for a stable price policy may stem from different reasoning. In a world of uncertainty firms do face an operating capital constraint (cf. footnote 16) which is divided between actual production activities and supporting precautionary and speculative reserves that are intended to promote the continuity of product flow at minimum cost. Thus, buffer input inventories may be held against the possibility of bottleneck shortages. Or, even when the bottleneck question is not at issue, input inventories may be speculatively purchased at low-price moments. In general, inventory costs tie up operating capital-time and thus reduce average rate of product flow (therefore also average quantity of inputs purchased). When these considerations obtain for the buyer of an intermediate good, the supplier firm may be led to prefer an intertemporally stable pricing policy which it supports by carrying sufficient inventory to meet 'normal' random fluctuations in demand; it is typically the case that the seller enjoys comparative cost advantages in the storage of its product. For the supplier firm such policy eliminates demand uncertainties that would otherwise derive from speculative assault on its product. On the other hand, such policy eliminates the speculative

[20] End of season 'clearance sales' do not fall under this rubric.

motive for the buyer and diminishes his concern with respect to the precautionary question. If the income effect that associates to a stable (as opposed to a random) price policy offsets the concomitant costs of storage for the supplier, then he will adopt a stable price policy which – under competitive influences – is preferred by all.

In addition to the foregoing there are reinforcing game-theoretic considerations that may also obtain regardless of the nature of the product. From the viewpoint of rival firms in an oligopolistic setting, price policy is one of the most ubiquitous and visible aspects of both firm operation and game-strategic 'intent'. Generally speaking, a history of stable prices not only reflects the feasibility of using standby capacity and operating inventories as buffers against random shocks, but may also suggest that rival firms find it mutually beneficial to thus weaken the possibility of spontaneous price warfare (regarded as suboptimal by all).

1.3. Structural Responses to Uncertainty

As with any other science, the wellspring for economic theory is, of course, the real world. Economic actors face, generate, influence and bear uncertainty in many ways. To explore the broad theoretical implications that follow from the presence of uncertainties we may be guided as to questions of significance by the structural mechanisms that exist for their manipulation. This is our approach in the sketch that follows.

To mitigate the impact of undesirable (and personally incident) contingencies, an actor may hope to trade the uncertainty atoms to which they associate in markets that are expressly constituted for this purpose; or he may attempt to influence contingency likelihoods and/or to mitigate contingency impacts through structural changes in the atom that are within his power to effect. On the collective level, society has an interest in lessening the impact of uncertainties on its members, and in particular for those who would appear to have the least ability to bear them. Legislation and other social mechanisms are institutionalized to protect such unfortunates, while to government falls the more active discretionary welfare role: through direct policy intervention, government promotes greater stability and less uncertainty in the economy and, when necessary, also acts as an insurer of last resort. We shall have a closer look at these market and non-market mechanisms in the subsections that follow.

1.3.1. The markets

An atom of uncertainty may associate to the future market value or profit stream of a real productive asset. If capital requirements for this asset are so large that no one actor is willing to assume the full burden of undesirable contingencies, then the securities market provides a mechanism for sharing the financial incidence of the atom however it may resolve. An investor simply chooses his own scale of incidence, according to his attitudes on risk-bearing and his perception of the atom, and in view of current market price and his own endowment[21]. The securities market thus plays an essential economic role with respect to uncertainty that (i) arises in the productive sector, and (ii) is predominantly borne by risk-averse investors; by diffusing the financial incidence of 'large-scale' atoms, risky activity is the more readily undertaken.

Of course the 'common' sharing of an uncertainty atom may still appear too risky for some, and a spectrum of financial instruments will frequently arise so as to discriminate prevailing risk-bearing attitudes and perceptive judgments in an 'optimal' way. Thus a corporation offers debt instruments as well as common stock. The former have priority in the event of bankruptcy, but will pay a prespecified interest however large are the returns to total capital. The terms of these instruments are adjusted by management so as to optimize the (uncertain) returns to capital (relative to the 'predominantly held' risk attitudes of its equity holders, say, and subject to other characteristics of the capital market).

Not every uncertainty atom of the asset type can, however, be shared, nor may an owner wish to have it shared. A 'sharing' market may not exist when the service-producing flow of an asset is intrinsically indivisible; home ownership is a case in point. But even when the securities market for such an asset does exist it may function quite weakly when the *control* of that asset is tied to a decision-making unit for which a significant question of moral hazard is involved. The case of human capital is an important example. Since indenture by contract is illegal,

[21] For example, when uncertainty associates to the future spot price of a given commodity, producers may wish to hedge by selling some part of their product on its current forward market.

the possibility of 'irresponsible' or incompetent behavior on the part of a debtor becomes a significant consideration for any lender. In consequence, loans are small scale and are offered only to preferred risks; left to itself, the capital market provides for an underallocation in human development. On the other hand, even when an uncertain asset is market-sharable, a decision maker may prefer to retain full ownership of that asset. Of course this will happen whenever its current market value is too low in view of the decision maker's speculative judgments concerning atomic resolution. The classic example is Knight's entrepreneur. To this person, who is more willing to bear the uncertainty than anyone else, falls whatever profit may obtain. An entrepreneur may also choose to retain full ownership if this is linked by investors to his latitudes for managerial control. Schumpeter's entrepreneur is one who earns his profits by transforming his uncertainty atom through inventive means that are not perceived by others at the crucial time; the point is that he must be free to implement these means.

Whether the decision maker purchases or sells an uncertainty atom, or some share of an atom, depends upon how he believes it may resolve. Before deciding, he may have some antecedent options for acquiring further information. He may simply let some time go by in order to observe the atom in evolution, as for example in the case of any new prospectus. Or he may perform some small scale test after the fashion of a Bayesian, to better determine what may already be the 'truth', not yet discovered; for example, an oil firm will test a new field by sending down a few taps. Or he may decide to search out possible substitute atoms, to better assess the merits of the one in question; the search for a better job opportunity and the search for the minimum price are familiar examples. Such measures for the acquisition of decision-relevant information are typically limited in principle, especially when an atom cannot be realized *except* over time. But even when it is possible to know an atom in perfect detail – as perhaps in the case of the oil firm above – this process may be too costly to effect in practice. A decision maker must weigh the costs of acquiring information against the perceptive benefits gained from it; he may accordingly choose to face some 'residual' uncertainty rather than incur the cost of further search.

A securities market provides for the common sharing of a *given* atom, however it may resolve. An insurance market, on the other hand, pools resources from a *class* of atoms that have common uncertainty characteristics in respect to an unwelcome and well-specified contingency E. More particularly, it is usually the case that random realizations of E within this class are thought to be 'reliably' governed by stochastic law, and that one such occurrence bears little or no causal relation to any other. Then members of this class may pool insurance premiums *ex ante* to spread the impact of realizations *ex post*; these would otherwise fall on the unfortunate few. Each atom bearer will pay a small known premium in exchange for mitigating the possible impact of a substantially larger loss. In most cases this function is orchestrated by insurance firms because specialization enables them to take better advantage of the law of large numbers, and it is this law on which the insurance idea rests. By reducing the financial burden of at least some sources of large-scale shock, the insurance markets thus allow both households and firms to specialize in their respective consumptive and productive activities without having to precommit large contingency reserves. Indeed, such requirements could be prohibitive in the absence of insurance possibilities, and some atoms might not be held in consequence. Insurance markets thus have a qualitative effect on aggregate scale and scope of economic activities similar to that which is promoted by the existence of securities markets.

The market for a particular contingency E may fail to form for a variety of reasons. It may fail to form if its base (those who choose to insure) is not sufficiently large. Since the 'risk (or loading)' component of an insurance premium varies inversely with the size of the base[22], and demand for insurance varies inversely with premium, this implies a threshold size for the class in question below which its market will not form.

An insurance market may also fail to exist by reason of moral hazard: according to Arrow, in situations (i) where the occurrence of E is in some measure subject to the behavioral influence of the insured, and (ii) when the insurance policy might (by its very availability) alter incentives and therefore the probabilities upon which the insurance company must rely. The existence of fire insurance, for example, might induce some

[22] This would follow, for example, if the decision rule for determining the load involved covering a selected 'loss interval' in the Neyman–Pearson sense. The point is that the distribution for average (per-atom) loss peaks more sharply toward its mean as the size of the base increases.

people to be less careful with matches. The point is this: in the absence of effective means for inspection and control, this generates a 'morally inflated' probability number $p(E)$. The market for E may function nevertheless, but if so the insurance premium is perforce inflated in both its actuarial and loading components, the latter because per-atom losses relative to the mean are inflated in the probability sense[23]. The effect on demand is as before. Indeed, an insurance market may not form at all if the effects of moral hazard are unknown in a stochastic sense (historical data do not exist) and when subjective assessments for $p(E)$ are, in consequence, too large to support the market. In the same way, insurance markets are typically nonexistent for contingencies that depend not only upon the behavior of a would-be insuree, but upon the behavior of other economic actors as well. Such endogenous uncertainties are usually singular for the situation at hand and, in general, cannot sensibly be described according to stochastic law.

It is sometimes possible to increase the base of an insurance market through the elimination or reduction of moral hazard, provided that the costs of doing so (these are typically passed on to the insuree) do not offset the demand effect of a lowered net premium. Common practices include inspection and control, and coinsurance. Insurance companies generally offer lower premiums to those that choose a policy with a deductible or some other form of coinsurance (of course there is no transaction or enforcement cost here). To purchase life and health insurances it is not uncommon that health examinations are required. In the case of fire insurance, buildings may be subject to periodic inspection and to the installation of appropriate fire prevention equipment.

When an insurance market fails to exist (or when its premium is prohibitively large), an atom bearer may choose to *self-insure* by filling his portfolio with assets that are both 'uniformly more secure' (than the atom) and 'sufficiently liquid'. Thus the savings activity is, in part, a homemade form of 'catch all' insurance. In the case of the nonexistence of a particular insurance market, savings may be less efficient than purchasing a more specialized insurance against E (were this possible)

---

[23] While the use of variance as a surrogate for 'risk' is well known to imply some strikingly counter-intuitive results, we may chance an appeal to it here for heuristic purposes: the variance of a binomial distribution increases as $p(E) \to \frac{1}{2}$. It is interesting to note that the typical insurance market is one for which $p(E)$ is 'small'.

but may be second best in the insurance sense. In the event that self-insurance is not feasible, then these uncertainty atoms will not be held and the associated economic activities will not be undertaken.

1.3.2. Private non-market mechanisms

While the factors that contribute to the resolution of an uncertainty atom may be exceedingly complex and in some intrinsic measure beyond the determination of its bearer, he may nevertheless have options for its strategic transformation; for the creation of a new isotope, as it were. The theory of the firm, as before, is a good vehicle for illustrating the conditional nature of uncertainty.

Firms have a variety of control instruments for manipulating the uncertainty characteristics of both revenue and cost streams. These include inventory, employment and price policies that are designed (i) to influence, channel or better define the behavior of those who populate the firm's markets, and (ii) to organize production in an optimal way. One of the most prevalent internal instruments for these purposes – basic for production processes that flow over time – is some form of the buffer inventory or standby capacity mechanism. By these devices the firm may smooth its product flow in the face of randomly fluctuating market conditions. When output is storable, say, then current sales are taken from buffer stock and this is replenished at the least-cost[24] convenience of the firm. Or the buffer role may be shared by the existence of standby capacity and other input stocks. When output cannot be stored, as in the production of electricity for example, then factor reserves assume this role alone. By thus internalizing 'small-scale' shock, the firm is able to achieve a more stable relationship with the outside world. Wages, input employment rate and output price – as 'external connectors' under firm control – may remain relatively stable. So far as the firm is concerned, such stabilities not only help to reduce production costs, but may also increase revenue by promoting the accumulation and conservation of goodwill capital (i.e. its markets).

On the input side of the picture, the firm may attempt to secure the relative smoothness of factor flows. By contracting with a supplier as to

[24] If, for example, random fluctuations in demand are sufficiently regular, and short-run average production cost is sensitive to scale, then a smooth-production-and-buffer-inventory policy will be cost minimizing in the stochastic sense.

quantities, prices and delivery dates, both parties tie down at least some aspects of their respective uncertainty atoms. The supplier is now more sure of demand; the firm, of supply. Such simultaneous contractual determination of both quantity and price may appear to be an *over*-determination from a *tatonnement* view of the market. But when markets flow over time and evolve under conditions of uncertainty, actors may choose to make *some* precommitments based upon their own judgments concerning priorities and their ability to meet these commitments, rather than chancing what might otherwise turn out to be 'opportunities forgone'. The point is that such contracts constitute only a *proper* subset of all transactions, and are undertaken between actors among whom a 'priority' relationship exists. The price-guided allocation mechanism continues to function, but its cutting edge is composed of those transactions that retain the 'spot' character of conventional market theory. These 'transactions on the intertemporal margin' absorb the full brunt of market uncertainties, while those under contract follow developing trends when the time for recontracting falls due.

The mechanism for securing labor supply is necessarily somewhat different because of the human non-indenture aspect noted earlier. But an entrepreneur may nevertheless be willing to enter a one-sided contract that obligates him in respect to wages and other conditions of employment (though typically not so far as long-term duration of employment is concerned, for obvious reasons). In particular, by offering a constant wage that does not depend upon random realizations of revenue flow, the firm thus accomodates the 'smooth consumption stream' preferences of its labor force. If, on the other hand, labor were paid according to its randomly fluctuating value of marginal product (as computed *ex post* according to actual sales), then higher turnovers and training costs might result with (random) short-run changes in demand; the grass may appear to be greener with respect to employment opportunities elsewhere. Of course random layoffs could result in even more dramatic losses in firm-specific human capital. Firm management is thus under a strong incentive to implement production techniques, inventory mechanisms and wage policies that promote the continuity of employment. When this continuity is threatened by labor itself, in the form of an organized threat to strike, then bargaining in good faith is a mechanism for working out management/union differences with no interruptions in product flow.

On the output side of the picture, as we have seen, firms will attempt to create income and goodwill effects for their consumers by maintaining a stable price policy. In addition, such a policy diminishes the likelihood of debilitating price competition.

1.3.3. Industrial and social organization

The ability of a firm to control future market conditions is of course limited. A particular market may have life and death characteristics of its own which are simply beyond the influence of the firm. Rather, the sense of influence is reversed, and this contributes in some essential ways to the organization of the productive sector. In general, firms have a tendency to structure so as to *internalize* and thus reduce the impact of intrinsic production and market uncertainties. More particularly, this is commonly achieved through the risk-smoothing characteristics of size and diversification.

With respect to the production process itself, for example, contingency reserves of one form or another (output and input inventories, spare parts and maintenance pools) increase proportionately less than scale because the uncertainties that call them forth increase in the same way.

With respect to the markets, there are a variety of typical organizational forms and optimal strategies that arise in response to characteristic market uncertainties. Market demand for any given commodity, for example, may be subject to sudden and permanent collapse; its market day in the sun may vanish under a shift in tastes or because of technological advances that render it economically obsolete. The firm diversifies its product mix so as to divorce its fate from that of any one product line and, more generally, in order to isolate its markets in the long run from uncertainties of the substitution type.

In addition, a firm may face the uncertainty of securing raw materials during times of prosperity and distributive outlets in times of recession. The impact of these contingencies depends upon relative market power in respect of global transformations of the competitive environment. Thus an increase in scale, perhaps through horizontal merger, may provide sufficient mass not only to weather such circumstances but indeed to capitalize on them to further improve competitive position. The firm may also consider shifting such uncertainties backward and forward through vertical integrations, at which loci they may pose a diminished threat to long-run survival. When this appears to be the case, such benefits must still be weighed against the costs – and, indeed, the uncertainties – of *entry* into new and specialized activities. In this connection it may prove optimal to arrange for something less than full integration; the franchise and principal/agent mechanisms are familiar examples[25]. By preserving the profit incentives that attend a decentralized structure, this sort of arrangement allows both parties to focus upon their comparative advantages for dealing with source-specific uncertainties, while at least some of the (dual) uncertainties that would otherwise obtain at their common transactions node are 'washed away' according to the provisions of the contractual arrangement. A national manufacturer, for example, has the size to buffer uncertainties that relate to the availability of inputs; its regional franchised outlets are better suited to dealing with local variations in demand. An individual franchise operator is assured of perfectly elastic supply up to some limit at prespecified price. The parent firm, on the other hand, enjoys demand stability through decentralization. Moreover, this decentralized and profit-sharing marketing structure promotes an optimal value for total profit, with tradeoff between short-run profit and total market share at the strategic option of the parent (through choice of costs to its franchised outlets and subject, of course, to uncertainties that associate to its own costs).

While vertical and/or horizontal reorganization thus enables a firm to insulate itself from market uncertainties, these uncertainties may remain undiminished for weaker rivals to bear in full. An oil shortage, for example, whatever its origin, impacts first upon the 'independent' companies; these depend upon allocations from the reserve inventories of the 'major' companies, and the order of priority is perfectly clear.

This last example illustrates a phenomenon that holds in greater generality. By way of reducing uncertainties that attend to their own environments, risk-averse actors often succeed in shifting them for others to bear. This contributes an interesting dimension for a welfare-conscious society, since uncertainties are typically shifted to and are borne by those for whom an unhappy realization would have the greatest relative impact[26]. We have noted above that in the productive sector

[25] Contractual arrangements between buyers and sellers are more restrictive in nature; they constitute a surrogate for vertical integration when the costs of such integration are considered to outweigh their benefits.

[26] Strictly speaking, the freedom to impose such externalities is implicit for a *laissez faire* system.

this sort of dynamic promotes the growth of monopoly power, and conversely. This tendency is countered to some extent through anti-trust law, which delimits at least some forms of integration. In this way it may be possible to prevent the shifting of contingency burdens and thus preserve a small firm sector that would otherwise tend to shrink.

Society sometimes provides institutional relief for individuals who act myopically or who may experience personal difficulties that could not 'reasonably' have been prevented. For example, Aid to Families with Dependent Children provides some help for single parents who find it impossible to rear their children with sufficient material sustenance; Medicare and Medicaid provide relief for the indigent, the aged and the poor. A welfare society might rationalize these and other such transfers as the 'indemnity' of a grand insurance scheme, with 'premiums' supplied through tax monies at large. Welfare-conscious individuals may collectively reason: there but for the vicissitudes of life goes any of us. Unemployment compensation is similar, except that employers are required to contribute to the fund; the productive sector thus assumes partial responsibility for insuring its work force against layoff. Social Security, on the other hand, requires individuals to assume responsibility for their own retirement years. This statistically rationalized and globally based insurance program provides minimal subsistence benefits for those elderly who will turn out to have suffered an inability to escape from a neighborhood of poverty.

A welfare society also challenges the traditional dictum of *caveat emptor*, on the argument that the subtleties of misrepresentation are easier to practice than they are to recognize, especially by individuals whose competence and resources for this purpose are naturally limited. Thus some information/watchdog activities may be more efficiently dispatched in the public sector. Pure food and drug legislation is a familiar example; the licensing or certification of a variety of specialized agencies (physicians, lawyers) is another.

Because of its enormous capacities for absorbing shock, the state often functions as an insurer of last resort. It is common practice, for example, to provide *ex post* relief for entire communities that have been struck by natural disaster. In addition, the state may subsidize some insurance programs that would not otherwise function adequately from the social point of view, perhaps because of significant possibilities for moral hazard, but in any case when nonsubsidized premiums would be so large as to discourage the participation of a socially optimal base. Health care insurance is a case in point.

Finally the state, through direct intervention, plays a unique role in maintaining the stability and reducing the uncertainty of the general economy. Through the use of automatic stabilizing mechanisms and other monetary and fiscal policy instruments, it thus attempts to 'correct' and stabilize random fluctuations in macrovariables, especially those which may significantly affect investor confidence, and to move the economy away from a suboptimal equilibrium. The state may also employ transfer mechanisms such as investment credit and a preferred rate for capital gains so as to mitigate uncertainties for risk-averse investors. It may focus on a particular target sector through the use of policy instruments that effectively reduce or reshape the uncertainties of that market. The price support of agricultural products and the soil bank program operate to stabilize activity in the farm sector, where production and investment are particularly sensitive to changing market conditions; in this case the concomitant manipulation of inventory reserves is, perforce, administered by the state.

PART 1

CHAPTER 2

ON THE FOUNDATIONS OF DECISION MAKING UNDER UNCERTAINTY

Peter C. Fishburn

2.1. Introduction

This paper is an exposition of the conditional subjective expected utility theory for decision under uncertainty developed in ref. [9]. The formulation used in the theory is based on three things: acts, states and extraneous scaling probabilities. The last of these is included for mathematical tractability and with an eye on the scaling of utilities as suggested by the expected utility theory of von Neumann and Morgenstern [6, 11, 25]. Further discussion of extraneous scaling probabilities will be deferred to section 2.3, following a more complete treatment of the basic act–state viewpoint that underlies our theory.

The act–state viewpoint that we shall adopt has early traces in the development of theories of games of chance and insurance, and is greatly influenced by Savage's formulation for personalistic decision theory [20]. In this formulation, an individual decision maker is to select an act (which might specify a sequence of actions to be implemented over a period of time) from a set F of acts when the holistic outcome or consequence of his decision depends both on the act selected and on which state in a set S of mutually exclusive and collectively exhaustive states of the world obtains (occurs, is realized, is the 'true state'). It is generally presumed that the decision maker is uncertain about which state is the true state and that this state (the state that obtains) is not itself affected by the act that is implemented. Jeffrey [13, 14] criticizes this latter aspect of causal independence between acts and states and develops a theory that allows for interdependence in the sense that the decision maker's probability measure on the states can differ depending on which act is selected. As noted elsewhere [6, 16] it is always possible to reformulate an apparently interdependent situation in such a way that the presumed type of independence is obtained, even though a much more cumbersome arrangement may result from such a transformation.

Although our formulation derives from Savage's, it differs in one major respect. In his theory, which is detailed in his book and in the final chapter of ref. [6], each act is a function from S into the set of consequences. We shall use act–state pairs rather than consequences. This change is caused by the fact that in most practical situations the specification of an act and a state will not determine a unique consequence. That is, there is residual uncertainty that is not removed by the states formulation. Savage's viewpoint can be seen as an idealization in which all relevant uncertainty is accounted for in the definitions of states. We have simply relaxed this ideal viewpoint.

Because of this relaxation, Savage's notions of constant acts (which assign the same consequence to each state) and other special types of acts are inapplicable. This may be just as well since these special acts do not often correspond to any real courses of action. Our approach therefore sidesteps one of the most criticized aspects of Savage's theory. But in so doing, it necessitates the adoption of substitute procedures to handle certain things that Savage deals with through the medium of his special acts.

For one thing, we shall posit direct preference comparisons between pairs of act–event pairs, of which act–state pairs are a special case. Savage applies the individual's preference relation to pairs of acts throughout his development: comparisons which correspond to our act–event comparisons are handled in his system by comparisons between special types of acts.

Second, we no longer have the natural utility comparisons among consequences under the different states that prove so useful in Savage's and others' [1, 5, 8, 18, 21] derivations of a subjective expected utility model. Because of this, we shall use a special structural axiom that permits some utility comparison between act–event pairs under different events.

These and other aspects of our formulation and theory will be developed more precisely in the next three sections. In the ensuing section we shall consider our act–event formulation and conditional subjective expected utility model apart from the use of extraneous scaling probabilities and mixed acts. Mixed acts will be introduced in section 2.3 and used in our basic set of axioms in section 2.4. Later sections present additional axioms and definitions which extend the usefulness of the basic model.

2.2. Act–Event Pairs

Since certain important aspects of the approach to decision under uncertainty that we shall examine can be discussed apart from considerations of extraneous scaling probabilities, we shall begin without the latter aspect. With this omission, two primitive sets of our theory remain. The first of these is a set F of acts $f, g, \ldots$, which are viewed as the actual or feasible courses of action open to the decision maker. The second is a set S of states of the world $s, s', \ldots$, each of which describes certain potential realizations of aspects of the decision maker's environment that are not subject to his control. Subsets of S are called *events*, which we denote by $A, B, C, \ldots$. The empty event is $\varnothing$. By 'A obtains' we mean that some $s \in A$ obtains, or that A contains the true state.

To effect some generality we shall not assume that all possible events are relevant to the concerns of the decision maker. Instead, we suppose that the set of relevant events is a Boolean algebra ε of subsets of S, with S itself in ε. This means that ε is closed under finite unions and complements: if $A, B \in \varepsilon$, then $A \cup B \in \varepsilon$, and if $A \in \varepsilon$, then $S - A = \{s\colon s \in S \text{ and } s \notin A\}$ is in ε also. Throughout, we shall let $\varepsilon' = \varepsilon - \{\varnothing\}$, so that ε' consists of all events in ε except for the empty event $\varnothing$. For $A \in \varepsilon'$, $\varepsilon(A) = \{A \cap B\colon B \in \varepsilon\}$ is the Boolean algebra on A induced by ε.

In this section the decision maker's preference relation $\succ$ is applied to act–event pairs in $F \times \varepsilon'$. We interpret $(f, A) \succ (g, B)$ as 'f given A is preferred to g given B', or that the decision maker would rather do f under the assurance that A contains the true state than do g under the assurance that B contains the true state.

There are several potential problems with this viewpoint. First, if $A \cap B = \varnothing$, the comparison between (f, A) and (g, B) may seem to require simultaneous suppositions that each of two incompatible events obtains. Actually, when comparing (f, A) and (g, B) we would expect the decision maker to imagine what might happen if f were used and some state in A were the true state, and then realign his thoughts to imagine what might happen if g were used and some state in B were the true state. His introspection about these two possibilities would then result in a preference (or indifference) judgment between the two. In this light, statements such as 'it is better to free Mr. Accused when he is in fact guilty than to convict Mr. Accused when he is in fact innocent' would be considered relevant in our approach.

Second, if A and B are not identical but have a nonempty intersection, then there is the possibility that the desired comparison between (f, A) and (g, B) might end up as a comparison between $(f, A \cap B)$ and $(g, A \cap B)$ due to a conscious or unconscious effort to reconcile the different conditioning events. Although such an effect might arise in practice, it can be minimized in the scaling procedure by avoiding comparisons between intersecting but different events[1].

There is a third aspect of our use of $\succ$ on $F \times \varepsilon'$ that, as far as I am aware, is unique in axiomatizations of subjective expected utility. This aspect involves the use of $\succ$ with pairs (f, A) in which A, while not empty, might be regarded as virtually impossible by the decision maker and have zero probability in his subjective probability measure on ε. So long as A is not logically impossible or self-contradictory, and we would expect this much of events in ε that are not equal to $\emptyset$, there seems to be no *a priori* reason to exclude such events from comparisons under $\succ$. For example, although you may assign zero probability to the event that New York City will disappear into the Atlantic Ocean before 1980, this event is not logically impossible (as of 1972) and it may not seem unreasonable to consider preferences between act–event pairs that include this event[2].

2.2.1. A conditional subjective expected utility model

The numerical representation model for $\succ$ on $F \times \varepsilon'$ that we propose consists of a real-valued utility function u on $F \times \varepsilon'$ and a finitely additive probability measure[3] P_A on $\varepsilon(A) = \{A \cap B\colon B \in \varepsilon\}$ for each $A \in \varepsilon'$ such that, for all $f, g \in F$ and $A, B, C \in \varepsilon'$,

---

1 Since Savage's theory applies $\succ$ to acts, it might seem that his approach avoids problems of this sort. However, it appears that his system poses similar if not more serious problems by the way in which relevant comparisons are assumed between special acts, which in many cases are unrealistic fictions that have little relation to available courses of action.

2 See ref. [20], p. 39, for further remarks on this aspect.

3 P_A is a finitely additive probability measure on the algebra $\varepsilon(A)$ if and only if $P_A(A) = 1$, $P_A(B) \geq 0$ for each $B \in \varepsilon(A)$, and $P_A(B \cup C) = P_A(B) + P_A(C)$ whenever $B, C \in \varepsilon(A)$ and $B \cap C = \emptyset$.

$(f, A) \succ (g, B)$ if and only if $u(f, A) > u(g, B)$, (2.1)
$u(f, A \cup B) = P_{A \cup B}(A)u(f, A) + P_{A \cup B}(B)u(f, B)$ when $A \cap B = \varnothing$, (2.2)
$P_C(A) = P_C(B)P_B(A)$ when $A \subseteq B \subseteq C$. (2.3)

Property (2.1) is the usual order-preserving property for utility.

Property (2.2) is an expectation equation. When A and B are disjoint, it says that the utility of f given $A \cup B$ equals the weighted sum of the utilities of f given A and f given B, where the weights are the decision maker's probabilities for A and B conditioned on their union $A \cup B$. (Note that $P_{A \cup B}(A) + P_{A \cup B}(B) = 1$, and that (2.2) allows the unconditional probabilities of A and B, $P_S(A)$ and $P_S(B)$, to be positive or zero.)

The natural extension of (2.2) is given by

$$u(f, A) = \int_A u(f, s)\mathrm{d}P_A(s). \tag{2.2*}$$

The correspondent of this for our mixed-act theory is discussed in section 2.6. According to (2.3), if $P(A) > 0$, where $P \equiv P_S$, then the expression displayed above is the same as

$$u(f, A) = \frac{1}{P(A)} \int_A u(f, s)\mathrm{d}P(s).$$

If $P(A) = 0$, this latter expression is inapplicable but the former expression is unaffected.

Property (2.3) is a natural chain rule for the P_A measures, and it plays an important role in extending (2.2) to (2.2*). For example, if A_1, A_2 and A_3 are three mutually disjoint events in ε' whose union equals A, then (2.2) gives

$$\begin{aligned} u(f, A) &= P_A(A_1 \cup A_2)u(f, A_1 \cup A_2) + P_A(A_3)u(f, A_3) \\ &= P_A(A_1 \cup A_2)[P_{A_1 \cup A_2}(A_1)u(f, A_1) + P_{A_1 \cup A_2}(A_2)u(f, A_2)] \\ &\quad + P_A(A_3)u(f, A_3). \end{aligned}$$

Property (2.3) applied to this then yields

$$u(f, A) = P_A(A_1)u(f, A_1) + P_A(A_2)u(f, A_2) + P_A(A_3)u(f, A_3),$$

which is a necessary prerequisite to the derivation of (2.2*).

It should be clear that we intend to obtain a probability measure P_A on $\varepsilon(A)$ for each $A \in \varepsilon'$ regardless of whether the unconditional probability of A, $P(A)$, is positive or zero. In fact it would be possible to have

$A_1 \supseteq A_2 \supseteq A_3 \ldots$ such that $P_{A_1}(A_2) = 0$, $P_{A_2}(A_3) = 0, \ldots,$ with $A_i \in \varepsilon'$ for each i. When $P(A) > 0$, P_A is completely determined from P by applying (2.3) to get $P_A(B) = P(B)/P(A)$ for each $B \in \varepsilon(A)$, but when $P(A) = 0$, P_A need not be prescribed by P. Nevertheless, it exists.

This has significance for the foundations of Bayesian decision theory. For example, the outcome of an information-producing experiment in a sequential process may have a smooth distribution with probability zero for each outcome value. Nevertheless, some value will be observed and further action will be based on this observation. Despite the zero probability for each outcome value, our theory tells us that there is a conditional probability measure over other aspects of the uncertain states for each conditioning outcome value.

One other aspect of our model deserves mention, and that is the question of uniqueness properties for u and the P_A. Clearly, if F and S are finite (or perhaps infinite), there may be more than one P that satisfies the model, and u need not be unique up to a positive linear transformation. For scaling purposes it seems desirable to ensure the usual uniqueness properties by some means or another, and we shall in fact do this through our later use of extraneous scaling probabilities.

2.2.2. Some implications of the model for $\succ$ on $F \times \varepsilon'$

Although formidable mathematical problems preclude presentation of a set of axioms for $\succ$ on $F \times \varepsilon'$ that imply the model of this section, it may be instructive to examine a few of its implications.

We shall say that a binary relation $\succ$ on a set K is a *weak order* (in the strict sense) if and only if $\succ$ on K is asymmetric [$a \succ b \Rightarrow$ not ($b \succ a$)] and negatively transitive. The latter property says that for all $a, b, c \in K$, [not ($a \succ b$) and not ($b \succ c$)] $\Rightarrow$ not ($a \succ c$) or, equivalently, $a \succ c \Rightarrow (a \succ b$ or $b \succ c)$. Define $\sim$ (indifference) and $\succsim$ (preference-or-indifference) from $\succ$ as follows:

$a \sim b$ if and only if not ($a \succ b$) and not ($b \succ a$),
$a \succsim b$ if and only if $a \succ b$ or $a \sim b$.

When $\succ$ is a weak order, $\sim$ is an equivalence (reflexive, symmetric and transitive) and $\succsim$ is transitive and complete ($a \succsim b$ or $b \succsim a$ for any $a, b \in K$).

Three simple implications of (2.1) and (2.2) are:

IMPLICATION 1. *$\succ$ on $F \times \varepsilon'$ is a weak order.*

IMPLICATION 2. *If* $A \cap B = \emptyset$ *then* $[(f, A) \succ (g, A) \,\&\, (f, B) \succ (g, B)] \Rightarrow (f, A \cup B) \succ (g, A \cup B)$, *and* $[(f, A) \succsim (g, A) \,\&\, (f, B) \succsim (g, B)] \Rightarrow (f, A \cup B) \succsim (g, A \cup B)$.

IMPLICATION 3. *If* $A \cap B = \emptyset$ *then* $(f, A) \succsim (f, B) \Rightarrow (f, A) \succsim (f, A \cup B) \succsim (f, B)$.

The first of these should make it clear that we are talking about an idealized individual with arbitrarily fine powers of preference discrimination, for in practice it seems likely [7, 15] that indifference may not be transitive, and there is evidence [22, 23] that even preference is not transitive in certain situations.

The second implication specifies two 'sure-thing' conditions [20]. With A and B two nonempty disjoint events, the first of these says that if you would rather do f than g when A is presumed to obtain, and would rather do f than g when B is presumed to obtain, then you would rather do f than g when some state in $A \cup B$ is presumed to be the true state. For example, if you would rather go to a movie than stay home if it should be snowing outside, and would rather go to a movie than stay home if it should not be snowing outside, then you would rather go to a movie than stay home. The second part of implication 2 replaces $\succ$ with $\succsim$ throughout.

The third implication is an 'averaging' condition [2] that tends to average out considerations due to different events. Suppose you consider act f and, if you could specify which of events A and B would obtain, you would specify A. That is, you prefer A to B given f, or $(f, A) \succ (f, B)$. Implication 3 then prohibits each of $(f, A \cup B) \succ (f, A)$ and $(f, B) \succ (f, A \cup B)$, where $(f, A \cup B) \succ (f, A)$ would mean that you would rather specify $A \cup B$ then A if you could 'pick' one of the two.

This might be clearer if we involve Howard's clairvoyant [12], who always knows what obtains and never tells a lie, and Savage's preference between news items [13, p. 72]. You are thinking of going to a movie tomorrow night (act f), and are concerned about whether it will snow. When you consult your clairvoyant, he will present you with one of three pronouncements:

A = 'it will not snow tomorrow night',
B = 'it will snow tomorrow night',
$C = A \cup B$ = 'I refuse to tell you what will happen tomorrow night'.

Suppose, under commitment to do f, that you would rather hear A than B. Implication 3 then says that you would just as soon hear A as C, and that you would just as soon hear C as B.

2.2.3. Two more implications

In concluding this section we mention two other implications of the model. These focus on the inclusion of nonempty events which are considered to be virtually impossible by the decision maker.

IMPLICATION 4. *If* $A \cap B = \emptyset$, *and if* $(f, A) \sim (g, A)$, $(f, B) \succ (g, B)$ *and* $(f, A \cup B) \sim (g, A \cup B)$, *then* $(h, A \cup B) \sim (h, A)$ *for all* $h \in F$.

IMPLICATION 5. *If* $A \cap B = \emptyset$, *if* $(f, A) \succ (f, B)$ *or* $(f, B) \succ (f, A)$, *and if* $(f, A \cup B) \sim (f, A)$, *then* $(h, A \cup B) \sim (h, A)$ *for all* $h \in F$.

These implications derive, respectively, from the extreme cases allowed by implications 2 and 3. According to (2.1) and (2.2), if the hypotheses of either implication 4 or 5 hold, then $P_{A \cup B}(B) = 0$, so that, by 2.2, $u(h, A \cup B) = u(h, A)$ for every h in F. Hence, by (2.1), $(h, A \cup B) \sim (h, A)$.

In terms of the clairvoyant–snow example given above, the hypotheses of implication 5, namely $(f, A) \succ (f, B)$ and $(f, A \cup B) \sim (f, A)$, signify that no snow is preferred to snow under a commitment to go to a movie tomorrow night, but that B (snow) is considered virtually impossible, which might be true if our decision maker lives in Miami and tomorrow is July 1. If this were the case then the novelty value of snow might well give $(f, B) \succ (f, A)$.

2.3. Extraneous Probabilities and Mixture Sets

The third and final set used in the axioms of this paper is $[0,1]$, the interval of real numbers from 0 to 1. The numbers in this set, denoted $\alpha, \beta, \ldots$, are viewed as probabilities for chance events that need have no direct connection to the events in ε. Appropriate extraneous events could be things like the event that a ball drawn at random from an urn containing five red and 95 green balls will be red, or the event that a pointer spun on a circular disk will come to rest within a specified arc of the disk.

Our use of extraneous events or extraneous probabilities is by no means novel in axiomatizations of subjective expected utility. Indeed, the first outline of such an axiomatization [19] uses an even-chance event (intended to have probability $\frac{1}{2}$) to scale utility, and Suppes' later 'completion' [21] of Ramsey's ideas employs the same device. Extraneous events or probabilities appear in a number of other theories [1, 4, 5, 8, 18], and even Savage [20], while not explicitly incorporating such things in his axioms, shows (pp. 38–39) how extraneous events can enter his formulation of the set of states of the world.

The extraneous probabilities in [0,1] will be used to construct mixed acts in precisely the way that is done in game theory [17, 25] or in statistical decision theory [3, 24]. With $f, g \in F$, $\frac{1}{2}f + \frac{1}{2}g$ is an even-chance mixed act, implemented by flipping a fair coin and using f if 'heads' or g if 'tails'. More generally, each mixed act will be represented by a simple probability distribution[4] x on F, with $x(f)$ the probability that f will be used if x should be adopted. For convenience we shall refer to such distributions as *Acts*.

If x and y are Acts and $0 \leq \alpha \leq 1$, then $\alpha x + (1 - \alpha)y$, the direct linear combination of x and y with $(\alpha x + (1 - \alpha)y)(f) = \alpha x(f) + (1 - \alpha)y(f)$ is also an Act. $\alpha x + (1 - \alpha)y$ could be implemented in one stage directly, or be broken into two stages by first choosing x or y according to the probabilities α and $(1 - \alpha)$ respectively and then implementing the chosen one of x and y.

2.3.1. *Mixture sets*

Having introduced Acts, whose main purpose will be to provide structural support through extraneous scaling probabilities for the derivation of the conditional subjective expected utility model, I will need no further to refer to acts since each act f is represented by the 'degenerate' Act x that has $x(f) = 1$.

Going one step beyond this in generality, we shall use the Herstein–Milnor [11] notion of a *mixture set*, which consists of a nonempty set X and a function from $[0,1] \times X \times X$ to X that satisfies the following three axioms for all $\alpha, \beta \in [0,1]$ and all $x, y \in X$:

$$1x + 0y = x, \tag{2.4}$$

$$\alpha x + (1 - \alpha)y = (1 - \alpha)y + \alpha x, \tag{2.5}$$

$$\alpha[\beta x + (1 - \beta)y] + (1 - \alpha)y = \alpha\beta x + (1 - \alpha\beta)y. \tag{2.6}$$

---

[4] $x(f) \geq 0$ for each $f \in F$, $\sum_F x(f) = 1$, and $\sum_G x(f) = 1$ for some finite subset G of F.

Our interpretation of X will of course be that α, β, ... are extraneous scaling probabilities and x, y, ... are Acts. Under this interpretation the set of Acts is clearly a mixture set. In our axioms, $\succ$ will be applied to the set $X \times \varepsilon'$ of all Act–event pairs in which the event is nonempty. (x, A) represents the composition of 'doing' Act x and having event A obtain, and $(x, A) \succ (y, B)$ is viewed in a manner analogous to our previous interpretation of $(f, A) \succ (g, B)$.

With X a mixture set and ε a Boolean algebra of events, the set $X(A) = \{(x, A): x \in X\}$ for $A \in \varepsilon'$ can be viewed as a mixture set with $\alpha(x, A) + (1 - \alpha)(y, A) = (\alpha x + (1 - \alpha)y, A)$: the event A simply tags along as an index. Thus $X \times \varepsilon'$ can be thought of as a family $\{X(A): A \in \varepsilon'\}$ of essentially similar mixture sets, one for each conditioning event in ε'. The importance of this viewpoint will be noted shortly.

In this connection it should be remarked that when $A \neq B$, the expression $\alpha(x, A) + (1 - \alpha)(y, B)$ has no meaning in our system. We have deliberately avoided any notion of mixing different events in ε', mainly because of the conflict that could result between the mixing probabilities and the decision maker's beliefs about the relative likelihoods of A and B containing the true state.

2.3.2. Linear utility

As a final prelude to our axioms, we recall the axioms of Herstein and Milnor [11] for $\succ$ on a mixture set. To make clear the connection between their axioms and ours, and to indicate how their axioms will be used in the derivation of our model, we state their axioms for $\succ$ on the mixture set $X(A)$ as follows:

$$\succ \text{ on } X(A) \text{ is a weak order,} \tag{2.7}$$

$$\{\alpha: (\alpha x + (1 - \alpha)y, A) \succsim (z, A)\} \text{ and } \{\alpha: (z, A) \succsim (\alpha x + (1 - \alpha)y, A)\}$$
$$\text{are closed (in the relative usual topology for } [0,1]), \tag{2.8}$$

$$(x, A) \sim (y, A) \Rightarrow (\tfrac{1}{2}x + \tfrac{1}{2}z, A) \sim (\tfrac{1}{2}y + \tfrac{1}{2}z, A). \tag{2.9}$$

For the given conditioning event $A \in \varepsilon'$, these are to hold for all x, y, $z \in X$. The second axiom, (2.8), is a form of continuity or Archimedean axiom that is required to obtain a one-dimensional (real-valued) linear order-preserving utility function. The third axiom is a weak form of sure-thing axiom. Presuming that A obtains, (2.9) says that if you are indifferent between Acts x and y, then you should be (or will be) indifferent between Acts $\frac{1}{2}x + \frac{1}{2}z$ and $\frac{1}{2}y + \frac{1}{2}z$.

As Herstein and Milnor prove, (2.7), (2.8) and (2.9) imply that there is a real-valued function u on $X(A)$ such that, for all $x, y \in X$ and $\alpha \in [0,1]$

$$(x, A) \succ (y, A) \text{ if and only if } u(x, A) > u(y, A),$$
$$u(\alpha x + (1 - \alpha)y, A) = \alpha u(x, A) + (1 - \alpha)u(y, A).$$

The first of these is similar to (2.1) in its order preservation. The second is a linearity or expectation form. When it holds, we say that u is *linear* on $X(A)$. If u on $X(A)$ satisfies the above expressions, then so does v on $X(A)$ if and only if $v = au + b$ for real numbers $a > 0$ and b. We abbreviate this by saying that u is *unique up to a positive linear transformation.*

The first three axioms in our system imply (2.7), (2.8) and (2.9) for each $A \in \varepsilon'$. Hence, for each $A \in \varepsilon'$, there is a linear order-preserving function on $X(A)$ which is unique up to a positive linear transformation. Our approach is to align these separate functions, by appropriate positive linear transformations, so that they provide one overall function u on $X \times \varepsilon'$ for which $(x, A) \succ (y, B)$ if and only if $u(x, A) > u(y, B)$. We pursue this further in the next section.

2.4. *The Basic Model*

Having developed the basic components of our theory and suggested the type of representation model that we are interested in, we turn to the axioms and their implications. Throughout this and succeeding sections, X is a mixture set, ε is a Boolean algebra of events, $\varepsilon' = \varepsilon - \{\varnothing\}$, and $\succ$ is a binary relation on $X \times \varepsilon'$.

Our general system includes eight axioms, the first six of which are put forth in this section. The seventh axiom enables the derivation of appropriate properties for probability measures and is discussed in the next section. The eighth axiom permits the extension of our basic model to an integral form like (2.2*) under conditions which are described in section 2.6.

2.4.1. The first six axioms

The following six axioms apply to all A, B, $C \in \varepsilon'$ and all $x, y, z, w \in X$:

AXIOM 2.1. *$\succ$ on $X \times \varepsilon'$ is a weak order.*

AXIOM 2.2. $\{\alpha : (\alpha x + (1 - \alpha)y, A) \succsim (z, B)\}$ *and* $\{\alpha : (z, B) \succsim (\alpha x + (1 - \alpha)y, A)\}$ *are closed.*

AXIOM 2.3. $[(x, A) \sim (z, B) \,\&\, (y, A) \sim (w, B)] \Rightarrow (\frac{1}{2}x + \frac{1}{2}y, A) \sim (\frac{1}{2}z + \frac{1}{2}w, B)$.

AXIOM 2.4. *If* $A \cap B = \varnothing$ *then* $(x, A) \succsim (x, B) \Rightarrow (x, A) \succsim (x, A \cup B) \succsim (x, B)$.

AXIOM 2.5. $(x, S) \succ (y, S)$ *for some* $x, y \in X$.

AXIOM 2.6. *If* $A \cap B = \varnothing$ *then* $(x, A) \succ (x, B) \,\&\, (y, B) \succ (y, A)$ *for some* $x, y \in X$.

The first three axioms are extensions of the Herstein–Milnor axioms (2.7), (2.8), (2.9) of the preceding section. Axiom 2.1 applies $\succ$ to all of $X \times \varepsilon'$ and clearly implies (2.7) for each $X(A) = \{(x, A) : x \in X\}$. Axiom 2.2 differs from (2.8) only to the extent that different conditioning events may be used: when $A = B$ in axiom 2.2, (2.8) results. Axiom 2.2 is the continuity or Archimedean axiom of our system.

The third axiom extends (2.9), since the latter is obtained from axiom 2.3 by setting $B = A$ and $w = y$. (Since $(y, A) \sim (y, A)$ is presumed, axiom 2.1 is involved also in a minor way.) Although axiom 2.3 is a fairly weak form of sure-thing axiom, based entirely on indifference, it is the only direct sure-thing axiom in our system with the exception of axiom 2.8 (in section 2.6). The correspondent in the $X \times \varepsilon'$ system to the sure-thing implication 2 of section 2.2 is the following:

If $A \cap B = \varnothing$ *then* $[(x, A) \succ (y, A) \,\&\, (x, B) \succ (y, B)] \Rightarrow (x, A \cup B) \succ (y, A \cup B)$, *and* $[(x, A) \succsim (y, A) \,\&\, (x, B) \succsim (y, B)] \Rightarrow (x, A \cup B) \succsim (y, A \cup B)$.

As can be seen from theorem 2.1 below, this is implied by axioms 2.1 through 2.6. Part of this sure-thing principle can be obtained from axioms 2.1 and 2.4. For example, if $A \cap B = \varnothing$, $(x, A) \succsim (y, A)$, $(x, B) \succsim (y, B)$ and if, in addition, $(x, B) \succsim (y, A)$ and $(x, A) \succsim (y, B)$, then axioms 2.1

and 2.4 imply $(x, A \cup B) \succsim (y, A \cup B)$. But, generally speaking, others of axioms 2.1 through 2.6 are required to establish the foregoing.

Axiom 2.4 is an averaging condition that is the direct copy of implication 3 for the present context. Our previous comments on implication 3 apply equally well to axiom 2.4.

The fifth and sixth axioms are structural conditions which, unlike the first four axioms, are not wholly necessary for the model presented in theorem 2.1. Axiom 2.5 is rather innocuous since it simply says that there are two nonindifferent Acts.

On the other hand, axiom 2.6 is a major structural restriction. It will fail if and only if there are disjoint events A and B in ε' such that $(x, A) \succsim (x, B)$ for *every* $x \in X$. That is, if there are nonempty disjoint events such that, for every Act, the decision maker would just as soon see the first event obtain as see the second obtain, then axiom 2.6 is false. It is not difficult to construct simple examples which violate the axiom, and in such cases the only way to salvage our system is to introduce some artificial prizes or penalties, or some artificial acts, that will rectify the difficulty.

Nevertheless, axiom 2.6 is generally less restrictive than axioms or constructions in other systems which serve a similar purpose, namely to allow comparison or direct alignment of preferences under different events. For example, Savage's formulation, which takes all consequences as relevant under every event, obtains complete interevent comparisons and, in effect, identical utility ranges under each event. In our theory, each pair A, B of disjoint events in ε' has an act z for which $(z, A) \sim (z, B)$, but there may be only one z which satisfies this indifference statement.

In light of our remarks at the end of the preceding section, axiom 2.6 (along with axioms 2.4 and 2.5) plays a crucial role in the alignment of the linear order-preserving utility functions defined on the $X(A)$. If something like this axiom is not used, then it may be impossible to obtain the basic model described in theorem 2.1. Further details on this point and on the alignment process are presented in section 3 of ref. [9].

2.4.2. The basic model

Our so-called basic model is summarized in the following theorem. Its proof and that for theorem 2.2 are given in ref. [9].

THEOREM 2.1. *Suppose that axioms* 2.1 *through* 2.6 *hold. Then there is a real-valued function u on* $X \times \varepsilon'$ *and, for each* $A, B \in \varepsilon'$ *for which* $A \cap B = \emptyset$, *there are unique nonnegative real numbers* $P_{A \cup B}(A)$ *and* $P_{A \cup B}(B)$ *that sum to* 1, *such that, for all* $x, y \in X$ *and all* $A, B \in \varepsilon'$,

(i) $(x, A) \succ (y, B)$ *if and only if* $u(x, A) > u(y, B)$,

(ii) *u is linear on* $X(A)$,

(iii) $u(x, A \cup B) = P_{A \cup B}(A)u(x, A) + P_{A \cup B}(B)u(x, B)$ *when* $A \cap B = \emptyset$.

Moreover, when (i)–(iii) *hold, u on* $X \times \varepsilon'$ *is unique up to a positive linear transformation.*

This basic model has many of the properties of the model discussed in section 2.2. Because of our use of Acts in place of acts (or the use of extraneous scaling probabilities), the model of theorem 2.1 has the linearity property (ii) along with the stated uniqueness properties.

The deficiency of the basic model when compared to the section 2.2 model lies in the omission of any mention of probability measures or the chain rule $P_C(A) = P_C(B)P_B(A)$ when $A \subseteq B \subseteq C$. To be sure, $P_{A \cup B}(A) + P_{A \cup B}(B) = 1$, so that these numbers behave somewhat like probabilities and could be estimated or scaled with the use of (iii) once u has been estimated, but axioms 2.1 through 2.6 *do not* imply that each P_A is finitely additive or that the chain rule holds. We explore this further in the next section.

2.5. *Finitely Additive Probability Measures*

To examine further the question of probabilities on events under axioms 2.1 through 2.6, we shall first present an example used in ref. [9].

The example has two acts and three states. We shall let x denote the Act that has probability x for the first act and probability $1 - x$ for the second. With A, B and C the three single-state events, $\varepsilon' = \{A, B, C, A \cup B, A \cup C, B \cup C, S\}$. For the quantities described in theorem 2.1 let

$$\begin{array}{ll} P_{A \cup B}(A) = 0.6, & P_{A \cup B}(B) = 0.4, \\ P_{A \cup C}(A) = 0.1, & P_{A \cup C}(C) = 0.9, \\ P_{B \cup C}(B) = 0, & P_{B \cup C}(C) = 1.0, \\ P(A) = 0.1, & P(B \cup C) = 0.9, \\ P(B) = 0, & P(A \cup C) = 1.0, \\ P(C) = 0.5, & P(A \cup B) = 0.5, \end{array} \qquad (P \equiv P_S)$$

and

$$\begin{aligned}
u(x, A) &= x,\\
u(x, B) &= 1 - x,\\
u(x, C) &= \tfrac{1}{2},\\
u(x, A \cup B) &= 0.4 + 0.2x,\\
u(x, A \cup C) &= 0.45 + 0.1x,\\
u(x, B \cup C) &= \tfrac{1}{2},\\
u(x, S) &= 0.45 + 0.1x.
\end{aligned}$$

It is easily checked that (ii) and (iii) of theorem 2.1 hold, and (i) holds on defining $\succ$ as dictated by u in (i). Moreover, it is not hard to see that all of axioms 2.1 through 2.6 hold. Hence the P values specified above cannot be altered. Note that these values are all nonnegative, and each row pair sums to 1 as specified in the theorem. However, P is not additive since $P(A) + P(B) = 0.1$ and $P(A \cup B) = 0.5$. In addition, the chain condition fails since $P(A) = 0.1$ and $P(A \cup B)P_{A \cup B}(A) = 0.3$.

2.5.1. Axiom 2.7

Clearly then, something more than axioms 2.1 through 2.6 is required to obtain the desired probability aspects of the model in section 2.2. To rectify this problem we shall use a seventh axiom that, like axiom 2.6, is a special structural condition.

Axiom 2.7. *If A, B and C are mutually disjoint events in ε', and if there is an $x \in X$ such that $(x, A) \sim (x, B)$, then there is a $y \in X$ such that exactly two of (y, A), (y, B) and (y, C) are indifferent.*

As noted in the preceding section, if $A \cap B = \varnothing$, then axioms 2.1 through 2.6 imply that $(x, A) \sim (x, B)$ for some $x \in X$. Hence, so long as ε' contains three mutually disjoint events, the hypotheses of axiom 2.7 are guaranteed for each such triple of events by our previous axioms. In each case, then, axiom 2.7 requires that there be an Act y, which may or may not equal x, such that exactly two of (y, A), (y, B) and (y, C) are indifferent (with the third preferred to or less preferred than the indifferent two).

The failure of axiom 2.7 in our example is easily noted. There is only one x there, namely $x = \frac{1}{2}$, for which two of $u(x, A)$ $u(x, B)$ and $u(x, C)$ are equal. But when $x = \frac{1}{2}$ *all three* of these utility values are equal, and therefore axiom 2.7 fails.

In essence, axiom 2.7 provides precisely the type of structure that, in the context of the previous axioms, is required to establish additivity for the P_A functions. Readers who are interested in the details of this should consult the final subsection of ref. [9].

With the addition of the new axiom we have

THEOREM 2.2. *Suppose axioms* 2.1 *through* 2.7 *hold, with u and the* P_A *having the properties described in theorem* 2.1. *Then* P_A *on* $\varepsilon(A)$ *is a finitely additive probability measure for each* $A \in \varepsilon'$, $P_C(A) = P_C(B)P_B(A)$ *whenever* $A \subseteq B \subseteq C$ *and* $A, B, C \in \varepsilon'$, *and if* $x \in X$ *and* $A_1, \ldots, A_n$ *are mutually disjoint events in* ε' *whose union equals* A, *then*

$$\text{(iv)} \qquad u(x, A) = P_A(A_1)u(x, A_1) + P_A(A_2)u(x, A_2) + \ldots + P_A(A_n)u(x, A_n).$$

As seen by the example that began this section, expression (iv), extending (iii) of theorem 2.1, may be false when only axioms 2.1 through 2.6 are assumed.

2.6. Measurable and Bounded Conditional Acts

If ε is finite then theorems 2.1 and 2.2 give a complete account of the conditional subjective expected utility model that arises from our axioms. However, when ε is infinite, it remains to consider the extension of (iv) to

$$\text{(v)} \qquad u(x, A) = \int_A u(x, s)\mathrm{d}P_A(s) \qquad \text{for } (x, A) \in X \times \varepsilon',$$

where (x, s) is an abbreviation for $(x, \{s\})$.[5] Since $u(x, s)$ is defined in our system only if $\{s\} \in \varepsilon'$, we assume $\{s\} \in \varepsilon'$ for all $s \in S$.

Special definitions will be used in our examination of (v). First, $(x, A) \in X \times \varepsilon'$ is *measurable* if and only if $\{s: u(x, s) \in I\} \cap A$ is an event in ε for each interval I of real numbers. Equivalently, (x, A) is measurable if and only if $\{s: s \in A \text{ and } u(x, s) < a\} \in \varepsilon$ and $\{s: s \in A \text{ and } u(x, s) > a\} \in \varepsilon$ for each real number a. It follows easily from this and our previous assumptions that (x, A) is measurable if and only if $A \cap \{s: (y, t) \succ (x, s)\} \in \varepsilon$ and $A \cap \{s: (x, s) \succ (y, t)\} \in \varepsilon$ for each $(y, t) \in X \times S$. If ε were the set of all subsets of S, as in Savage's theory, then each (x, A) would

5 See section 10.3 of ref. [6] and chapter 9 of ref. [10] for definitions of the integration process involved here.

be measurable. If some (x, A) were not measurable in our system, then the last characterization of measurability given above indicates how ε could be expanded to ensure measurability without necessarily requiring all subsets of S to be in ε.

Measurability is introduced for technical reasons, for if (x, A) were not measurable then, even under axiom 2.8 presented below, the integral in (v) need not be well defined.

A measurable (x, A) is *bounded below* (with probability 1) if and only if $P_A(\{s: s \in A \text{ and } u(x, s) \geq a\}) = 1$ for some real number a, and *bounded above* if and only if $P_A(\{s: s \in A \text{ and } u(x, s) \leq b\}) = 1$ for some real number b. Finally, (x, A) is *bounded* if it is bounded above and below.

2.6.1. Axiom 2.8

The axiom that we use in a partial extension of (iv) to (v) is a general sure-thing or dominance principle that is related to Savage's final axiom (P7) and to axioms in refs. [5], [6] and [8] that we developed together. Later in this section we comment on another condition that ensures (v) for all measurable (x, A). The following applies to all $x, y \in X$ and all $A, B \in \varepsilon'$.

AXIOM 2.8. $[(x, s) \succ (y, B)$ *for all* $s \in A] \Rightarrow (x, A) \succsim (y, B)$,
$[(x, A) \succ (y, s)$ *for all* $s \in B] \Rightarrow (x, A) \succsim (y, B)$.

The first part of this says that if x given s is preferred to y given B for each s in A, then x given A is preferred or indifferent to y given B. The second part is similar. Axiom 2.8 is implied by our other axioms when S is finite, but is not implied by them when S is infinite.

THEOREM 2.3. *Suppose axioms* 2.1 *through* 2.8 *hold, with u and the* P_A *as specified in theorems* 2.1 *and* 2.2. *If* $(x, A) \in X \times \varepsilon'$ *is measurable then*

(*a*) $\int_A u(x, s) \mathrm{d}P_A(s)$ *is well defined and finite,*

(*b*) $u(x, A) \geq \int_A u(x, s) \mathrm{d}P_A(s)$ *if* (x, A) *is bounded below,*

(*c*) $u(x, A) \leq \int_A u(x, s) \mathrm{d}P_A(s)$ *if* (x, A) *is bounded above, and*

(*d*) $u(x, A) = \int_A u(x, s) \mathrm{d}P_A(s)$ *if* (x, A) *is bounded.*

As far as I have been able to determine, axioms 2.1 through 2.8 do not imply that u is bounded, although boundedness does arise in some other theories [5, 8], including Savage's (see ref. [6], p. 206). Of course if u is bounded then (v) holds for every measurable (x, A) as noted in part (*d*) of theorem 2.3.

The proof of part (*d*) is essentially the same as the proof of theorem 2 in ref. [9]. The role of axiom 2.8 in this proof lies in its implication that $u(x, A) \geq a$ when $P_A(\{s\colon u(x, s) \geq a\} \cap A) = 1$, and $u(x, A) \leq b$ when $P_A(\{s\colon u(x, s) \leq b\} \cap A) = 1$.

The other parts of theorem 2.3 follow easily from part (*d*). Since $u(x, A) = P_A(A^+)u(x, A^+) + P_A(A^-)u(x, A^-)$ for an arbitrary measurable (x, A) when A^+ and A^-, defined by $A^+ = A \cap \{s\colon u(x, s) \geq 0\}$ and $A^- = A \cap \{s\colon u(x, s) < 0\}$, are nonempty, it will suffice to consider an A for which $u(x, s) \geq 0$ for all $s \in A$.

Thus, suppose that (x, A) is measurable, that $u(x, s) \geq 0$ for all $s \in A$, and that (x, A) is unbounded above. Let

$$A_n = A \cap \{s\colon u(x, s) < n\},$$
$$B_n = A \cap \{s\colon u(x, s) \geq n\},$$

for $n = 1, 2, \ldots$, with $A = A_n \cup B_n$, $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$, $\bigcup_{n=1}^{\infty} A_n = A$, and $u(x, A) = P_A(A_n)u(x, A_n) + P_A(B_n)u(x, B_n)$ provided that A_n, $B_n \in \varepsilon'$. Since $B_n \in \varepsilon'$ for all n (by measurability and unboundedness) and $A_n \in \varepsilon'$ for all sufficiently large n, we can assume with no real loss in generality that A_n, $B_n \in \varepsilon'$ for all n.

Now axiom 2.8 guarantees that $u(x, A_n) \geq 0$ and $u(x, B_n) \geq n$ for all n, so that $u(x, A) \geq nP_A(B_n)$ for all n. Since $u(x, A)$ is finite, this requires that $P_A(B_n) \to 0$ as $n \to \infty$, so that $P_A(A_n) \to 1$ as $n \to \infty$. By definition,

$$\int_A u(x, s)\mathrm{d}P_A(s) = \sup\{\int_{A_n} u(x, s)\mathrm{d}P_A(s) + nP_A(B_n)\colon n = 1, 2, \ldots\},$$

where the expression in braces does not decrease as n increases, and

$$\int_{A_n} u(x, s)\mathrm{d}P_A(s) = P_A(A_n) \int_{A_n} u(x, s)\mathrm{d}P_{A_n}(s) = P_A(A_n)u(x, A_n),$$

using part (*d*) of theorem 2.3. Thus $\int_A u(x, s)\mathrm{d}P_A(s)$ equals the least upper bound of $\{P_A(A_n)u(x, A_n) + nP_A(B_n)\}$, and since

$$u(x, A) = P_A(A_n)u(x, A_n) + P_A(B_n)u(x, B_n) \geq P_A(A_n)u(x, A_n) + nP_A(B_n),$$

it follows that $u(x, A) \geq \int_A u(x, s)\mathrm{d}P_A(s)$, which in effect verifies part (*b*) of theorem 2.3. Part (*c*) is proved in a similar way, and part (*a*) follows from these.

2.6.2. A final condition

The foregoing analysis, besides establishing the parts of theorem 2.3 that do not appear in ref. [9], shows what is required in addition to axioms 2.1 through 2.8 to obtain (v) for all measurable Act–event pairs, bounded or unbounded. I will state this condition in terms of u and the P_A although it is not hard to see how the u part can be replaced by statements which use $\succ$.

CONDITION 2.1. *Suppose $A, A_1, A_2, \ldots \in \varepsilon'$ with $A_1 \subseteq A_2 \subseteq \ldots$ and $\bigcup_{n=1}^{\infty} A_n = A$. Then $u(x, A_n) \to u(x, A)$ as $n \to \infty$ if $P_A(A_n) \to 1$ as $n \to \infty$.*

With $\{A_n\}$ an increasing sequence of events whose limit is A, this says that if the likelihood that A_n obtains, given that A obtains, approaches certainty as n gets large, then the utility of x given A_n approaches the utility of x given A as n gets large. Hence condition 2.1 has some intuitive appeal. The condition says nothing about the behavior of $\{u(x, A_n)\}$ relative to $u(x, A)$ when $P_A(A_n)$ does not approach unity, as could happen if P_A is not countably additive.

The effect of condition 2.1 on a measurable, unbounded-above (x, A) with $u(x, s) \geq 0$ for all $s \in A$ is easily noted. Since $P_A(A_n) \to 1$ in our previous analysis of this case, condition 2.1 requires $u(x, A_n) \to u(x, A)$, and hence that $P_A(A_n)u(x, A_n) \to u(x, A)$. Since $nP_A(B_n) > 0$ for all n, we conclude that $\sup\{P_A(A_n)u(x, A_n) + nP_A(B_n)\} \geq u(x, A)$, or $\int_A u(x, s)\mathrm{d}P_A(s) \geq u(x, A)$. Along with part (*b*) of theorem 2.3, this gives (v).

It follows that (v) holds for all measurable $(x, A) \in X \times \varepsilon'$ when axioms 2.1 through 2.8 and condition 2.1 hold.

References

[1] Anscombe, F. J. and Aumann, R. J. A definition of subjective probability. *Annals of Mathematical Statistics,* **34** (1963), 199–205.
[2] Bolker, E. D. A simultaneous axiomatization of utility and subjective probability. *Philosophy of Science,* **34** (1967), 333–340.
[3] Ferguson, T. S. *Mathematical Statistics: a decision theoretic approach.* Academic Press: New York (1967).
[4] Fishburn, P. C. Preference-based definitions of subjective probability. *Annals of Mathematical Statistics,* **38** (1967), 1605–1617.
[5] Fishburn, P. C. A general theory of subjective probabilities and expected utilities. *Annals of Mathematical Statistics,* **40** (1969), 1419–1429.
[6] Fishburn, P. C. *Utility Theory for Decision Making.* Wiley: New York (1970).
[7] Fishburn, P. C. Intransitive indifference in preference theory: a survey. *Operations Research,* **18** (1970), 207–228.
[8] Fishburn, P. C. Subjective expected utility with mixture sets and Boolean algebras. *Annals of Mathematical Statistics,* **43** (1972), 917–927.
[9] Fishburn, P. C. A mixture-set axiomatization of conditional subjective expected utility. *Econometrica,* **41** (1973), 1–25.

[10] Fishburn, P. C. *Mathematics of Decision Theory.* UNESCO: Mouton (1972).
[11] Herstein, I. N. and Milnor, J. An axiomatic approach to measurable utility. *Econometrica,* **21** (1953), 291–297.
[12] Howard, R. A. The foundations of decision analysis. *IEEE Transactions on Systems Science and Cybernetics,* **SSC-4** (1968), 1–9.
[13] Jeffrey, R. C. *The Logic of Decision.* McGraw-Hill: New York (1965).
[14] Jeffrey, R. C. Ethics and the logic of decision. *Journal of Philosophy,* **62** (1965), 528–539.
[15] Luce, R. D. Semiorders and a theory of utility discrimination. *Econometrica,* **24** (1956), 178–191.
[16] Luce, R. D. and Krantz, D. H. Conditional expected utility. *Econometrica,* **39** (1971), 253–271.
[17] Luce, R. D. and Raiffa, H. *Games and Decisions.* Wiley: New York (1957).
[18] Pratt, J. W., Raiffa, H. and Schlaifer, R. The foundations of decision under uncertainty: an elementary exposition. *Journal of the American Statistical Association,* **59** (1964), 353–375.
[19] Ramsey, F. P. Truth and probability. In F. P. Ramsey, *The Foundations of Mathematics and other logical essays.* Harcourt, Brace and Co.: New York (1931). (Reprinted in Kyburg and Smokler (eds.), *Studies in Subjective Probability.* Wiley: New York (1964).)
[20] Savage, L. J. *The Foundations of Statistics.* Wiley: New York (1954).
[21] Suppes, P. The role of subjective probability and utility in decision-making. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955,* **5** (1956), 61–73.
[22] Tversky, A. Intransitivity of preferences. *Psychological Review,* **76** (1969), 31–48.
[23] Weinstein, A. A. Individual preference intransitivity. *Southern Economic Journal,* **34** (1968), 335–343.
[24] Weiss, L. *Statistical Decision Theory.* McGraw-Hill: New York (1961).
[25] von Neumann, J. and Morgenstern, O. *Theory of Games and Economic Behavior,* 2nd edition. Princeton University Press: Princeton, New Jersey (1947).

COMMENTS

*On recent developments in subjective expected utility**

Michael Balch

C2.1.1. Introduction

The theory of subjective expected utility that Professor Fishburn proposes here[1] amounts essentially to an axiomatic recasting of the classical Savage [8] paradigm for decision making under uncertainty, and is significant for several reasons.

First, the Fishburn variant sidesteps a structural requirement of the Savage system that had markedly circumscribed the applicability of the paradigm. To see that limitation, recall that in the subjectivist framework the subject chooses (and implements) an *act* in the face of uncertainty, uncertainty is·resolved when nature 'chooses' a *state*, and the conjunction of these results in a *consequence* to the subject. In the Savage system, acts are regarded as mappings from a set of states to a set of consequences, and the subject is required (this is the delimiting structural condition noted above) to express binary preference over the set of *all* such *formal* mappings. The subject is thus called upon to make sense of the generic 'act' which associates an arbitrarily prescribed consequence c to an arbitrarily prescribed state s. The difficulty is that this does not allow for the analytic treatment of 'consequences' which cannot be logically characterized (conceived of) *except through the occurrence of a given distinguished state.* And of course, this is the commonplace; for example, the 'consequence' to me of buying a bicycle today depends intrinsically upon whether or not my right leg is broken tomorrow. Fishburn's solution is simple and to the point: regard acts and states as the basic contextual primitives (as indeed they are), and

---

* This is an expanded version of a comment on the paper presented by Peter Fishburn [4] at the Third NSF–NBER Conference on Decision Rules and Uncertainty, Iowa City, May 1972. I am indebted to Sam Wu for his critical comments on the preliminary draft, and to Bert Schoner for some highly stimulating discussion throughout.

[1] This theory is developed, complete with proofs, in ref. [3].

require binary preference over probability mixtures of their conjunctions. We shall have an interpretive look at this in the sections to follow; suffice it to say here that the system does enjoy conceptual integrity.

Second, the Fishburn development achieves an explicit axiomatic synthesis between the subjectivist and frequentist traditions by incorporating the streamlined behavioral assumptions of Herstein and Milnor [5] in very nearly pure form; this is perhaps not surprising, given the sketch which closes the paragraph above. These axioms are instrumental in relating attitudes toward 'objectively' defined risk and subjective 'degree of belief' judgments for more general uncertainty contexts (those which are not necessarily presented to the subject in 'lottery' form). Of course the Savage system accomplishes this end as well (indeed, it blazed that trail), but in a way that (a) makes essential use of 'acts' which I cannot (in general) comprehend, and (b) blurs the distinction between uncertainty which arises in 'natural' fashion from the particular decision context at hand and uncertainty which for scaling purposes is 'fair-spinner induced'. Still, his explicit Herstein–Milnor substructure notwithstanding, it is by no means obvious from Fishburn's existence proof how it is that his subjective probability measures are *conceptually* generated. In the final section of this paper I will show[2] that the Fishburn mixture system extends in a unique way to one which admits an elementary conceptual mechanism by which subjective probabilities are *directly* scaled.

Finally, the Fishburn theory bears some close similarities to – and thus helps to cast new light upon – the recent 'conditional' model of Luce and Krantz [7]. This work addresses the long-standing need for a general theory of choice under uncertainty which does not require (as does the Savage theory) that acts be without (causal) influence in determining the state that obtains. The Luce–Krantz framework, however, incorporates a structural condition (the strong algebraic closure properties required for the choice set of the theory) that leads at once to an interpretive paradox. More specifically[3], the set of alternatives

[2] Professor Fishburn informs me that he had worked through a similar extension before arriving at the final version of his theory, but rejected it at the time in the face of the Bolker objection (to be discussed in the final section of this paper). We shall see, however, that one can provide an appropriate set of mind experiments with respect to which the Bolker objection is sidestepped.

[3] We shall have a closer look at this in the next section, after the Luce–Krantz system has been properly introduced.

with respect to which choice is exercised is required to be closed under two operations (disjoint union and non-null restriction, by name) which together with the semantic understanding given for the primitives of the system imply a model-theoretic paradox of the Russell type. For to begin with, every alternative is itself to be regarded as atomic in the sense that, once implemented, the subject has no further influence over what may be the outcome. But this understanding is at once contradicted by the structural requirements just noted: non-null restriction creates a new atomic alternative by 'decomposing' what was supposed to have been an *atomic* alternative itself; and disjoint union goes the other way around by 'combining' two atomic alternatives to create what is supposed to be a new *atomic* one. Since these closure properties are structurally necessary for the Luce–Krantz theory, we might hope to reinterpret the primitives of their system in order to avoid the paradox just noted. But somehow or other we must clearly identify the atomic alternatives; otherwise we have model-theoretic confusion over which (uncertain) environments a decision maker may or may not choose to inhabit. Although Fishburn avoids this kind of difficulty, and despite some close resemblances between his *act–state* pairs and the *conditional decisions* of Luce and Krantz (these are the (not quite) atomic alternatives of their system), it is nevertheless the case that the Fishburn primitives are not fully conditional in the Luce–Krantz sense. In the post-conference paper which follows this one, Balch and Fishburn [1] develop a theory that fills this gap.

Before leaving these introductory remarks, it is worth noting that both the Savage and Luce–Krantz difficulties, though somewhat different, appear in common to owe to a basic conflict between competing model-theoretic imperatives. On the one hand, if we aim at the kind of representational mode that is characteristic for models of subjective expected utility, then we must provide sufficient structural richness within the system. On the other hand, there is the argument that the subject should express binary preference *only* over the set of atomic alternatives, rather than over the (larger) set of possible consequences. Indeed, this is the tradition, beginning already with the axiomatic cornerstone of von Neumann and Morgenstern. Fishburn departs (as do Balch and Fishburn) from this requirement at some cost in model-theoretic 'testability'; i.e. if a consequence cannot be presented as 'available for choice' (but rather has the character of a 'contingency', given choice of some act), then binary preference between *such* things cannot be put to the *behavioral* test. But this seems 'second order' to me in comparison with the conceptual hiatus that appears when a decision maker is presented with an atomic alternative that he finds impossible to interpret[4].

C2.1.2. The paradox

We now take a closer look at the Luce–Krantz system. The primitive $\mathscr{A}$ is an algebra of subsets, called conditioning events, of a given set of chance outcomes[5]. The primitive $\mathscr{D}$ is a set of mappings, called conditional decisions, each of which is defined on some element of $\mathscr{A}$ and takes image in a set of consequences $\mathscr{C}$ (also primitive). The generic conditional decision f_α is indexed by its domain of definition $\alpha \in \mathscr{A}$; every element in $\mathscr{D}$ is presumed available to the decision maker. The idea for regarding f_α to be defined only on α is the authors' observation that choice of an action invariably has an effect on the chance outcomes that can occur; for example, if I choose to travel from 'here' to 'there' by car (rather than by airplane, say) then I cannot die on this trip in an airplane crash. Once an action f_α is chosen, the uncertainty as to which outcome obtains (will actually occur) does not disappear, it simply 'redistributes' over the subsets of α in the appropriate way. Axiom 1 of Luce and Krantz requires that $\mathscr{D}$ enjoy algebraic closure with respect to the operations of:

(i) Disjoint union: if $f_\alpha, g_\beta \in \mathscr{D}$ and $\alpha \cap \beta = \emptyset$, then

$$f_\alpha \cup g_\beta \underset{Df}{=} \begin{cases} f_\alpha \text{ on } \alpha \\ g_\beta \text{ on } \beta \end{cases} \quad \text{is also in } \mathscr{D}.$$

(ii) Non-null restriction: if $f_\gamma \in \mathscr{D}$ and $\delta \subset \gamma$ (and $\delta \in \mathscr{A}$ is not regarded as impossible by the subject), then the restriction of f_γ to δ is also in $\mathscr{D}$.

4 *Note added after receiving the Krantz and Luce reply* [6] *that appears later in this volume:* Krantz and Luce correctly point out that the 'atomic closure paradox' noted above (and further discussed below) vanishes if $\mathscr{D}$ contains 'lotteries' *only* (this solves the twin riddles: (1) when is the 'union' of two atomic alternatives itself atomic? and (2) how may an atomic alternative have consequences that are also atomic?). To preserve the 'atomic alternative only' view of the Luce–Krantz $\mathscr{D}$, then, it would seem that uncertainty must appear in 'objective' form only, as a well-specified and universally understood set of 'probability experiments'.

5 I have kept to the notational formats in refs. [4] and [7], though some of the symbols have been changed to avoid confusion between the primitives of the two systems. In identifying $\mathscr{A}$, I have reported the authors' phraseology without change.

The authors posit a preference relation $\succ$ on $\mathscr{D}$, and together with additional behavioral and structural axioms sufficient to their purpose, prove a representation theorem the conclusion of which reads: there exists a function $u: \mathscr{D} \rightarrow Re$ and a probability measure P on $\mathscr{A}$ such that

$$f_\alpha \succ g_\beta \quad \text{iff} \quad u(f_\alpha) > u(g_\beta), \tag{C2.1}$$

and

$$\text{if } \alpha \cap \beta = \varnothing, \text{ then} \\ u(f_\alpha \cup g_\beta) = P(\alpha \mid \alpha \cup \beta)u(f_\alpha) + P(\beta \mid \alpha \cup \beta)u(g_\beta). \tag{C2.2}$$

Of course (C2.1) is just the order-preserving property of u, the first requirement of any utility function. It is the decomposition expression (C2.2) that contains the essence of what is meant by conditional expected utility: it says that the utility assignment for the action $f_\alpha \cup g_\beta$ (which, once taken, guarantees an outcome in $\alpha \cup \beta$) is the weighted sum of utility assignments for f_α and g_β, the weights $P(\alpha \mid \alpha \cup \beta)$ and $P(\beta \mid \alpha \cup \beta)$ $(= 1 - P(\alpha \mid \alpha \cup \beta))$ reflecting the subject's personal probability as to which of α or β does in fact obtain (will in fact occur). The intended spirit here is, of course, that once he chooses $f_\alpha \cup g_\beta$ the decision maker has already exercised his 'conditioning influence' over nature; it then remains for God to have the final say (and for the subject to await its revelation). But now comes the conceptual difficulty. If $f_{\alpha \cup \beta} \in \mathscr{D}$, with $\alpha \cap \beta = \varnothing$, then axiom 1 (ii) requires that the restrictions to α and β, say f_α and f_β respectively, *also* be available for choice. And if, say, $f_\alpha \succ f_\beta$, the decision maker will simply choose f_α for sure; but then what was the meaning of $f_\alpha \cup f_\beta$ $(= f_{\alpha \cup \beta})$ in the first place, that is, what is the meaning of the 'probability' weight $P(\alpha \mid \alpha \cup \beta)$ in (C2.2)? On the other hand, if f_α and g_β are separately available for choice, with $\alpha \cap \beta = \varnothing$, then axiom 1 (i) requires that $f_\alpha \cup g_\beta$ is also available, but then what is it that *generates* the uncertainty as to which of α or β will obtain? The point is that the answer to 'which of f_α or g_β is (will be) operative in $f_\alpha \cup g_\beta$?' should either be under the control of the decision maker or not; the twin conditions (i) and (ii) of axiom 1 imply that we can have it both ways.

The Fishburn framework avoids this difficulty altogether, while preserving something of the spirit of the Luce–Krantz approach[6]. The

---

[6] As noted above, cf. ref. [1] for a theory that starts from a set of fully conditional primitives.

problem we have observed centers on a confusion over which (uncertain) environments the subject may *choose* to inhabit, and which circumstances are beyond his ability to influence. Fishburn returns to Savage (on this question) by insisting that what one shall mean by an *event* is something that is subject to the 'choice' of God alone.

Fishburn starts with a set of acts (or strategies) F which are presumed available to the decision maker, where the consequence of any given act f depends upon which state of nature s in a set S of mutually exclusive and collectively exhaustive states actually obtains. It is assumed that the 'true' state is unknown to the subject at the moment when he must exercise choice and, as noted above, the description of S is such that no act (once chosen) can affect whatever is the true state. It is important to note that although Fishburn starts with these Savage primitives, there is no 'sufficient richness' requirement in this theory for S, and only a modest[7] such requirement for F (i.e. we are essentially free to describe these sets according to the context of the decision making situation at hand). Fishburn further departs from Savage in denoting 'the consequence of act f when state s obtains' by the pair (f, s), rather than by the functional notation $f(s) \in \mathscr{C}$ (where in the Savage system, as noted earlier, the set of consequences $\mathscr{C}$ was primitive and F was the set of maps from S to $\mathscr{C}$)[8]. More generally, the set of *events* $\mathscr{E}'$ is a Boolean algebra of subsets of S with the empty set missing, and for $A \in \mathscr{E}'$, (f, A) is whatever may happen when f is chosen and it is presumed that *some* $s \in A$ obtains. Fishburn specifies a preference relation on $F \times \mathscr{E}' = \{(f, A) \mid f \in F,\ A \in \mathscr{E}'\}$, but takes great care to point out that such act–event pairs are not actually available as objects of choice (unless $A = S$, or unless A is regarded by the subject as virtually certain to obtain), since events (as conceived here) are not 'appropriable' on demand; the subject is simply asked to express his 'druthers' (to God, if to no one else) between a commitment to do act f under the supposition that some

---

[7] This is necessary in order to treat certain anomalies that might be loosely described as arising with 'model-theoretic measure zero'; a complete discussion is undertaken in ref. [1].

[8] At face, of course, this is no more than a different name for the same thing (even when $f \to (f, s)$ is regarded as a correspondence, as Fishburn provides for to accommodate 'residual uncertainty'), but the older notation has long been associated with a theory that does not treat 'consequences' which are characterizable only in relation to distinguished states.

$s \in A$ obtains and a commitment to do act g under the supposition that some $s \in B$ obtains[9].

Now several things should be noted. First, (f, A) pairs have something (but not all) of the character of both 'conditioning events' and 'conditional decisions', since they reflect the subject's partial influence over his environment through f. Indeed, if f is the act 'travel by car' and $S = A \cup \bar{A}$, where A is the event 'airplane crashes' and $\bar{A}$ its complement, then (f, S) does not contain the chance circumstance 'I die in an airplane crash'. Moreover, although we must give up on the idea that (f, A) can be appropriated by the subject on demand, we do allow (indeed require) him to consider the circumstance. And if A *should* in fact obtain (say as the first part of a sequential process) then (f, A) does, after all, have similar effect (from that point forward) to the corresponding 'conditional decision' of Luce and Krantz.

Second, the Fishburn counterpart to the decomposition expression (C2.2) is

$$\text{if } A \cap B = \varnothing, \text{ then } u(f, A \cup B) = P_{A \cup B}(A)u(f, A) + P_{A \cup B}(B)u(f, B). \quad \text{(C2.3)}$$

Here there is no question that the probability weights are associated with natural environmental uncertainty. To be sure, the Fishburn counterpart of LK axiom 1 (ii) must be relaxed in the sense noted above: in general, neither $(f, A \cup B)$ nor its restrictions (f, A) and (f, B) are actually available to the subject on demand. He is being asked to muse over 'what ifs', and it is just this price that buys the candle.

Third, a counterpart for LK axiom 1 (i) is required by Fishburn as well, through the introduction of extraneous lotteries over act–event pairs with the event held fixed. Recall that the interpretive difficulty for LK axiom 1 (i) had to do with the meaning of $f_\alpha \cup g_\beta$ when both f_α and g_β were separately available to the decision maker, i.e. what then was the nature of the uncertainty associated with $\alpha \cup \beta$? Luce and Krantz recognized this question, and proposed informally that $f_\alpha \cup g_\beta$ represent a laboratory experiment, in the form of a well-defined lottery with 'prizes' f_α and g_β. What Fishburn has done is to incorporate this idea formally by embedding $F \times \mathscr{E}'$ in $X \times \mathscr{E}' = \{(x, A) \mid x \in X, \ A \in \mathscr{E}'\}$, where X is the mixture set of simple lotteries with prizes in F. It fol-

9 This idea had already been employed by Bolker [2] in a more simplified decision-theoretic context.

lows at once that $X \times \mathscr{E}'$ can be written (and conceived) as the union

$$\bigcup_{A \in \mathscr{E}'} \{(x, A) \mid x \in X\}$$

of a family of mixture sets, each of which is X suppositioned on a fixed $A \in \mathscr{E}'$. That is, the 'prizes' of any lottery suppositioned on an event A are themselves lotteries (perhaps degenerate) suppositioned on the same A. Fishburn does not find it necessary (nor conceptually palatable) to consider lotteries of the form '(x, A) if heads, (y, B) if tails' when $A \neq B$; indeed, they are not elements of $X \times \mathscr{E}'$. I shall have more to say on this in the next section[10]. Of course the preference relation $\succ$ is now assumed primitive over $X \times \mathscr{E}'$, and is instrumental in relating attitudes toward *risk* (as realized by trials of a fair spinner) and subjective judgments on the 'distribution of uncertainty' over events in $\mathscr{E}'$.

C2.1.3. The Bolker objection and Supergenie

Axioms 1–3 in ref. [3] are those 'natural' extensions of the corresponding Herstein–Milnor axioms for ordered mixture sets which are appropriate to the present system. They are necessary as well as a subset of the sufficient conditions for Fishburn theorem 1 to hold; there is no hope for a suppositional expected utility theory without them. And once the conceptual framework provided by axiom 1 is accepted, axioms 2 and 3 seem as palatable as their simpler parents. The extensions chosen serve double duty: they Herstein–Milnor order every $\{(x, A) \mid x \in X\}$ separately, and also provide considerable cardinal linkage across elements of $\mathscr{E}'$ (cf. lemmas 1–5 in ref. [3]). Recall in this latter connection that 'lotteries' of the form $\frac{1}{2}(x, A) + \frac{1}{2}(y, B)$ are not elements of X when $A \neq B$. Fishburn declines the construction of a system that includes such things on the Bolker argument that 'this creates a direct conflict between the mixture (scaling) probabilities and the decision maker's beliefs about the relative likelihoods of A and B' (from p. 10 in ref. [3]). The idea is that our decision maker would have every right to challenge (place little faith in) an experimenter's ability to 'award' A or B on the toss of a coin. And yet the interesting mathematical fact here is:

It is possible to embed the Fishburn system in an axiomatic structure that encompasses all such 'forbidden' lotteries, and from which emerges a

[10] For now, compare footnote 2 and the paragraph to which it refers.

cardinal utility that agrees (after adjustment of zero and unit) with the Fishburn utility indicator on the Fishburn restriction; moreover, the (unique) Fishburn conditional probability measures P_A for $A \in \mathscr{E}'$ can in fact be computed directly from such lotteries.

To show this I will first describe the set of mind experiments that justifies the extended system on conceptual grounds. Thus let F and $\mathscr{E}'$ be as before, and suppose that you are owned by a Supergenie who can bring any event to pass with a snap of his fingers (which lends some force to the notion of ownership!). His particular sport is to place you in binary choice situations (from which you may not politely withdraw) between pairs of alternatives of the form $\lambda(f, A) + (1 - \lambda)(g, B)$, with the following understanding: you spin a fair pointer (the Sg agrees not to interfere in its outcome) and if, say, that predetermined arc associated with the probability number λ obtains, the Sg 'wills' A and requires your commitment to act f. The Sg makes no value judgments on your tastes, and is entirely ethical with respect to the rules of his game. With this procedural understanding between you and the Sg, the Bolker objection is obviated, and it is to your advantage to pairwise order all such alternatives[11].

Mathematically, this amounts to assuming a preference relation $\succ$ on $\mathscr{M}(F \times \mathscr{E}')$, the mixture set of lotteries with prizes in $F \times \mathscr{E}'$. Note that this set contains the Fishburn primitive

$$\mathscr{M}(F) \times \mathscr{E}' = \bigcup_{A \in \mathscr{E}'} \mathscr{M}(F) \times \{A\}$$

as a special subset (where I have written $\mathscr{M}(F)$ for X). Now assume the usual Herstein–Milnor axioms for $\succ$ on $\mathscr{M}(F \times \mathscr{E}')$; these reduce to Fishburn's axioms 1–3 for the restriction of $\succ$ to $\mathscr{M}(F) \times \mathscr{E}'$. Suppose moreover that Fishburn axioms 4–6 also hold for the restriction of $\succ$ to $\mathscr{M}(F) \times \mathscr{E}'$, so that Fishburn Theorem 1 applies. It follows from the uniqueness part of this theorem (more precisely, from lemma 6 of ref. [3]) that the Fishburn system and Herstein–Milnor supersystem have utility indicators that agree (after adjustment of 0 and unit) on $\mathscr{M}(F) \times \mathscr{E}'$. Thus u is extended to all of $\mathscr{M}(F \times \mathscr{E}')$, and is linear on

---

[11] Of course this little fantasy need not have been spun as a drama (against a backdrop having game-theoretic overtones); Supergenie is nothing more than our decision maker's conceptual mechanism for expressing his 'druthers'.

that set (i.e. $u(\lambda\xi + (1 - \lambda)\eta) = \lambda u(\xi) + (1 - \lambda)u(\eta)$ for $(\lambda, \xi, \eta) \in [0,1] \times \mathcal{M}(F \times \mathcal{E}') \times \mathcal{M}(F \times \mathcal{E}')$). To see how the probability assignments[12] of theorem 1 are generated by an elementary canonical mind-experiment involving choice between Supergenie lotteries in $\mathcal{M}(F \times \mathcal{E}')$, let A, $B \in \mathcal{E}'$ with $A \cap B = \emptyset$, and suppose that $x \in X$ is such that

$$(x, A) \succ (x, B). \tag{C2.4}$$

This construction is guaranteed by part of axiom 6. By theorem 1, there exist unique non-negative numbers $P_{A \cup B}(A)$ and $P_{A \cup B}(B)$ summing to 1 such that

$$u(x, A \cup B) = P_{A \cup B}(A)u(x, A) + P_{A \cup B}(B)u(x, B). \tag{C2.5}$$

On the other hand, using construction (C2.4), axiom 4 and the Archimedean property of a Herstein–Milnor order (on $\mathcal{M}(F \times \mathcal{E}')$), there exists a unique $\lambda \in [0,1]$ such that

$$(x, A \cup B) \sim \lambda(x, A) + (1 - \lambda)(x, B). \tag{C.26}$$

By the order preserving and linearity property of the extended u this implies, together with (C2.4) and (C2.5), that $\lambda = P_{A \cup B}(A)$. In other words, given the commitment to do x in any case, the objective probability for obtaining A that I would require of my Supergenie in the lottery on the right hand side of (C2.6), in order to be just indifferent to taking my 'natural' chances on A under the supposition that $A \cup B$ obtains, is just precisely the $P_{A \cup B}(A)$ that emerges from the Fishburn theory. Moreover, the argument did not depend on x (except in so far as $(x, A) \sim (x, B)$ is false; otherwise any mix would do and we could not then infer $P_{A \cup B}(A)$ by this procedure).

References

[1] Balch, M. and Fishburn, P. C. Subjective expected utility for conditional primitives. In this volume, pp. 57–69.

[2] Bolker, E. D. A simultaneous axiomatization of utility and subjective probability. *Philosophy of Science*, **34** (1967), 333–340.

[12] As noted earlier, the (intricate) Fishburn derivation of these probabilities provides little insight into their meaning though, given his commitment to economize on requirements of a 'perception-theoretic' nature, this is hardly surprising. But the point of this present section is that, from the purely formal view, there is no economy whatever; i.e. the subject of the Fishburn theory behaves precisely *as if* Supergenie were alive and well in his mind. And as we shall see below, the heuristic appeal of the Supergenie mechanism for generating 'degree of belief' judgments is immediate.

[3] Fishburn, P. C. A mixture-set axiomatization of conditional subjective expected utility. *Econometrica*, **41** (1973), 1–25.
[4] Fishburn, P. C. On the foundations of decision making under uncertainty. In this volume, pp. 25–44.
[5] Herstein, I. N. and Milnor, J. An axiomatic approach to measurable utility. *Econometrica*, **21** (1953), 291–297.
[6] Krantz, D. H. and Luce, R. D. The interpretation of conditional expected-utility theories. In this volume, pp. 70–73.
[7] Luce, R. D. and Krantz, D. H. Conditional expected utility. *Econometrica*, **39** (1971), 253–271.
[8] Savage, L. J. *The Foundations of Statistics*. Wiley: New York (1954).

CHAPTER 3

SUBJECTIVE EXPECTED UTILITY FOR CONDITIONAL PRIMITIVES

Michael Balch and Peter C. Fishburn

3.1. Introduction

The classical Savage paradigm [12] for decision making under uncertainty requires that 'states-of-the-world' be so conceived as to remain uninfluenced by any act the decision maker may choose to implement. In this paper we develop an expected utility theory which relaxes this conceptual requirement by casting primitives in conditional form. Although the need for a conditional theory has been recognized for some time, the first general treatment along these lines has been given only recently by Luce and Krantz [9]. Unfortunately, this latter theory is haunted by a structural requirement that leads to an interpretive paradox for its primitives. It is partly in response to this difficulty – a discussion of which appears in ref. [1] – that our present effort has emerged. Our model descends directly from the one developed in ref. [5] (see also ref. [6]) which, though successful in avoiding conceptual paradox, did not itself start from a set of conditional primitives.

We briefly recall the case for a conditional theory. For the decision maker of the Savage theory, the consequence of implementing an act depends upon whatever is (happens to be) the true 'state-of-the-world', but this true state is not known with certainty at the moment when choice is exercised. The subject is to envisage a mutually exclusive and collectively exhaustive set S of *candidate* states-of-the-world; each of these is at least logically possible, and of course one of them in fact obtains (is true). Now the Savage theory requires that this set S *must in principle* be so described by the subject (according to his own perceptions) that *if* an arbitrary $s \in S$ were indeed the true state, then no act which he may choose to implement can alter that fact[1]. The practical difficulty with this requirement is that there are typical decision contexts for which the formulation of an appropriate Savage set S is intrinsically not obvious and, indeed, exceedingly complex. For an elementary example, consider a salesman who must decide on advertising expenditure for the coming marketing period. The determinants of demand are of course impossibly intricate, but the decision maker believes that the sales volume $s \in [0, \infty)$ which is realized by the end of the period will indeed have been somehow influenced by the advertising level $x \in [0, \infty)$ that was selected at the start of the period. An appropriate set of Savage states for this problem is by no means transparent.

Instead, we shall associate to each act its own 'natural' set of act-conditional outcomes[2], and develop an expected utility theory for these conditional primitives; i.e. the theory will not rely on a set of Savage states. The advantage of this approach is that it is usually obvious how to describe these sets from the context of the particular decision making situation at hand[3]. And since it is always possible to formally generate a Savage set from a full collection of act-conditional outcome sets (as illustrated for the advertising/sales example in footnote 3), our shift in framework provides a gain for conceptual simplicity and analytic

---

[1] Thus Professor Savage, by way of indicating possible directions for the extension of his theory, asks [13, p. 307]: 'Is it good, or even possible, to insist, as this preference theory does, on a usage in which (a) acts are without influence on events, and (b) events without influence on well-being?' In (a), Savage is calling for a conditional theory; in (b), for a theory that does not rely on 'conceiving' acts which award *any* given consequence for *any* given state (the point here is that, when 'consequences' are holistically interpreted and thus inextricably related to the state that obtains (this is, strictly speaking, always the case), then it defies logic to 'conceive' an act that, say, awards for state s a consequence c which depends holistically on state t ($\neq s$)). The present theory is delimited by neither (a) nor (b).

[2] An earlier version of this idea appears in ref. [3], pp. 21–30, 36–39.

[3] Thus it is natural for the advertising/sales example above to take $S_x = [0, \infty)$ as the x-conditional outcome set for every $x \in X = [0, \infty)$. The fact that these S_x are identical (for this very simple example) is merely incidental; $[0, \infty)$ cannot serve as a Savage set of states-of-the-world (i.e. the salesman of this example perceives otherwise). Now we may formally construct a Savage set Σ as the set of mappings

$$\Sigma = \{\sigma : X \to \bigcup_{x \in X} S_x \,|\, \sigma(x) \in S_x\}.$$

Thus a 'state' $\sigma \in \Sigma$ consists in naming a sales volume $\sigma(x) \in S_x$ for *every* advertising level $x \in X$. It is clear that the function set Σ has the Savage property, but note its formidable size and consequent conceptual complexity. The point is that Σ has no role to play in the conditional theory.

tractability with no attendant loss of generality. The more unwieldy set of Savage states appropriate to any given situation may rest comfortably behind the scenes.

Before presenting the details of our theory, some further remarks are in order. It is the case for all theories of expected utility known to us that, in order to guarantee existence of and uniqueness for the desired representation, there must be sufficient structural richness within the system. This will not surprise anyone who has reflected upon the truly ambitious nature of that representation, especially when the theory also seeks to account for 'natural' uncertainty (those situations which are not already presented to the decision maker in 'lottery' form and, indeed, for which 'statistical' (or 'risk') interpretations simply do not apply)[4]. Now with respect to the systemic placement of requisite structural richness, the first consideration must be for the conceptual integrity of the system; otherwise, the theory may be incomprehensible at its very foundation. We achieve this desideratum and satisfy the structural imperative noted above by asking the subject of our theory to express his 'druthers' as between various conceivable (as *he* understands the term) resolutions of natural uncertainty. That is, he is to express preference between suppositioned circumstances of the form 'act f is implemented and the f-conditional event A obtains'. Indeed, we shall ask him to consider 'lotteries' which offer such suppositioned circumstances as 'prizes'; for this purpose the subject is to imagine a Supergenie prizemaster whose authority over nature is complete (we shall give full interpretive details below). It is important to understand that while we do require the subject thus to exercise his imagination (and in a normatively consistent way), we shall never present for his consideration a suppositioned circumstance that *he* does not comprehend as logically possible[5].

A final remark. Our theory provides for a *direct conceptual link* between objective probabilities (which associate to a partitioning of the circumference of a fair spinner, say) and subjective probabilities (which

---

4 In the present theory, for example, we shall elicit a probability measure for each act (on subsets of its outcome set) which may be interpreted as reflecting the subject's personal appraisal of relative likelihoods (with respect to the question: which of two events (conditioned by the same act) is 'more likely' to contain the true state?).

5 We have already remarked in footnote 1 on a major criticism of the Savage theory in this vein. See also discussions in refs. [1], [5] and [9].

may be interpreted as reflecting 'degree of belief', as noted earlier). Indeed, we employ the former in a canonical mind experiment to scale the latter directly[6]. This experiment, which is elementary in conception and of evident heuristic appeal (once the notion of a Supergenie lottery is understood), will automatically reflect whatever 'prize-distortion' or 'superstition' effect is *intrinsically operative* for our subject. That is, the subject is allowed to feel that 'something as good as "this" must (could never) happen to me'. This sort of thing is usually ruled out in other theories of subjective expected utility (in the Savage theory, for example) on the grounds that whether or not a given event obtains has nothing to do with an arbitrarily superimposed prize structure. Of course this is true. But subjective *feelings* concerning relative likelihoods are not God-given; they are, after all, subjective, and it is these feelings which our theory elicits.

3.2. Theory Core

We have in mind a given decision making situation under uncertainty. The only way in which our subject can attempt to influence his environment is by implementing an act. In general, he is uncertain as to what will be the outcome of any given act at the moment when choice is exercised, though he does have a subjective picture of what *might* happen.

Thus let the primitive $\mathscr{F}$ be the set of acts presumed available for choice, and for each $f \in \mathscr{F}$, let S_f be a mutually exclusive and collectively exhaustive set of logically possible 'states-of-the-world-as-conditioned-by-f'. The subject understands that precisely one of these *f-conditional states* obtains (is true, will be realized when f is implemented), though he cannot (in general) *a priori* identify the true f-conditional state with certainty. More generally, let $\mathscr{E}_f$ be a Boolean algebra of subsets of S_f; for convenience we put $\mathscr{E}'_f = \mathscr{E}_f - \{\varnothing\}$. An element $A \in \mathscr{E}'_f$ is called

[6] This idea is employed in ref. [11] for a decision context which is circumscribed, in part, by the full quotation in footnote 1. The authors stop just short of 'proclaiming' Supergenie (a conceptual imperative, in general, for this kind of mind experiment), preferring rather to illustrate by specific examples in which the role played by Supergenie is easily anthropomorphized (for example by the 'sporting' President of the XYZ Company, pp. 8–9 in chapter 2).

an *f-conditional event*, and is said to obtain if some f-conditional state in A obtains. We treat the collection $\{\mathscr{E}_f\}_{f\in\mathscr{F}}$ of act-conditional Boolean algebras (therefore also $\{S_f\}_{f\in\mathscr{F}}$) as primitive.

The holistic consequences of our theory are act–event pairs (f, A), where for any such pair it is always understood that $A \in \mathscr{E}'_f$. The subject is to think of (f, A) as that circumstance, perhaps uncertain still, which is conditioned by choice of f and suppositioned on the occurrence of A. That is, he is to imagine what might happen if indeed the true f-conditional state were in A. In particular, $(f, \{s\})$ is whatever may happen if f is implemented and the f-conditional state s obtains (the 'atomic' state s need not describe a condition of certainty), while (f, S_f) is whatever may happen when f is implemented. We call (f, A) a *suppositioned circumstance*, and let $\mathscr{C}$ denote the set $\{(f, A)\,|\, f\in\mathscr{F},\ A\in\mathscr{E}'_f\}$ of all such pairs.

Let $\mathscr{M}(\mathscr{C})$ be the mixture set[7] of all simple probability measures on $2^{\mathscr{C}}$. A generic element in $\mathscr{M}(\mathscr{C})$ may be written as a convex combination

$$\sum_{i=1}^{n} \alpha_i(f_i, A_i)$$

of elements (f_i, A_i) in $\mathscr{C}$($\alpha_i \geq 0$ for $i = 1, 2, \ldots, n$, and $\sum_{i=1}^{n} \alpha_i = 1$).[8] Then our final primitive is a preference relation $\succ$ on $\mathscr{M}(\mathscr{C})$, with the following operational meaning: the subject is to (a) interpret every element of $\mathscr{M}(\mathscr{C})$ as a 'lottery' (perhaps degenerate) with 'prizes' in $\mathscr{C}$, and then (b) pairwise order such lotteries as if they were actually available for choice. Of course most[9] of these lotteries must be viewed as hy-

---

7 The *power set* $2^{\mathscr{C}}$ is the set of all subsets of $\mathscr{C}$. A probability measure $x \in \mathscr{M}(\mathscr{C})$ is *simple* if $x(C) = 1$ for some finite $C \in 2^{\mathscr{C}}$. $\mathscr{M}(\mathscr{C})$ is closed under a natural *mixture* operation for probability measures: if $x, y \in \mathscr{M}(\mathscr{C})$ and $\lambda \in [0, 1]$, then $\lambda x + (1 - \lambda)y$ is also in $\mathscr{M}(\mathscr{C})$, where

$$(\lambda x + (1 - \lambda)y)(C) \underset{Df}{=} \lambda x(C) + (1 - \lambda)y(C)$$

for all $C \in 2^{\mathscr{C}}$. For more on mixture sets see, for example, ref. [4], pp. 110, 111.

8 Strictly speaking, a generic $x \in \mathscr{M}(\mathscr{C})$ should be written

$$x = \sum_{i=1}^{n} \alpha_i \hat{x}_{(f_i, A_i)}$$

where, for each $c \in \mathscr{C}$, $\hat{x}_c$ is that degenerate probability measure for which $\hat{x}_c(\{c\}) = 1$. It is standard practice to relax this strict pedagogic form in the interest of notational economy.

9 The sub-mixture set of lotteries which are indeed real-world available is $\mathscr{M}(\{(f, S_f)|f\in\mathscr{F}\})$.

pothetical constructs, inasmuch as $\mathscr{C}$ is no 'ordinary' prize set to begin with: no mortal can guarantee 'delivery' of the generic suppositioned circumstance (f, A) (Everyman's influence over nature extending no further than his ability to implement f). But we shall suppose that our subject has no difficulty in imagining his real world to be embedded in one which is also inhabited by a Supergenie. A Supergenie, by definition, can bring any logically possible act-conditional event to pass with a snap of his fingers. In particular, the lottery $\alpha(f, A) + (1 - \alpha)(g, B)$, if 'offered' by the Supergenie and 'chosen' by the subject, is to be understood according to the following procedure: the subject determines the 'prize' by means of an ordinary probability experiment, and if the prize is (f, A), say, then the subject implements act f and the Supergenie 'wills' event A (that is, guarantees that the true f-conditional state is indeed *somewhere* in A). We need hardly emphasize that for this interpretation of the primitive $\succ$ on $\mathscr{M}(\mathscr{C})$, game-theoretic considerations (that might be thought to obtain 'between' the subject and Supergenie) such as 'cheating', 'reneging', 'outguessing', and so forth are simply not relevant here; Supergenie, after all, lives in the mind of the subject and is conjured up *by him* for the purpose of expressing his 'druthers'.

Of course the preference order $\succ$ on $\mathscr{M}(\mathscr{C})$ is to satisfy requirements of the usual type for normative consistency. We thus assume throughout that the primitives $(\mathscr{F}, \{\mathscr{E}_f\}_{f \in \mathscr{F}}, \succ)$ satisfy

$$\succ \text{ is a weak order[10] on the mixture set } \mathscr{M}(\mathscr{C}). \tag{A3.1}$$

$$\text{For all } x, y, z \in \mathscr{M}(\mathscr{C}), \text{ the sets } \{\alpha \in [0,1] \mid \alpha x + (1 - \alpha)y \succsim z\} \text{ and } \{\alpha \in [0,1] \mid \alpha x + (1 - \alpha)y \precsim z\} \text{ are closed.} \tag{A3.2}$$

$$\text{For all } x, y, z \in \mathscr{M}(\mathscr{C}) \text{ and for all } \alpha \in [0,1],\quad x \sim y \Rightarrow \alpha x + (1 - \alpha)z \sim \alpha y + (1 - \alpha)z. \tag{A3.3}$$

$$\text{For every } f \in \mathscr{F} \text{ and } A, B \in \mathscr{E}'_f \text{ with } A \cap B = \varnothing,\quad (f, A) \succsim (f, B) \Rightarrow (f, A) \succsim (f, A \cup B) \succsim (f, B). \tag{A3.4}$$

---

[10] A binary relation $\succ$ on a set X is a *weak order* (in the strict sense) if it is asymmetric ($x \succ y \Rightarrow$ not $y \succ x$) and negatively transitive (not $x \succ y$ and not $y \succ z \Rightarrow$ not $x \succ z$); transitivity ($x \succ y$ and $y \succ z \Rightarrow x \succ z$) follows. *Indifference* ($\sim$) and *preference-or-indifference* ($\succsim$) are defined

$$x \sim y \quad \text{iff} \quad \text{not } x \succ y \text{ and not } y \succ x,$$
$$x \succsim y \quad \text{iff} \quad x \succ y \text{ or } x \sim y.$$

When $\succ$ is a weak order, $\sim$ is an equivalence (reflexive, symmetric, transitive) and $\succsim$ is transitive and complete ($x \succsim y$ or $y \succsim x$ for all $x, y \in X$).

A3.1–A3.3 are the well-known Herstein–Milnor axioms [7] for an ordered mixture set, and A3.4 is that version of the Bolker [2]–Fishburn [5] averaging condition which is appropriate for the present conditional framework. The Herstein–Milnor axioms have been extensively discussed in the literature[11] for 'ordinary' prize sets $\mathscr{C}$ from the viewpoint of their model-theoretic efficacy; this discussion is no way altered for the present paradigm, given our conceptual understanding for $\succ$ on $\mathscr{M}(\mathscr{C})$. With respect to their structural role for the present theory, the Herstein–Milnor axioms contribute in two essential ways. The first (lemma 3.1 below) provides the utility function u: $\mathscr{M}(\mathscr{C}) \rightarrow Re$ of the theory; u is order-preserving, linear, and unique up to positive affine transformation. The second (lemma 3.2 below) is an Archimedean property which, together with A3.4, is central for the determination of subjective probabilities (as will be evident from lemma 3.3 below).

LEMMA 3.1. (*Herstein–Milnor*). *Let* $\succ$ *on* $\mathscr{M}(\mathscr{C})$ *satisfy* A3.1–A3.3. *Then there is a function* $u: \mathscr{M}(\mathscr{C}) \rightarrow Re$ *such that for all* $x, y \in \mathscr{M}(\mathscr{C})$ *and* $\alpha \in [0,1]$

(i) $x \succ y$ *iff* $u(x) > u(y)$, *and*

(ii) $u(\alpha x + (1-\alpha)y) = \alpha u(x) + (1-\alpha)u(y)$.

Moreover, if v: $\mathscr{M}(\mathscr{C}) \rightarrow Re$ *also satisfies* (i) *and* (ii), *then* $v = au + b$ *for some* $a, b \in Re$ *with* $a > 0$.

Lemma 3.1 is a restatement of theorems 7 and 8 in ref. [7].

LEMMA 3.2. (*Herstein–Milnor*). *Let* $\succ$ *on* $\mathscr{M}(\mathscr{C})$ *satisfy* A3.1–A3.3, *and consider* $x, y, z \in \mathscr{M}(\mathscr{C})$ *such that* $x \succsim y \succsim z$. *If* $x \succ z$, *then there exists a unique* $\alpha \in [0,1]$ *such that* $y \sim \alpha x + (1-\alpha)z$. *If* $x \sim z$, *then* $y \sim \beta x + (1-\beta)z$ *for all* $\beta \in [0,1]$.

Lemma 3.2 is a restatement of theorems 1 and 6 in ref. [7]. The intuitive appeal of the Archimedean property expressed in this lemma is obvious. With $x, y, z \in \mathscr{M}(\mathscr{C})$ appropriately specialized, it provides the canonical dichotomous choice algorithm by which our paradigm generates subjective probability numbers for act-conditional events. We introduce some convenient terminology.

---

[11] See refs. [7] and [10]. A related set of axioms is discussed in ref. [8] and in ref. [4], pp. 107–110. These are all modifications of the basic von Neumann–Morgenstern [14] system.

If A, $B \in \mathscr{E}'_f$ and $A \cap B = \varnothing$, then the two-element set $\{A, B\} \subseteq \mathscr{E}'_f$ is called a *dichotomy* of events for f (note that a dichotomy, in this usage, is not necessarily a partition of S_f). If $\{A, B\} \subseteq \mathscr{E}'_f$ is a dichotomy, and $(f, A) \sim (f, B)$ is false, then f is said to *bifurcate* the dichotomy $\{A, B\}$.

LEMMA 3.3. *Let* $\succ$ *on* $\mathscr{M}(\mathscr{C})$ *satisfy* A3.1–A3.4, *and let* $\{A, B\} \subseteq \mathscr{E}'_f$ *be a dichotomy for f. If f bifurcates* $\{A, B\}$, *then there exists a unique* $\alpha \in [0,1]$ *such that*

$$(f, A \cup B) \sim \alpha(f, A) + (1 - \alpha)(f, B).$$

If f does not bifurcate $\{A, B\}$, *then* $(f, A \cup B) \sim \beta(f, A) + (1 - \beta)(f, B)$ *for every* $\beta \in [0,1]$.

PROOF. Suppose without loss of generality that $(f, A) \succsim (f, B)$. Then $(f, A) \succsim (f, A \cup B) \succsim (f, B)$ by A3.4, and lemma 3.3 follows from lemma 3.2 by identifying (cf. footnote 8) (f, A), $(f, A \cup B)$, and (f, B) with x, y and z.

For the case that f bifurcates the dichotomy $\{A, B\}$, the interpretation of the unique probability number α in lemma 3.3 is clear: given the subject's commitment to implement f in any case, α is that (objective) probability for obtaining A that he would truly require of Supergenie (if there really were such a fellow) in the hypothetical lottery $\alpha(f, A) + (1 - \alpha)(f, B)$, in order to be just indifferent to taking what he perceives to be his 'natural' chances on the occurrence of A under the supposition that $A \cup B$ obtains. We give this uniquely determined α the name $P^f_{A \cup B}(A)$ (similarly, $P^f_{A \cup B}(B) = 1 - \alpha$). It then follows at once from

$$(f, A \cup B) \sim P^f_{A \cup B}(A)(f, A) + P^f_{A \cup B}(B)(f, B), \tag{3.1}$$

and the order-preserving and linearity properties of $u \colon \mathscr{M}(\mathscr{C}) \to Re$, that

$$u(f, A \cup B) = P^f_{A \cup B}(A)u(f, A) + P^f_{A \cup B}(B)u(f, B). \tag{3.2}$$

This is the fundamental decomposition formula for subjective expected utility. It expresses the utility assignment for the suppositioned circumstance $(f, A \cup B)$ as a convex combination of the utility assignments for the 'component' suppositioned circumstances (f, A) and (f, B), the weight $P^f_{A \cup B}(A)$ reflecting subjective degree of belief that the true f-conditional state is to be found in A under the supposition that it lives already in $A \cup B$.

The core of our theory is now in hand, though it does remain to tie up some loose ends. To begin with, we must consider the indeterminacy that arises when f fails to bifurcate a dichotomy $\{A, B\}$. For in this case, the canonical experiment of lemma 3.3 loses the power for discrimination by which $P^f_{A \cup B}(A)$ is otherwise uniquely determined. In the next section we shall reduce this 'knife-edge' anomaly in a manner that 'faithfully' reflects the subject's underlying appraisal of relative likelihoods. We then go on to establish the obvious generalization of (3.2) for multichotomies of f-conditional events; in the process, we show that the generic set function P^f_D: $\{A \cap D \mid A \in \mathscr{E}_f\} \to Re$ is indeed a probability measure.

3.3. Theory Closure

In order to treat the anomalous 'degree of belief' indeterminacy that arises for unbifurcated dichotomies, we introduce the notion of *Savage equivalence.* We shall define this below (in terms of primitives already introduced) as an equivalence relation $*$ on $\mathscr{F}$ in such a way that those acts which belong to a given $*$-equivalence class (a) share a common event algebra, and (b) are indistinguishable with respect to the probability numbers they may (separately) uniquely determine (according to the canonical experiment (3.1)) for *any* dichotomy in that algebra. Of course it may happen that a given act does not bifurcate a given dichotomy; indeed, the idea is then to generate the appropriate probability numbers for this case by means of some (any) Savage-equivalent act which *does* bifurcate that dichotomy. Finally, this aspect of the model is closed by assuming (A3.5 below) that every Savage class in $\mathscr{F}/*$ is sufficiently 'bifurcation rich' in the sense just described[12].

---

[12] A3.5 and A3.6 (to follow) are conditional (and somewhat weakened) versions of comparable robustness conditions developed for the 'uniformly Savage' framework of ref. [5].

This structural robustness, which we formally require to close the model, may (if necessary) be informally conceived in a way that does not burden our understanding for $\mathscr{F}$ as the set of real-world acts presumed available for choice. Thus we may have the following kind of subject-implemented and 'Savage-invariant' generating device in mind: for $A \in \mathscr{E}_f$, imagine an act f_A which is precisely the same as f except that the subject himself is to free-dispose of (say) a ten dollar bill *if A should in fact obtain.* To incorporate this in a formal way would involve axiomatizing a free-disposal commodity

To define $*$ on $\mathscr{F}$ we first construct an antecedent binary relation $\#$ on $\mathscr{F}$ according to

For all $f, g \in \mathscr{F}$, $f \# g$ iff

(a) $\mathscr{E}_f = \mathscr{E}_g$,

(b′) if $\{A, B\}$ is a dichotomy in $\mathscr{E}'_f$,
if f bifurcates $\{A, B\}$, and
if $(f, A \cup B) \sim \alpha(f, A) + (1 - \alpha)(f, B)$,
then $(g, A \cup B) \sim \alpha(g, A) + (1 - \alpha)(g, B)$,

(b″) same as (b′) with f, g interchanged.

The relation $\#$ on $\mathscr{F}$ is reflexive and symmetric, but not necessarily transitive[13]. Then in terms of this relation we define $*$ on $\mathscr{F}$ according to

For all $f, g \in \mathscr{F}$, $f * g$ iff

$$f \# h \Leftrightarrow g \# h, \text{ for all } h \in \mathscr{F}.$$

It is easily verified that this relation is indeed an equivalence; we denote the generic Savage class in $\mathscr{F}/*$ by $[f]$, and the common class-conditional event algebra for $[f]$ by $\mathscr{E}_{[f]}$. As anticipated above, we assume

For every $f \in \mathscr{F}$ *and dichotomy* $\{A, B\} \subseteq \mathscr{E}'_{[f]}$, *there is some* $g \in [f]$ *which bifurcates* $\{A, B\}$. (A3.5)

Then for the case that f itself does not bifurcate a dichotomy $\{A, B\}$, we just put

$$P^f_{A \cup B}(A) \underset{Df}{=} P^g_{A \cup B}(A)$$

for the $g \in [f]$ asserted by A3.5. That $P^f_{A \cup B}(A)$ is thus well defined (does not depend on choice of $g \in [f]$) follows from the fact that $*$ on $\mathscr{F}$ is an equivalence. We accordingly relabel $P^f_{A \cup B}(A)$ as $P^{[f]}_{A \cup B}(A)$. Thus for *any* dichotomy $\{A, B\} \subseteq \mathscr{E}'_{[f]}$ we have

$$(f, A \cup B) \sim P^{[f]}_{A \cup B}(A)(f, A) + P^{[f]}_{A \cup B}(B)(f, B) \tag{3.3}$$

(say money) of which 'more is preferred to less', and an assumption that, indeed, a 'prize-distortion' or 'superstition' effect (on subjective probabilities) does not obtain for 'small' free-disposal perturbations of the type just illustrated. But from the structural point of view, the model is complete on the basis of A3.1–A3.6.

[13] Suppose, for example, that $f \# h$ and $h \# g$ and that $(h, A) \sim (h, B)$ for some dichotomy $\{A, B\}$; we may still have uniquely determined $P^f_{A \cup B}(A)$ and $P^g_{A \cup B}(A)$ which do not agree.

and

$$u(f, A \cup B) = P^{[f]}_{A \cup B}(A)u(f, A) + P^{[f]}_{A \cup B}(B)u(f, B) \tag{3.4}$$

for uniquely determined $P^{[f]}_{A \cup B}(A)$.

More generally, let $\{D_1, D_2, \ldots, D_n\}$ be an n-chotomy in $\mathscr{E}'_{[f]}$; that is, $D_i \in \mathscr{E}'_{[f]}$ for $i = 1, 2, \ldots, n$ and $D_i \cap D_j = \varnothing$ for $i \neq j$. Let $\mathscr{E}(\{D_i\})$ denote the algebra generated by $\{D_1, D_2, \ldots, D_n\}$, put $D = \bigcup_{i=1}^{n} D_i$, and define $P^{[f]}_D(D) = 1$, $P^{[f]}_D(\varnothing) = 0$. To complete our theory we must extend (3.4) to

$$u(f, D) = \sum_{i=1}^{n} P^{[f]}_D(D_i)u(f, D_i), \tag{3.5}$$

and show that the set function $P^{[f]}_D$: $\mathscr{E}(\{D_i\}) \rightarrow Re$ is finitely additive, and therefore a probability measure.

It suffices to consider[14] the case $n = 3$. Thus let $\{A, B, C\}$ be a trichotomy in $\mathscr{E}'_{[f]}$ and, for arbitrary $g \in [f]$, let $\alpha, \beta, \ldots, \nu$ be the unique probability numbers for which

$$\begin{aligned}
u(g, A \cup B \cup C) &= \alpha u(g, A \cup B) + (1 - \alpha)u(g, C),\\
u(g, A \cup B) &= \beta u(g, A) + (1 - \beta)u(g, B),\\
u(g, A \cup B \cup C) &= \gamma u(g, A) + (1 - \gamma)u(g, B \cup C),\\
u(g, B \cup C) &= \delta u(g, B) + (1 - \delta)u(g, C),\\
u(g, A \cup B \cup C) &= \mu u(g, B) + (1 - \mu)u(g, A \cup C),\\
u(g, A \cup C) &= \nu u(g, A) + (1 - \nu)u(g, C).
\end{aligned}$$

Telescoping these by successive pairs we have

$$\begin{aligned}
u(g, A \cup B \cup C) &= \alpha\beta u(g, A) + \alpha(1 - \beta)u(g, B) + (1 - \alpha)u(g, C),\\
u(g, A \cup B \cup C) &= \gamma u(g, A) + \delta(1 - \gamma)u(g, B) + (1 - \gamma)(1 - \delta)u(g, C),\\
u(g, A \cup B \cup C) &= \nu(1 - \mu)u(g, A) + \mu u(g, B) + (1 - \mu)(1 - \nu)u(g, C).
\end{aligned} \tag{3.6}$$

Now suppose for the moment that these decompositions for $u(g, A \cup B \cup C)$ are in fact identical; we shall guarantee this in A3.6 below. Then, in particular, we have $\gamma = \alpha\beta$ and $\mu = \alpha(1 - \beta)$. Together, these

(a) give $\alpha = \gamma + \mu$, or
$P^{[f]}_{A \cup B \cup C}(A \cup B) = P^{[f]}_{A \cup B \cup C}(A) + P^{[f]}_{A \cup B \cup C}(B)$,
which proves *additivity* for $P^{[f]}_{A \cup B \cup C}$;

[14] The development to follow is that modification of the one given in ref. [5], pp. 39–41, which is appropriate for the present theory; we include it here for completeness.

(b) translate directly to the *chain rule*

$$P^{[f]}_{A\cup B\cup C}(A) = P^{[f]}_{A\cup B\cup C}(A \cup B)P^{[f]}_{A\cup B}(A),$$

$$P^{[f]}_{A\cup B\cup C}(B) = P^{[f]}_{A\cup B\cup C}(A \cup B)P^{[f]}_{A\cup B\cup C}(B); \quad \text{and}$$

(c) imply the generalization of (3.4) to

$$u(g, A \cup B \cup C) = P^{[f]}_{A\cup B\cup C}(A)u(g, A) + P^{[f]}_{A\cup B\cup C}(B)u(g, B) + P^{[f]}_{A\cup B\cup C}(C)u(g, C) \qquad \text{for all } g \in [f].$$

On the other hand, if the decompositions in (3.6) were *not* identical, we should have some non-zero triple of real numbers (ξ, η, ζ) for which

$$\xi + \eta + \zeta = 0, \text{ and}$$
$$\xi u(g, A) + \eta u(g, B) + \zeta u(g, C) = 0 \qquad \text{for all } g \in [f].$$

We therefore preclude this by positing

For every $f \in \mathscr{F}$ *and trichotomy* $\{A, B, C\} \subseteq \mathscr{E}'_{[f]}$, *there exist* $h, k \in [f]$ *such that the set of* 3-*tuples*

$$\{(1, 1, 1), (u(h, A), u(h, B), u(h, C)), (u(k, A), u(k, B), u(k, C))\}$$

is linearly independent. (A3.6)

Of course this structural condition is invariant under positive affine transformation of u. We have stated A3.6 in terms of derived rather than primitive concepts (we need not have); by so doing we realize a handsome notational saving, and perhaps communicate more effectively the idea that A3.6 simply performs the same kind of yeoman service as did A3.5 (by sidestepping yet another possible 'knife-edge' anomaly). For an informal way to achieve A3.6, recall the discussion in footnote 12. We summarize:

THEOREM. *Let* $(\mathscr{F}, \{\mathscr{E}_f\}_{f\in\mathscr{F}}, \succ)$ *satisfy* A3.1–A3.6. *Then there is a function* u: $\mathscr{M}(\mathscr{C}) \to Re$ *and a finitely additive probability measure* $P^{[f]}_D$: $\{E \cap D \mid E \in \mathscr{E}_{[f]}\} \to Re$ *for every* $[f] \subseteq \mathscr{F}/*$ *and* $D \in \mathscr{E}'_{[f]}$ *such that*

(i) *u preserves* $\succ$ *on* $\mathscr{M}(\mathscr{C})$,

(ii) *u is linear on* $\mathscr{M}(\mathscr{C})$,

(iii) *if* $B, C \in \mathscr{E}'_{[f]}$ *and* $B \subseteq C$, *the measures* $P^{[f]}_B$ *and* $P^{[f]}_C$ *satisfy the chain rule*

$$P^{[f]}_C(A) = P^{[f]}_C(B)\, P^{[f]}_B(A)$$

for every $A \in \mathscr{E}_{[f]}$ *such that* $A \subseteq B$,

(iv) $u(f, A \cup B) = P^{[f]}_{A \cup B}(A)u(f, A) + P^{[f]}_{A \cup B}(B)u(f, B)$ *for every* $f \in \mathscr{F}$ *and dichotomy* $\{A, B\} \subseteq \mathscr{E}'_{[f]}$.

Moreover, the measures $P^{[f]}_{D}$ *are unique and u is unique up to positive affine transformation.*

References

[1] Balch, M. On recent developments in subjective expected utility. In this volume, pp. 45–55.
[2] Bolker, E. D. A simultaneous axiomatization of utility and subjective probability. *Philosophy of Science,* **34** (1967), 333–340.
[3] Fishburn, P. C. *Decision and Value Theory.* Wiley: New York (1964).
[4] Fishburn, P. C. *Utility Theory for Decision Making.* Wiley: New York (1970).
[5] Fishburn, P. C. A mixture-set axiomatization of conditional subjective expected utility. *Econometrica,* **41** (1973), 1–25.
[6] Fishburn, P. C. On the foundations of decision making under uncertainty. In this volume, pp. 25–44.
[7] Herstein, I. N. and Milnor, J. An axiomatic approach to measurable utility. *Econometrica,* **21** (1953), 291–297.
[8] Jensen, N. E. An introduction to Bernoullian utility theory: I. Utility functions. *Swedish Journal of Economics,* **69** (1967), 163–183.
[9] Luce, R. D. and Krantz, D. H. Conditional expected utility. *Econometrica,* **39** (1971), 253–271.
[10] Luce, R. D. and Suppes, P. Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush and E. Galanter (eds.), *Handbook of Mathematical Psychology,* vol. 3. Wiley: New York (1965).
[11] Pratt, J., Raiffa, H. and Schlaifer, R. *Introduction to Statistical Decision Theory.* McGraw-Hill: New York (1965).
[12] Savage, L. J. *The Foundations of Statistics.* Wiley: New York (1954).
[13] Savage, L. J. Difficulties in the theory of personal probability. *Philosophy of Science,* **34** (1967), 305–310.
[14] von Neumann, J. and Morgenstern, O. *Theory of Games and Economic Behavior,* 2nd edition. Princeton University Press: Princeton (1947).

COMMENTS

The interpretation of conditional expected-utility theories

David H. Krantz and R. Duncan Luce

C3.1.1. Discussion

Balch [1] alleges a logical paradox arises in specifying an empirical interpretation for the conditional expected-utility theory of Luce and Krantz [5], which the work of Fishburn [4] and, especially, that of Balch and Fishburn [2] is said to overcome. In our view, however, it is at least as easy to specify an empirical interpretation for the Luce and Krantz (LK) axioms as for the Balch and Fishburn (BF) axioms, as we now argue.

A *conditional decision* is a function that specifies an outcome for each state in a conditioning event. To facilitate comparability with BF, we make a trivial change in our notation for conditional decisions: let f_A denote the restriction to the domain A of a function f whose domain includes at least the event A. This makes the LK object f_A exactly comparable to the BF object (f, A). Our notation permits some brevity in specifying new functions. For example, if A and B are disjoint events and f, g are conditional decisions with domains respectively including A, B, then we write $f_A \cup g_B$ for the function which is restriction to $A \cup B$ of the relation $f \cup g$.

In specifying an empirical interpretation, difficulties arise from the structural axioms since they force the set of objects over which preference, $\succsim$, is defined to be larger than is comfortable. The LK system includes two non-trivial groups of structural conditions: closure ones and solvability ones. The closure conditions, which are those Balch criticized, state that if f_A and g_B are conditional decisions, with $A \cap B = \varnothing$, then $f_A \cup g_B$ is also a conditional decision; and that, if $f_{A \cup B}$ is a conditional decision, then so is f_A. (For simplicity here, we exclude the possibility of null events other than $\varnothing$, and we assume any event symbolized by A, B, etc., is non-null.) The solvability conditions (which were not criticized, but which in fact pose practical difficulties for finding an acceptable empirical interpretation) specify, first, that given any A and any g_B, there exists f_A such that $f_A \sim g_B$; and, second, given

$$h_A^{(1)} \cup g_B \succsim f_{A \cup B} \succsim h_A^{(2)} \cup g_B,$$

there exists h_A with $h_A \cup g_B \sim f_{A \cup B}$.

The BF system likewise has two groups of structural conditions: mixture space conditions and a continuity axiom. The first states that all lotteries on finite sets of conditional decisions are in the domain of the preference relation. The second states, in effect, that there are no discontinuities in preference as a function of lottery probabilities.

Balch's criticism of the LK closure axioms is, in brief, the assertion that f_A, g_B, and $f_A \cup g_B$, $A \cap B = \varnothing$, cannot simultaneously be available for choice, and so one part or the other of the closure axiom is necessarily invalid. The argument is that if f_A and g_B are available for choice, then the decision maker controls whether A or B occurs. So $f_A \cup g_B$ is difficult to interpret, since whether or not the event A occurs, given $A \cup B$, is not determined by any natural mechanism external to the decision maker. For example, if one is free to choose between traveling by air, f_A, or by bus, g_B, then it is 'silly' both to decide to travel and to let the choice of A or B depend on a chance mechanism. On the other hand, if $f_A \cup g_B$ is a 'natural' option (for example, one decides to make the trip, but without knowing whether air reservations are available) then f_A simply is not, at that time, available as a unitary act.

To this criticism we reply that, if f_A and g_B are natural options, but $f_A \cup g_B$ is not, then our closure condition necessitates that we form a lottery, with the conditional probability of A given $A \cup B$ fixed by a chance mechanism at the disposal of an experimenter. This is certainly less restrictive than the BF mixture space requirement, which entails the consideration of not one, but arbitrary lotteries between f_A and g_B. In the other case, where $f_{A \cup B}$ is a natural option but f_A cannot be delivered, we are perfectly prepared to delete the event A from our algebra of events, incorporating $A \cup B$ and $f_{A \cup B}$ into a simpler structure for purposes of the analysis. (Alternatively, we may be able to keep the structure intact and use other axioms to make inferences about the preference for the unavailable f_A. For example, if it is known that $f_{A \cup B} \succsim f_B$, this certainly entails, with the other axioms, that $f_A \succsim f_{A \cup B}$.)

In short, for mathematical reasons, the natural alternatives, whose utilities we wish to analyze, must be embedded in a structure with the proper closure conditions. This means, in practice, that the natural alternatives must sometimes be supplemented by artificial ones in order to attain a measurement structure. This is no different in principle from other artificial constructions in measurement, for example, sets of standard weights used in connection with a pan balance. And certainly it compares favorably with the mixture space apparatus of the BF system, which has lotteries run by a 'Supergenie'.

The issue of behavioral observations *versus* hypothetical choices seems an extraneous one. It is perfectly true that if a man is free to choose travel by air or bus, f_A or g_B, then it may be difficult to persuade him to choose instead between $f_A \cup g_B$, where whether A or B occurs is determined by throwing a die, and h_B, going by bus with a bonus of \$100, for example. He simply may not submit to having only those two choices actually available to him, and so we may have to pose the choice as a hypothetical one and hope that the utilities and subjective probabilities determined in this way will predict actual choices in other situations. But this is a very general problem in the theory of preference, and it has no special bearing in the present framework. The LK axioms do not commit us, by their formal structure, to any fixed method of determining the $\succsim$ relation.

So far, we conclude that the closure restrictions in the LK system are less bothersome for empirical interpretation than the mixture-space apparatus of BF. The picture is a little less clear if we take account of the other structural restriction in LK, namely the solvability properties.

The solvability properties require that we be able to manipulate f while holding A fixed – that is, systematically vary the outcomes associated to different states of A. This is not always 'natural'. It means that to measure the utility of a natural decision, f_A, we must introduce artificial modifications of the assignment of outcomes to states. (For example, we could harass weary commuters on the Long Island Railroad by asking their preferences between $f_A \cup g_T$ and h_T, where g_T is a train trip and h_T is the same trip supplemented with a \$0.25 rebate for each minute behind schedule.) The BF system has no such requirement because of the great richness of its mixture space with numerical probabilities.

It seems to us that the chief difference between the two systems lies right at this point: the LK system imposes great richness on its outcome structure, but can get on with as few as three atomic events (or even two, with a little extra effort); the BF system can deal with any set of basic action alternatives, but utilizes the elaborate mixture-space apparatus. Presumably the latter apparatus could be made more qualitative by moving in the direction of the Savage axioms [7]; but in any case, what is required is a very fine-grained structure of events or probabilities. The essence of this difference is familiar from the contrast between the utility measurement procedure of refs. [3] and [6]. Does one best measure utility by trading off value and probability or by trading off value against value? The latter has more face validity and is more easily generalized to situations where the subjective probabilities are not well behaved; it is the method of Davidson *et al.* and of the LK system. The former method gets along with a much simpler structure of basic options; it is the method of Von Neumann and Morgenstern, Mosteller and Nogee, Savage, and others, and is the basis of the BF system.

It seems to us that both methods have their uses, and that the real contribution of the BF system is to extend the mixture space analysis to cover intrinsic conditional probabilities associated with decisions. However, in our opinion it does not have the advantage in interpretation over the LK system claimed by its authors.

References

[1] Balch, M. On recent developments in subjective expected utility. In this volume, pp. 45–55.

[2] Balch, M. and Fishburn, P. C. Subjective expected utility for conditional primitives. In this volume, pp. 57–69.

[3] Davidson, D., Suppes, P. and Siegel, S. *Decision Making: An Experimental Approach*. Stanford University Press: Stanford (1957).

[4] Fishburn, P. C. A mixture-set axiomatization of conditional subjective expected utility. *Econometrica*, **41** (1973), 1–25.

[5] Luce, R. D. and Krantz, D. H. Conditional expected utility. *Econometrica*, **39** (1971), 253–271.

[6] Mosteller, F. and Nogee, P. An experimental measurement of utility. *Journal of Political Economy*, **59** (1951), 371–404.

[7] Savage, L. J. *The Foundations of Statistics*. Wiley: New York (1954).

Frameworks for preference

Richard C. Jeffrey

C3.2.1. Discussion

If I now prefer p to q, what sorts of entities are the prospects p and q? In 1954, Savage [10] answered this question in what he took to be behavioristic terms: prospects – the terms of the preference relation – are functions from a set S of *states of nature* to a set C of *consequences*. He called such functions 'acts'. A decade or so later, Bolker and I answered it differently: prospects are *propositions*, i.e. (nearly enough) sets of possible states of nature where the human agent is taken to be part of nature and his acts are thus ingredients in states of nature. Now, branching off from work of Luce and Krantz, Balch and Fishburn propose a hybrid answer: prospects are act–event pairs and probability mixtures of such pairs. With Savage, they treat man and nature – acts and events – dualistically. Their treatment of acts is far more satisfactory than Savage's, and they are to be commended for the step toward holism which they take in dropping Savage's extraneous set C of consequences. But I will argue for the fully holistic or naturalistic position in which preference is a ranking of events (or propositions), some of which are acts.

With Balch and Fishburn, I find Savage's system unacceptable. Choice ought to reveal preference to at least this degree: the prospect which the agent chooses ought to be one of the highest in his preference ranking which he takes to be options for him. But in Savage's system, prospects seem to be entities of a sort among which finite beings cannot choose. For Savage, choice of an 'act' is choice of a scheme which associates a definite consequence with each of the infinity of possible states of nature. If these are acts, then only God could know what act is being performed. After performance, the human agent may learn what consequence his act associated with the actual state of nature, but neither before nor after its performance can he be expected to know what consequences the act associates with the rest of the possible states. Thus, for us, choice of an act cannot be choice of a particular function from states to consequences. To make Savage's scheme humanly applicable, one would have to modify it so as to make preference a relation between *sets* of functions from states to consequences. Choice of an action would then be choice of such a set, an unknown member of which will be realized. God *knows* which function is realized, and the human agent has *opinions* about the matter, expressed by a subjective probability distribution over the chosen set.

Balch and Fishburn seem to be making just this sort of complaint about Savage's theory in their footnote 3, where they contrast their theory with Savage's by temporarily formulating their proposal in something like his terms. The humanly available acts are various possible expenditures on advertizing, the consequences are various possible incomes from sales, and it is the *states* which are functions, namely, all functions from expenditures to incomes. Here human uncertainty has its proper object: not the identity of the act which is being performed, but the identity of the actual state of nature. I see this move as a definite conceptual improvement over Savage's representation of matters, and over the alternative just noted, of viewing preference as a relation between sets of Savage-style acts.

But Balch and Fishburn note this move only in passing, as a possibility. Their own move is to scrap Savage's consequences along with his functions from states to consequences. Hesitantly holistic, they take the basic prospects to be pairs (f, A) where f is a humanly possible act (not one of Savage's functions) and A is a subset of the set S_f of all possible *f-conditioned states of nature* (which need not be functions either). I take it that the agent prefers act f to act g when he prefers the pair (f, S_f) to the pair (g, S_g): human choice is among unconditioned acts, i.e. among vacuously conditioned acts. The point of including such pairs as (f, A) and $(f, S_f - A)$ in the preference ranking is presumably to allow the agent to analyze his attitude toward f itself, i.e. toward (f, S_f), as a function of his degree of belief in f given A and of his attitudes toward f as they would be if he knew that the true state would be in A and if he knew that the true state would not be in A.

But why not be a bit more holistic, and view the agent as part of nature? A state of nature would then specify what act the agent performs, along with everything else one usually takes it to specify. Preference would be a relation on a Boolean algebra of subsets of the set S of all such holistic states. Underlying the preference ranking would be functions u and P on states and sets of states respectively: A would be preferred to B if and only if the conditional expected utility $E(u \mid A)$ of u

on A were greater than $E(u \mid B)$, where both expectations are computed according to the probability measure P. Then prospects are propositions: sets of states. For the most part, prospects would be outside the agent's power to affect, for example, he might prefer fine weather tomorrow to rain tomorrow, even though there are no acts he can perform to realize either prospect. But among the prospects there would be certain propositions which he can make true or false as he pleases, and such propositions do duty as acts, for example, the act of taking his umbrella as he leaves the house in the morning would be represented by the set of states in which he does just that. Remember: the agent is part of nature, and his acts are ingredients in states of nature. In terms of the more conventional representation, in which the states of nature do not specify the agent's acts, my set S might be thought of as the cartesian product of the set of states with the set of acts (where the acts are not thought of as functions). But I prefer to think of acts as ingredients in states *ab initio*.

Such is the view which I put forth in my book and article of 1965 [6,5]. The thing would have been impossible but for the prior mathematical work of Bolker, set forth in full in ref. [2], condensed in ref. [3] and presented without technical details in ref. [4]. Bolker's existence and uniqueness theorems are novel and important from a measurement-theoretical point of view, as are his methods, which look to von Neumann's work on continuous geometries rather than to Holder's theorem on ordered groups.

It was the work of von Neumann and Morgenstern [9] which persuaded economists and statisticians that cardinal utilities do, after all, make sense, and it was the work of Savage [10] which persuaded them that subjective probabilities make sense. In each case, the process of persuasion took some time, and in each case there was prior work (by Ramsey and DeFinetti) which might have done the job of persuasion but did not: it was von Neumann–Morgenstern and Savage who finally got the ear of the public. By now, one no longer has to earn the right to deliberate in terms of subjective probabilities and utilities by first rebutting the ordinalists of the 1930s. Therefore I think it proper to characterize the notion of an ideally satisfactory preference ranking as one for which there exist a random variable u and a probability measure P relative to which the conditional expected utilities $E(u \mid \cdot)$ mirror preferences: A is at least as high as B in the ranking if and only if

$E(u \mid A) \geq E(u \mid B)$. From this characterization one can deduce that the preference relation is transitive, connected, etc. (Under 'etc.' we have the averaging condition: if A and B are disjoint prospects, then their union lies in the closed interval between them, in the preference ordering). One might take the set of all such consequences to be the general theory of preference. This is not to deny the importance of existence theorems like Bolker's, which give conditions on the preference ranking and on the algebra of prospects from which one can deduce the existence of functions u and P as above. On the contrary: it is because we have such existence theorems that the foregoing procedure seems feasible. Note, however, that Bolker's conditions are not intended as axioms for the general theory of preference. Those conditions restrict the algebra of prospects in important ways, and make certain special assumptions about the preference ranking, so that their consequences neither exhaust nor lie wholly within the general theory of preference as defined above. But that is inevitable: one cannot expect the conditions to be necessary as well as sufficient for existence of the functions u, P.

But there remains a uniqueness problem, even if we sidestep the existence problem as I have suggested above. Consider the preference relation which is determined by a particular pair u, P, where u is bounded above or below. One might expect that any other pair u', P' which determined the same preference relation would be related to the first pair by the conditions $P' = P$ and $u' = au + b$ where a is positive. But in fact these strong conditions do not hold. In fact (Bolker's uniqueness theorem) the relevant group of transformations for u is not simply positive affine: it is a certain more comprehensive subgroup of the projective transformations (with positive determinant). Nor is P uniquely determined by the preference relation: there will be a certain 'quantization' or uncertainty about the probabilities of propositions which appear above or below S in the preference ranking.

This underdetermination of u and P by the preference relation (unless u is unbounded above and below) is fascinating, but may be seen as a flaw. (Must one have preferences of unlimited intensities in order to have a perfectly sharp subjective probability measure?) If so, the flaw is removable, for example, by using *two* primitives: preference and comparative probability. With these primitives, one ought to be able to drop some of Bolker's restrictions on the algebra of prospects in favor of conditions on comparative probability and conditions connecting preference and comparative probability. I would expect that in this way one could get significantly closer to an existence theorem in which the conditions are necessary as well as sufficient for existence of u and P, while obtaining the usual uniqueness result: P is unique, and u is determined up to a positive affine transformation. It would be a job worth doing.

To see the situation clearly, let us now think of prospects as probability measures over a measure algebra of subsets of S. On S there is a fixed random variable u which assigns utilities to possible states, and one prospect is preferred to another when the (unconditional) expected utility of u is greater when computed according to the first probability measure than when computed according to the other. If we take the set of prospects to include all probability measures over the underlying measure algebra, the preference relation determines u up to a positive affine transformation as in the von Neumann-Morgenstern theory. But in the Bolker–Jeffrey theory, the prospects form a much thinner set than that: they stand in one-to-one correspondence with the probability measures $P(\cdot \mid A)$ which are obtainable from a fixed measure P (the agent's actual subjective probability measure) by conditionalization relative to the various measurable subsets A of S to which P assigns positive values. (Note that the unconditional expectation of u relative to the measure $P(\cdot \mid A)$ is the conditional expectation of u on A relative to the measure P.) Preferences among prospects in this thin set determine u only up to a wider set of transformations, and the basic measure P is not uniquely recoverable from preferences if u is bounded above or below.

But to get the stronger determination of u, one need not fatten the set of prospects very much: it would be enough to have one additional prospect Q which is not of form $P(\cdot \mid A)$ for any measurable subset A of S but which *is* known to be (say) a 50–50 mixture of two such prospects: $Q = \frac{1}{2}P(\cdot \mid A) + \frac{1}{2}P(\cdot \mid B)$ for measurable subsets A and B of S where $P(A) \neq 0 \neq P(B)$ and $P(\cdot \mid A)$ is preferred to $P(\cdot \mid B)$. One would then know that the expected value of u relative to Q is exactly half way between the expected values of u relative to $P(\cdot \mid A)$ and to $P(\cdot \mid B)$, and could use this fact to determine P uniquely and determine u up to a positive affine transformation. In these terms, the Balch–Fishburn sort of move would be to fatten the Bolker–Jeffrey set of probability measures by closure under *all* mixing operations: $aP(\cdot \mid A) + (1 - a)P(\cdot \mid B)$ with $0 \leq a \leq 1$ would be a prospect whenever $P(\cdot \mid A)$ and $P(\cdot \mid B)$ are.

These comments have been frankly tendentious and partisan. With Bolker, I applaud the Balch–Fishburn constructions as a real advance over Savage's approach, and over that of Luce and Krantz in so far as the latter continues to treat acts as functions, albeit partial functions, on a set of act-free states of nature. But I see the step from the Balch–Fishburn framework to Bolker's and mine as a further, natural advance to a truly holistic standpoint. If I have been noisy in my advocacy it is because that standpoint has not previously been called to the attention of foundationally-minded economists and because Bolker's methods have not yet found their place in the toolkits of measurement theorists.

References

[1] Balch, M. and Fishburn, P. Subjective expected utility for conditional primitives. In this volume, pp. 57–69.

[2] Bolker, E. D. *Functions resembling quotients of measures.* Dissertation, Harvard University (April 1965).

[3] Bolker, E. D. Functions resembling quotients of measures. *Transactions of the American Mathematical Society,* **124** (1966), 292–312.

[4] Bolker, E. D. A simultaneous axiomatization of utility and subjective probability. *Philosophy of Science,* **34** (1967), 333–340.

[5] Jeffrey, R. C. New foundations for Bayesian decision theory. In Y. Bar-Hillel (ed.), *Logic, Methodology and Philosophy of Science.* North-Holland: Amsterdam (1965), 289–300.

[6] Jeffrey, R. C. *The Logic of Decision.* McGraw-Hill: New York (1965).

[7] Luce, R. D. and Krantz, D. H. Conditional expected utility. *Econometrica,* **39** (1971), 253–271.

[8] von Neumann, J. *Continuous Geometry.* Princeton University Press: Princeton (1960).

[9] von Neumann, J. and Morgenstern, O. *Theory of Games and Economic Behavior.* Princeton University Press: Princeton (1944).

[10] Savage, L. J. *The Foundations of Statistics.* Wiley: New York (1954).

Remarks on 'Subjective expected utility for conditional primitives'

Ethan D. Bolker

C3.3.1. Discussion

I am pleased to have this chance to add to the remarks I made several years ago [4] on the problems addressed here by Michael Balch and

Peter Fishburn. The style and mathematical level of ref. [2] splendidly combine rigor and accessibility. The model they choose has two striking advantages. First, it is truly conditional: acts do affect the state of the world. Second, consequences are holistic: your reward for action is the world as you have helped to make it.

There are, however, aspects of their theory which I find less satisfactory. The 'Bolker objection' (which could just as well have been named the Jeffrey objection) says that it is unreasonable to ask a decision maker to express preferences about events or lotteries he feels cannot occur. We must ask when the Balch–Fishburn lottery

$$\alpha(f, A) + (1 - \alpha)(g, B) \tag{C3.1}$$

legitimately lies in the domain of the decision maker's preference order $\succ$. The most restrictive answer would be 'Only when $f \neq g$, $A = S_f$ and $B = S_g$'. Lottery (C3.1) is then merely a mixed strategy on the set $\{f, g\}$. But with so few allowed gambles there might be no theory. I see little objection to lottery (C3.1) provided $f \neq g$, even when $A \neq S_f$ and $B \neq S_g$. To consider (f, A) the decision maker must ask himself how he would feel if he chose act f and event A occurred. That should not overtax his imagination even though he can in no way cause the logically possible event A to come to pass.

When $f \neq g$ the events (f, A) and (g, B) occur in different conditional worlds, no *a priori* subjectively correct α accompanies lottery (C3.1) and so our decision maker might agree to consider (C3.1) for every value of α. However, when $f = g$ he may say 'I cannot conceive of lottery (C3.1), because in choosing f I have already changed the world as much as I can. I don't believe that then $P^f_{A \cup B}(A) = \alpha$. I cannot conjure up a Supergenie who could make it so, since I do not believe in magic and my subjective feelings about probabilities in $\mathscr{E}_f$ are as much a part of that world as its events.' That is the Bolker objection conditioned by f, and it still stands. The objectionable lotteries are just what Balch and Fishburn use to determine subjective probabilities in $\mathscr{E}'_f$, for if $A \cap B = \varnothing$ and there is a unique α for which

$$(f, A \cup B) \sim \alpha(f, A) + (1 - \alpha)(f, B)$$

then they sensibly define $P^f_{A \cup B}(A)$ to be α. That seems to me a superficial deduction of numerical subjective probabilities from ordinal utilities. It is tantamount to asking the decision maker directly for his estimate of $P^f_{A \cup B}(A)$.

It may be that the subjective probabilities of events in $\mathscr{E}'_f$ can be recovered from preferences among the allowed mixed strategies. If such were the case Balch and Fishburn could counter my objection and avoid the *ad hoc* introduction of direct scaling for subjective probabilities. But I would still prefer a lottery-free theory, one in which no reference was made even to those non-objectionable gambles. It may not be necessary to construct such a theory, however, since presumably real as well as hypothetical decision makers are familiar with lotteries nowadays, and Herstein and Milnor's results [7] cry out for application. I once attempted a lottery-free theory [3, 4]. That theory was technically cumbersome; it used countably additive measures and a continuum of states. Balch and Fishburn avoid both these psychologically unrealistic requirements. There are some structural similarities between my theory and theirs.

I recognized one difficulty they encounter, that of finding a consistent definition for $P^f_{A \cup B}(A)$ when $(f, A) \sim (f, B)$. Their solution is to postulate Savage classes satisfying a robustness axiom A3.5 and an independence axiom A3.6. I can imagine situations in which both axioms fail because $[f] = \{f\}$: the Savage class is a singleton. In fact, that would seem to me more typical than not, for in a truly conditional theory such as theirs $\mathscr{E}_f \neq \mathscr{E}_g$ when $f \neq g$ and so, *a fortiori*, $g \notin [f]$. I find the alternative to Savage classes sketched in footnote 12 unconvincing: if the hypothetical side payment really exists it is already counted in the construction of $\mathscr{E}_f$; if it does not, the Bolker objection prevails.

I resolve unbifurcated dichotomies by finding a C disjoint from $A \cup B$ for which f bifurcates $\{A, B \cup C\}$. A technical observation shows that it suffices to bifurcate those dichotomies, for which there is such a C. My Axiom ii (impartiality), like Balch and Fishburn's A3.6, guarantees that the probabilities thus obtained are both well defined and additive. A similar idea might work in Balch and Fishburn's theory. I think they could modify their model and their methods so as to answer the objections I have raised without sacrificing the clarity and elegance they have achieved.

My final reservations apply to all theories of this kind, including my own. I wonder if such axioms really describe behavior, whether 'practical' military and political decision makers really use these theories and, if they do, whether they ought to. Perhaps what we need instead is an axiomatization of morality.

References

[1] Balch, M. On recent developments in subjective expected utility. In this volume, pp. 45–55.
[2] Balch, M. and Fishburn, P. C. Subjective expected utility for conditional primitives. In this volume, pp. 57–69.
[3] Bolker, E. D. Functions resembling quotients of measures. *Transactions of the American Mathematical Society*, **124** (1966), 292–312.
[4] Bolker, E. D. A simultaneous axiomatization of utility and subjective probability. *Philosophy of Science*, **34** (1967), 333–340.
[5] Fishburn, P. C. On the foundations of decision making under uncertainty. In this volume, pp. 25–44.
[6] Hampshire, S. Morality and pessimism. *New York Review of Books*, **19** (January 25, 1973), 26–33.
[7] Herstein, I. N. and Milnor, J. An axiomatic approach to measurable utility. *Econometrica*, **21** (1953), 291–297.
[8] Jeffrey, R. C. *The Logic of Decision*. McGraw-Hill: New York (1965).

*Some comments on some axioms for decision making under uncertainty**

John W. Pratt

C3.4.1. Introduction and miscellany

To the extent that these comments are not *ex cathedra*, they are about only the three papers by Balch and Fishburn in this volume [1–3]. To bound my task, and for other reasons, both good and base, I have not delved into the antecedents of these papers or into axiom systems for decision theory generally. A convenient cover for some preliminary remarks is, however, provided by mentioning some desiderata which might be invoked for such axiom systems, particularly when the resulting model is not at issue.

---

* This originated as a comment on the paper presented by Fishburn [3] at the Third NSF–NBER Conference on Decision Rules and Uncertainty, Iowa City, May 1972. Oversights may have occurred in adapting it to take account of the two subsequent papers [1, 2]. If so, I apologize. I am grateful for the support of the National Science Foundation (Grant GS-2994) and the John Simon Guggenheim Memorial Foundation.

C3.4.1.1. Simplicity is one. Unfortunately there is no clear way to define it. If one axiom system actually contains another which leads to the same conclusion, then the situation is clear. If two systems overlap, we might try to talk about some axioms implying others, in the presence of still others. But even if we succeed in avoiding a quagmire, is this kind of thinking relevant to the simplicity that matters? Suppose two systems are identical except that two axioms in system A are replaced by one in system B. Is system B necessarily better? Not if its one axiom is incomprehensible and A's two are intuitively appealing. The simplicity that matters is not a technical concept.

C3.4.1.2. Elegance. Whatever the definition, Savage certainly gets high marks for having fundamentally internal calibrating events rather than grafting on or implicitly assuming the necessary calibration. If, however, the decision maker is in fact going to calibrate by reference to an external or hypothetical event, this elegance is of peripheral importance. Thus I am sympathetic to bringing in mixed acts. Once this is done, however, both the set of acts and the set of events are fundamentally as infinite as Savage's, since the acts and events comparable to Savage's include the mixed acts and the mixing events. It is unfair to compare just the unmixed acts and internal events with Savage's [1, p. 50].

C3.4.1.3. Mathematical difficulty of proofs. This is really irrelevant to the value of an axiom system, if not to the chances of getting it published.

C3.4.1.4. Weakness. Provided it gives the desired result, a weaker system is generally better, though even this isn't universal if the weaker system is incomprehensible; see section C3.4.1.1. Though weakness and simplicity are undoubtedly related, they are certainly not identical in relation to one of the issues of occasional concern here: what is the most innocuous enrichment of the available acts and events enabling one to prove the expected utility theorem? It is tempting to trumpet others' enrichments while hiding one's own in footnotes and obscure axioms. Balch and Fishburn rarely succumb [2, axioms A3.5 and A3.6 and perhaps footnote 12], but comparing innocuousness is an inherently vexing problem. Are Savage's requirements more or less nocuous than Supergenie lotteries? Furthermore, some writers may have included unnecessary acts or events for simplicity or through oversight when the richness seemed inconsequential to them. Fairness then requires pruning their systems where this is easy but not where it is hard – a delicate distinction. What acts does Savage really need and what could be eliminated?

More fundamentally, enrichment has one desirable aspect: it generalizes the conclusions. Among other advantages, this may facilitate assessment of probability and utility; see section C3.4.1.5 below. To play the elimination game properly, then, one would have to distinguish carefully the minimum needed to prove existence from additions needed to extend the domain of applicability.

Most authors don't bother, and I don't blame them. I doubt it would be worth the candle. Whatever the professional players' other purposes, the main purpose of the axioms is to convince intelligent amateurs that expected utility ought to guide decision making. This calls for compromise in many respects, including elimination. Nice distinctions don't help. Once convinced, even for a limited class of problems, decision makers will extend the method as necessary without hesitation. They will even adopt act-dependent probabilities when appropriate. It would surprise me if anyone has truly been deterred from accepting the faith by scruples over the unavailability of any of the acts commonly introduced, or over act-dependent probabilities, even if these are occasional debating points for those committed to other faiths.

C3.4.1.5. Constructiveness. If a decision maker is to use subjective expected utility theory to reach decisions, what assessments or hypothetical choices will he make? Many axiom systems are like Balch and Fishburn's, which might lead him to think, if he didn't know better, that he was expected to preference-order the entire mixture set $\mathcal{M}(\mathcal{C})$, or at least to assess a utility function on all act–event pairs. They drop contrary hints in the accompanying text, of course, including several references to scaling, albeit as a concession, and an interest in obtaining eventually a true expectation representation (3, p. 29 (2.2*), p. 40 (iv), (v), etc.; 2, p. 67 (3.5), but obscured in the later theorem]. Perhaps in papers for professionals only it goes without saying that, regardless of how the axioms look, the decision maker will actually assess only enough probabilities to determine those he needs, and the utilities of the consequences or, here, the act–state pairs (both to an adequate approximation). It would be unfair to say of any axiom system that 'the subject is required to express binary preference over the set of *all*' acts, act–event pairs, or whatever [1, p. 45; see also 2, p. 61]. Still, it is an advantage if an axiom system bears some relation, or even better, provides some guidance to the construction of probability and utility, beyond merely asserting their existence. This, and the similar idea of requiring only simple judgments and deducing more complex ones, were very much in our thinking when Raiffa, Schlaifer and I played this game [5, 6]. The simplest judgments to make relate to simple but hypothetical acts such as constant acts (consequences) and two-outcome lotteries. The whole object is to use these simple judgments to arrive at judgments of complex, real acts. From this point of view, it is desirable, as well as elegant, to draw as rich a body of conclusions from as little input data as possible; thus the difficulties of the elimination game (section C3.4.1.4) apply to preference relations as well as to the sets of acts, events, etc.

To summarize the discussion so far, it is far from obvious how to judge the desirability of various features of an axiom system. A lot depends on what game you are playing, by what rules. Many of Balch and Fishburn's comparisons should be taken with plenty of salt. Furthermore, as long as everyone is getting the same model, the preceding desiderata are of rather little importance.

C3.4.1.6. Intuitive acceptability, on the other hand, is vital. If we don't feel compelled to accept the axioms, they are pointless. I don't find Fishburn's axiom 2.3 or either version of axiom 2.4 (see also A3.4) [2, 3] easy to accept unless I think in terms of consequences (see below). Fishburn's axioms 2.6 and 2.7 are superficially quite unacceptable: they don't even permit an irrelevant, three-outcome event to be included. I presume they can be made acceptable by adding hypothetical acts – just the kind of enrichment his paper is trying to avoid. Balch and Fishburn [2] embrace considerable enrichment, but their A3.5 is not directly comprehensible, and even a mathematician might find it difficult to accept or reject their A3.6 on the face of it. Their reference to 'handsome notational savings' [2, p. 68] suggests no very nice form of the latter exists, so perhaps I may be forgiven inadequacies and imprecisions in the following attempt to state the essence of it: Given any act f and any disjoint events A, B, C such that, if you chose f, you would prefer A to B and B to C, there exists an act g such that:

(i) the mixing probability which would make you indifferent between B and a probability mixture of A and C if you were choosing f would not do so for g; and

(ii) for all disjoint events D and E, the same mixing probability would make you indifferent between $D \cup E$ and a probability mixture of D and E whether you were choosing f or g.

Thus the acts f and g have (i) different prizes but (ii) the same probabilities (and the same event algebras). A3.6 is not 'simply ... sidestepping another possible "knife-edge" anomaly' but is guaranteeing adequate, though not complete, independence of acts and events ('recall the discussion in footnote 12' with a vengeance) [2, p. 68]. Since the significant difference from other systems resides exactly here, I believe, it is unfortunate that rigorous statement is so formidable and elucidation so elusive.

To look at it another way, let $U(A) = u(f, A)$ be the (conditional expected) utility resulting from a fixed act f as a function of the event A. Any real-valued set function U whatever would be compatible with Balch and Fishburn's A3.1–A3.3. Their A3.4 requires only that $U(A \cup B)$ lie between $U(A)$ and $U(B)$ for disjoint A and B. We want $U(A)$ to be the conditional expectation given A of some function under some probability measure $P = P^f$. If such a probability exists, the values $P(A \mid A \cup B)$ can be determined from U whenever $U(A) \neq U(B)$ by

$$U(A \cup B) = P(A \mid A \cup B)U(A) + (1 - P(A \mid A \cup B))U(B)$$

for disjoint A and B [2, p. 64 (3.2)]. A3.5 extends this to $U(A) = U(B)$. Nothing so far, however, implies that the values thus determined should behave like probabilities. More remains than 'to tie up some loose ends' [2, p. 65]. A3.1–A3.5 are satisfied, for instance, if there is a [see also 4] single act f, all non-empty subsets of $\{0, 1, 2\}$ are events, and preferences correspond to

$$U\{i\} = i^2, \qquad U\{i, j\} = (i^2 + j^2)/2, \qquad U\{0, 1, 2\} = 1.2$$

for $i, j = 0, 1, 2$. If a probability existed, however, it would have to satisfy

$$U\{0, 1, 2\} = P\{i\}U\{i\} + (1 - P\{i\})U(\{0, 1, 2\}\text{-}\{i\}) \qquad (i = 0, 1, 2)$$

and hence $P\{0\} = 0.52$, $P\{1\} = 0.80$, $P\{2\} = 0.20$. That's a bit much.

Probably no axiom system can entirely avoid the need for care in interpreting its primitive terms (acts, events, consequences, etc.) in the real world. For example, unexceptionable as axiom 2.4 (A3.4) seems in either form [2, 3], it would not apply to Mr. Jetsetter's choice of two weekend invitations, one to a beach, the other to visit Mr. Rich, who will fly his guests to either a skiing area or a golf course but won't decide which in advance. Mr. Jetsetter prefers either skiing or golf to the beach, but can't carry both skis and golf clubs on the plane he must take to Mr. Rich's city, and prefers the beach to the risk of being ill-equipped. He would rather hear either 'Mr. Rich will choose skiing' or 'Mr. Rich will choose golf' than 'I refuse to tell you what Mr. Rich will choose' [cf. 3, p. 31]. To salvage axiom 2.4, he must define three acts: go to the beach, go to Mr. Rich's taking skis, go to Mr. Rich's taking golf clubs. Acts cannot involve reaction to events.

I have gotten so far off course that it is time to start a new section.

C3.4.2. Consequences, constant acts and Supergenies

Before discussing the elimination of consequences and constant acts and the conjuring up of Supergenies, let me clear away one point: usual theories do not really require that a consequence possess no uncertainty, whatever their authors may have said. The residual uncertainty Fishburn rightly calls attention to [3, p. 32] is or perfectly well could be allowed. A consequence is generally an uncertain prospect faced by the decision maker when he has chosen a particular act and those uncertainties of nature which he had included in his model of states have been resolved but others have not. There is a trade-off between the complexity of the probability model and the complexity of the consequences whose utility must be assessed. The utilities we assess usually could have been 'derived' from utilities defined on more refined consequences. What Savage actually said is far more than Fishburn indicates or I can summarize, but a representative quotation [7, p. 84] is: 'I therefore suggest that we must expect acts with actually uncertain consequences to play the role of sure consequences in typical isolated decision situations.'

C3.4.2.1. Eliminating consequences. How is one to think about the utility of act–state or act–event pairs? Are they really 'more general' or different in a 'major respect' [3, p. 26]? When saying 'the use of a particular act when a particular state is assumed to obtain is preferred to . . .' would it be any different to say 'result of using' or 'prospect resulting from using' or 'consequence of using' in place of 'use of'? Would there be any real difference between 'imagine what might happen if' [3, p. 27;

2, p. 61] and 'imagine the possible consequence if'? I don't think so. Balch may or may not think so [1, p. 45 or footnote 7, p. 50]. Indeed, Balch and Fishburn even go so far as to speak of 'prizes' belonging to a set suggestively denoted $\mathscr{C}$ [2, p. 61], though they stop just short of introducing consequences. While consequences may not be a 'conceptual imperative' [2, footnote 6], they are implicitly present, and must be thought about in the way consequences usually are in applying the theory, so it might be clearer to get them out in the open. It would reduce the danger of consequences being misinterpreted as opportunity losses. (It would have been clearer if Fishburn had said 'it is better that Mr. Accused be guilty and we free him than that he be innocent and we convict him' [3, p. 28].) It would also require specifying whether consequences correspond to act–event pairs for possibly rich and complicated mixed acts and compound events, or only for unmixed acts and atomic events, i.e. states. It is the latter we want to assess utility as a function of.

C3.4.2.2. Constant acts. Because the preference relation must apply to more than consequences, it is technically convenient to identify them with some other element of the system, and in some systems, with constant acts, for example, ref. [7], p. 25, and ref. [5], p. 365. (The constant acts are, of course, no more constant than the consequences, possessing the same residual uncertainty [cf. 3, p. 26].) This identification can be dropped, at some cost in cumbersomeness, if the hypothetical nature of constant acts is disturbing. But it is surely easy enough (all too easy) to *imagine* a hypothetical act certain to lead to a broken leg tomorrow, or death in an airplane crash, whether or not such an act is actually available. If anyone really had trouble identifying consequences with constant acts, I would worry about his contaminating his utility assessments with his probabilities, though it's his privilege if he wants to.

C3.4.2.3. Supergenie. Since it had never occurred to me that there was any real difficulty imagining the *hypothetical* constant acts and gambles over consequences which the theories require, Supergenie seems to me not essentially different or new, but just an anthropomorphization (or deification) of what everyone should have been thinking about all along. Thus I am closer to the view that 'this little fantasy need not have been spun' than to the view that Supergenie is 'a conceptual imperative for this kind of mind experiment', but he is clear and vivid, and if he helps anyone, I am all for him (quoting out of context [1, footnote 10] and [2, footnote 6]).

In sum, these differences from usual theories seem to me notational, descriptive, and expository rather than fundamental. (At least this mitigates Fishburn's self-criticism [3, pp. 27–28].) They may also facilitate some improvements in weakness (section C3.4.1.4). Balch and Fishburn do not address this question squarely, however, and it is certainly too complicated for me to, though what I suspect is the main improvement is discussed in the next section.

C3.4.3. Utility and probability

It is probably clear already that, as far as utility is concerned, Balch and Fishburn seem to me to provide little change from usual theories. The domain of definition of utility and the problem of assessing it are not changed in any essential way, and as they indicate, their A3.1–A3.3, which imply its existence, are essentially classical.

Probability, however, is quite another matter. The joint paper [2] leads to a probability measure conditional on the act chosen, for each possible act, without requiring the existence of any joint probabilities across acts. This is a substantial difference from usual theories, and on the face of it an advantage. To reach a decision on the basis of expected utility, one clearly needs a utility whose meaning crosses acts, but only act-conditional probabilities. Though this is, of course, evident from the conclusion of other theories, it is an advantage from the point of view of 'constructiveness' (section C3.4.1.4 above) to reflect it in the axioms.

There are, however, tempering considerations in some situations. Assessing probabilities conditional on acts could lead to 'act-distortion' effects. (Perhaps Balch and Fishburn would favor these as well as 'prize-distortion' effects [2, p. 60], though if I identified either of these kinds of effects in myself, I would try to purge them as irrational.) And there are sometimes other reasons for bringing the 'Savage set of states-of-the-world' on stage instead of letting it 'rest comfortably behind the scenes' [2, p. 59]. For example, rather than assessing a probability distribution of sales for each advertising level individually, it may be both simpler and more robust to assess a probability model of the relationship between advertising and sales, with some unknown parameters, and a subjective probability distribution for the parameters, perhaps based on some data, and to deduce from these a distribution on sales for each advertising level. This would be even more true in problems where decision trees are much simpler than direct consideration of all possible strategies, such as the design of a marketing experiment and subsequent choice of advertising level.

Of course, a conditional theory would permit a states-of-the-world model as a special case. When Savage was writing, incidentally, the clarification brought to statistical inference by states-of-world models was not so old that it would have been easy, or perhaps even safe, to give it up.

It would indeed be an advantage to have also the '*direct conceptual link* between objective probabilities... and subjective probabilities' which Balch and Fishburn claim to provide [2, p. 59], but I cannot find it in their axioms. The usual link involves defining new acts with standard prizes on an event and its complement, but that seems to be ruled out if events and probabilities are act-conditional. In fact, probabilities seem to be derivable only by way of utilities, as indicated in section C3.4.1.6 above, and even then may require some act-independence, smuggled in through the definition of the equivalence class $[f]$ and A3.5 [2, p. 66]. Of course, once the expected utility theorem is proved, independent scaling experiments can be incorporated immediately and lotteries involving them rated by expected utility. Presumably this is just Balch's embedding [1, pp. 52–53], although I would have expected the axioms to be satisfied automatically.

Incidentally, the definition of $[f]$ appears too inclusive and A3.5 consequently too strong. One should not be required to assess common probabilities over the widest possible set of acts allowed by the utility assessments. This could probably be handled outside the axiom system by redefining the event algebras, or within it by allowing the classes $[f]$ to partition the equivalence classes or by amalgamating events when all subsets of their union are equally desirable for a given act.

This is as good a place as any for a few remarks on conditional probabilities given events of probability zero. Fishburn says [3, p. 28, 30] that they exist and are unique, even when not prescribed by the unconditional probability measure, and that they are needed for continuous outcomes. I would say almost the opposite: they need not exist unless the decision maker wants them to, and he might want them in the theory for unobservable, continuous parameters but not for observable outcomes. A known outcome, information received, is fundamentally discrete, even in a sequential experiment. Continuous distributions of observables and conditional probabilities given observed events of probability zero are merely approximations and should be handled outside the axiom system. However, a decision maker might *want* to assess a continuous distribution for an unobservable parameter, such as the parameter of a Bernoulli process or an advertising–sales relationship, and conditional distributions of observables given this parameter. He then needs an additional axiom to get his whole unconditional probability distribution. Raiffa, Schlaifer and I once treated this subject along these lines, somewhat unenthusiastically and perhaps inadequately, because it was impeding our primary purpose [6, especially chapter 10]. If Fishburn's theory [3] really implies the existence and uniqueness of all conditional probabilities given non-empty events of probability zero, then apparently, since they are not determined by the unconditional probability distribution, the decision maker will be forced to make a great many hypothetical, not to say ineffable, judgments which he might not want to make.

C3.4.4. Summary and apology

The burden of my remarks has been that many of the criticisms Balch and Fishburn make of other axiom systems, and the distinctions they draw, have much less substance than their words suggest, while the important differences lie elsewhere and are not adequately brought out or explicated. I have dwelt, perhaps unduly, on the former because I suspect others have similarly overplayed matters of form, though I am not prepared to cite chapter and verse. The important differences are still sufficiently problematical and in need of further work so that extensive discussion would perhaps be premature.

I want to make clear, however, that my comments have not given a balanced view of Balch and Fishburn's papers, or attempted to, but have concentrated on what seem to me points of weakness or misemphasis, because they are, unsurprisingly, what I had most to say about. Just as Balch and Fishburn basically agree with Savage, while criticizing some aspects of his work, so I basically agree with them, though they have not yet succeeded as well as Savage. I agree heartily with their conclusions, and their justifications of them are new in significant respects and will doubtless be improved in the future. I am all in favor of any argument which will convince anyone not already convinced that maximizing expected utility is the only behavior worth rational consideration. In as much as I am already convinced, my satisfaction with traditional arguments should be discounted.

References

[1] Balch, M. On recent developments in subjective expected utility. In this volume, pp. 45–55.
[2] Balch, M. and Fishburn, P. C. Subjective expected utility for conditional primitives. In this volume, pp. 57–69.
[3] Fishburn, P. C. On the foundations of decision making under uncertainty. In this volume, pp. 25–44.
[4] Fishburn, P. C. A mixture-set axiomatization of conditional subjective expected utility. *Econometrica,* **41** (1973), 1–26.
[5] Pratt, J. W., Raiffa, H. and Schlaifer, R. The foundations of decision under uncertainty: an elementary exposition. *Journal of the American Statistical Association,* **59** (1964), 353–375.
[6] Pratt, J. W., Raiffa, H. and Schlaifer, R. *Introduction to Statistical Decision Theory.* McGraw-Hill: New York (1965).
[7] Savage, L. J. *The Foundations of Statistics.* Wiley: New York (1954).

REPLY TO COMMENTS*

M. Balch and P. C. Fishburn

The 'Bolker–Jeffrey objection'[1] arises when a decision maker finds it impossible to conceive a fair-spinner lottery $\alpha x + (1 - \alpha)y$ because he simply does not believe that the lottery-master can guarantee prize delivery[2]. In particular, the objection carries full force when the 'prizes', as in our theory, are intrinsically characterized in terms of natural contingencies over which no mortal may sensibly be expected to exercise control (for example, no one can guarantee good weather tomorrow on

* References in this reply are to the list given on p. 69.

[1] We thank Professor Bolker for pointing out that the 'objection' (which appears in ref. [2]) had been jointly formulated with Professor Jeffrey.

[2] On face, at least, the lottery $\alpha x + (1 - \alpha)y$ is 'supposed to' offer prize x with (objective) probability α and prize y with complementary probability.

the toss of a coin). One way to sidestep this objection (and there are compelling model-theoretic reasons for wanting to do so, as discussed in our paper) is to provide a conceptual framework, *necessarily hypothetical*, within which such lotteries make sense. The device that we structured for this purpose is a 'druthers' relation over 'Supergenie' lotteries, but embedding the real world in a (logically consistent) imaginary one is not enough; the reader must agree to enter that world as well, perhaps after the fashion of a Walter Mitty or an Alice (or, indeed, of some non-fictional characters with whom we are acquainted). Of course we accept that some people may find this difficult, even impossible[3]. But so long as we insist on representing preference orders *only* over alternatives that are actually available in the real world, then it would appear that we cannot have a 'full-property' theory of subjective expected utility which is also logically consistent (for a discussion of how the Savage [12] and Luce–Krantz [10] theories shortfall in this respect, see the Balch and Balch–Fishburn papers in this volume, and a further comment on 'constant acts', below).

There is another point of some interest. Professor Bolker correctly notes that we elicit subjective probability numbers of the form $P^f_{A \cup B}(A)$ (for $A \cap B = \varnothing$) by simply asking the subject a direct question (but this involves a Supergenie lottery $\alpha(f, A) + (1 - \alpha)(f, B)$ of the type that Bolker still finds objectionable); he goes on to suggest that we might find a more palatable way to cull such numbers from the underlying preference relation. We have two remarks. First, our 'direct question'[4] – which, note, is well put within a Supergenie framework – provides a most elementary[5] conceptual linkage between the subjectivist ('degree of belief') and objectivist ('statistical') interpretive traditions in probability theory; this might be regarded as a plus. Second, the Bolker suggestion has in fact been carried through for the case of an unconditional framework by Fishburn [5, see also 6] in a way that *makes no use whatever* of Supergenie lotteries. But the remarkable fact is that the Fishburn utility indicator and subjective probability measures are precisely those that emerge when his system is 'Supergeniefied' (this

---

3 The earlier Fishburn approach [5] may alleviate this difficulty for some, but note our remark on 'mathematical equivalence' below.

4 This idea had already been advanced (somewhat less formally) in ref. [11]; cf. footnote 6 in the Balch–Fishburn paper in this volume.

5 Cf. remarks in the introductory section of the Balch paper in this volume.

extension procedure is canonical and unique; see the final section of ref. [1], and footnote 11 in particular). Moreover, the Fishburn axiom system could be generalized to the present conditional framework without essential difficulty (though it would be even more technically involved than its unconditional prototype). Instead, we chose the Supergenie route in our (subjective!) application of an Occam's razor. The tradeoff was between two systems that are *mathematically equivalent* (up to canonical extension/restriction): one might be held to be the more conceptually palatable, but produces subjective probabilities in somewhat arcane fashion; the other produces the *same* numbers by means of a direct mind-experiment but demands, in turn, a more delicate exposition.

Finally, we agree with Professor Bolker that the typical 'bare bones' decision context that one meets in practice is often one for which our generic Savage equivalence class $[f]$ is a singleton $\{f\}$. On the other hand, it is desirable (from the viewpoint of theory closure, as discussed at length in our paper) to provide a sort of 'robustification' correspondence for $[f]$. The device we chose (as described in footnote 12) embeds the original decision context in one which (i) is now sufficiently rich, and (ii) has no (behavioral) effect whatever on its antecedent. But due to a descriptive oversight on our part, Professor Bolker challenges that $\mathscr{E}_{f_A} = \mathscr{E}_f$, where $\mathscr{E}_f$ is the event algebra for a 'basic' act f and $\mathscr{E}_{f_A}$ is the event algebra for its 'free-disposal modification' to f_A (see footnote 12). The point is that these algebras agree *by definition*: the only difference between these acts, after all, is that for f_A the subject is to dispose of something 'desirable' only *after* the 'f-natural' event A has been realized.

In brief reply to the comments of Professors Krantz and Luce, see the added footnote which now closes the introductory section in ref. [1].

We shall reply to just a few of the comments made by Professor Pratt.

Perhaps the first thing to say is that some of the formal distinctions that we have drawn with other variants of the expected utility paradigm were motivated, in part, by some difficulties that we experienced in attempting to interpret their primitives from the viewpoint of The Subject (whose behavior and cognitive abilities the paradigm purports to model). The classical Savage notion of a constant act, for one fairly widely recognized example, simply was not comprehensible to us as a well-defined 'something' that an actor might 'do'[6], or even imagine 'doing', except in decision contexts that appeared to us to be unduly restrictive. How, for example, to conceive of boarding a given airplane flight that will result in a safe arrival *regardless* of whether or not the airplane crashes[7]? Of course the subject of this paradigm family is idealized in the sense that he is assumed to compute faultlessly and choose consistently (i.e., 'rationally'), but then – and especially then – what could it mean to also suppose that he can make sense of a mathematical construction that translates for his decision context as a *logical* impossibility? The simple recasting of primitives that we have proposed avoids this sort of insensibility (by taking its cue from the actual choice set at hand), and provides an axiomatic formulation of the paradigm for some more general situations of analytic interest (cf. footnote 1 in our paper).

In our model, the consequences of action are referred to as suppositioned circumstances for reasons that have to do with the contingency-*only* manner in which some of them may come to be realized, and with the way in which we have recast the Savage primitives. Pratt seems to think that these ('pure') consequences are somehow hidden from plain view, but for every $f \in \mathscr{F}$ they are just the sub-collection of holistically interpreted symbols $\{(f, \{s\}) \mid s \in S_f\}$.

Professor Pratt 'salvages' A3.4 by applying it correctly; viz., separately, to each well-defined decision alternative in turn. The difficulty he encounters in what he takes to be a counter-illustration occurs because his 'act' f: {go to Mr. R's for vacation} cannot be interpreted as an holistically well-defined alternative for Mr. J. without further qualification as to the details of its implementation. In particular, if Mr. J. should go on such a visit, then he *cannot do so without* either taking golf clubs, or skis, or neither, or both. Moreover, whatever shall obtain as the *holistic* consequence (for Mr. J.) of {going to Mr. R's} depends upon which of these qualifying arrival modes is necessarily chosen by Mr. J. (as well as upon whether Mr. R. then decides to go golfing or skiing). We may say it this way: the suppositioned circumstances (f, A) for the (unqualified) Pratt f are non-holistically described

---

[6] Cf. the preceding comment by Professor Jeffrey, which airs much the same difficulty.

[7] This is logically different from imagining that you board that flight *and* that it does not crash, which is the sort of mind-image that the subject of our model is assumed to be able to form for *ex ante* consideration.

because an evaluation-relevant aspect of *decision* has been left un-described, simply, and the interpretive spirit of the weak ordering A3.1 is thus violated. In other words, Mr. J.'s subjective 'picture' of (go to Mr. R.'s, Mr. R. chooses to ski), say, and his binary evaluation of that picture against other such things, depends in general upon the way in which f is necessarily 'completed' for the context at hand. A3.4 applies for each such completion without difficulty, as Pratt observes. Of course the Pratt f may be regarded as a (non-singleton) *set* of acts that are each holistically well-described, and then given the obvious utility assignment $u(f) \underset{Df}{=} \sup_{g \in f} u(g)$, but such act-unions contribute nothing of behavioral substance to the EUP. Pratt concludes his illustration with the curious statement that '[Balch–Fishburn] acts cannot involve reaction to events'. But Mr. R.'s decision can be known to Mr. J. only after Mr. J. arrives on the scene; this is simply the sequential structure of Mr. J.'s decision context. Mr. J. is free to react to Mr. R.'s decision in any one of a variety of ways ('go back home', 'stay, and "make the best" of an unhappy circumstance', ...). What Mr. J. cannot do, of course, is to change the historical fact of his arrival mode.

PART 2

CHAPTER 4

VENTURES, BETS AND INITIAL PROSPECTS*

Clifford Hildreth

4.1. Introduction

This paper is concerned with dichotomous decisions under uncertainty and with one-dimensional families of choices. In the dichotomous cases, a decision maker chooses one of two random variables whose values represent alternative levels of wealth for the decision maker. One random variable, called the initial prospect, indicates how his wealth will be related to events outside his control if he retains his present assets and carries out his present plans and commitments. The other gives wealth as a function of developments in his environment if he chooses to modify his current position by signing a new contract (which could be an agreement to cancel an old contract), purchasing or selling some securities, buying or canceling insurance, expanding his business, placing a bet, or some other dealing that can be expected to affect his wealth, at least under some circumstances. The possible undertaking which would modify his initial prospect is called a venture.

The model for dichotomous choice is developed in section 4.2. It differs from some models appearing in the literature in posing a choice between two uncertain prospects rather than between a fixed level of wealth and an uncertain prospect. This is seen to be of some significance since an important aspect of any venture is its statistical relation to the initial prospect. If one alternative is a fixed level of wealth, this degenerate random variable is statistically independent of the venture.

* Earlier versions of this paper were presented at a Department of Economics Seminar at the University of Minnesota in April 1972, and at the Third NSF–NBER Conference on Decision Rules and Uncertainty at the University of Iowa in May 1972. The author is indebted to Peter Clark, James C. Hickman, Daniel McFadden and S. Y. Wu for useful suggestions. Research was conducted under Grant GS-3317 of the National Science Foundation.

The relevance of dependence between the initial prospect and the venture is easy to indicate in a vague fashion. If the venture is such that it tends to add help if things otherwise go badly and to impose a cost if things otherwise go well, then it tends to stabilize the decision maker's outlook and, if he is a risk averter, such a venture will be more valuable than one which offers similar probabilities of gain and loss but is independent of the initial prospect. Alternatively, but still speaking loosely, a risk averter should prefer, other things equal, a venture that is negatively correlated with his initial prospect.

In section 4.3, a decomposition of the change in utility due to undertaking the venture is suggested. Components due to expected gain, spread and dependence of the venture are distinguished. Interpretations of and possible approximations to elements of the decomposition are discussed.

One-dimensional families of ventures are introduced in section 4.4. Such a family is generated by a venture which can be undertaken in various amounts or sizes represented by multiples of a basic random variable. Favorable and optimal choices are related to derivatives of utility with respect to a variable representing size of the venture. The first derivative is decomposed in a fashion corresponding to the decomposition of utility in section 4.3.

Bets are defined and considered in a formal way in section 4.5. The large number of mutually favorable potential bets between pairs of individuals is noted. In section 4.6 some conjectured reasons that most people don't engage in betting are discussed and some hypothetical circumstances in which intelligent betting might have merit are suggested. Examples are cited which suggest that analyzing hypothetical bets may sometimes suggest less expensive insurance arrangements or more effective pricing policies.

The paper raises issues rather than settling any. The theoretical model does seem to serve the purpose of permitting a first approximation analysis of the effect of dependence between the venture and the initial prospect, but many questions will require a more elaborate model. Trying to express the relevant outcome of the decision process as a real variable (called wealth) is clearly a gross simplification whose effects should be explored. For many purposes, explicitly dynamic models will be needed. As we learn more about the form of utility functions, better approximations to utilities of particular kinds of ventures for particular kinds of decision makers should be possible.

4.2. Utility of Gain

Let $(\Omega, \mathscr{F}, P)$ be a probability space where an element ω of Ω represents a possible sequence of developments in a decision maker's environment and P represents the decision maker's subjective probability.

S, with typical element s, is a set of possible strategies or actions for the decision maker. A particular strategy and sequence of developments in his environment determines an outcome x:

$$x = \xi(\omega, s).$$

For given s, $X_s(\omega) = \xi(\omega, s)$ is a random mapping to the outcome space. In this paper the outcome is wealth measured in money and X_s is therefore a random variable.

φ is a von Neumann–Morgenstern utility of wealth and the decision maker's problem is to

$$\underset{s \in S}{\text{maximize}}\; E\varphi(X_s).$$

It is assumed throughout that

$$\varphi' > 0,\ \varphi'' \text{ is continuous, } |E\varphi(X_s)| < \infty\ \forall\, s \in S. \tag{4.1}$$

We begin by considering a simplified case in which only two alternative choices are feasible. The corresponding random variables are denoted X and $X + Y$.

X is called the initial prospect. It determines the distribution of outcomes that prevail if the decision maker carries out this current plans and commitments. Y is called the new venture and indicates how level of wealth will be affected under each alternative ω if a currently available bet, contract, or other undertaking is accepted. The problem, then, is to compare $E\varphi(X + Y)$ with $E\varphi(X)$.[1]

---

[1] Applications of this formulation are more general than may appear at first glance. If the prospective venture 'interacts' with the current prospect so that they are not additive, one can always define Z as the prospect that will result from taking the new venture and then set $Y = Z - X$. For example, if X represents mainly returns from an existing processing plant and the new venture is an additional plant whose construction would interfere in some ways with operation of the existing plant, then possible returns from existing and prospective plants would not be additive. However, one could define Z as prospective returns from both facilities and then define $Y = Z - X$ as the new venture for purposes of analysis. In fact, given a dichotomous choice from any source, one can always label the alternatives X and $X + Y$ and proceed to compare.

$$\begin{aligned} E\varphi(X + Y) &= \int\varphi(X + Y)\mathrm{d}P = \iint\varphi(x + y)F_{XY}(\mathrm{d}x, \mathrm{d}y) \\ &= \int F_Y(\mathrm{d}y)\int\varphi(x + y)F_{X \mid Y=y}(\mathrm{d}x) \\ &= \int F_X(\mathrm{d}x)\int\varphi(x + y)F_{Y \mid X=x}(\mathrm{d}y), \end{aligned} \tag{4.2}$$

where F is the distribution function of the random variable or variables indicated by subscripts.

In general, the decision depends intimately on φ, X, Y, P. Because it is frequently possible to know Y more completely than X it is of interest to know under what conditions there might exist a reasonably stable function of gains or losses in wealth such that knowledge (exact or approximate) of this function and the distribution of the new venture might be sufficient to determine the choice.

This point of departure differs from some in posing a choice between two prospects both of which are uncertain. In the real world everyone constantly faces a variety of contingencies that might affect wealth – property may be stolen, damaged or inherited; liabilities may be accidentally incurred; currency in one's billfold may turn out to be counterfeit. I doubt that certain wealth is ever experienced. If this is true it is more realistic to think of a new venture as an opportunity to modify an existing uncertain prospect rather than an opportunity to exchange a certain prospect for an uncertain one.

Note that if we define

$$\beta(y) = \int\varphi(x + y)F_{X \mid Y=y}(\mathrm{d}x)$$

then

$$E\varphi(X + Y) = \int\beta(y)F_Y(\mathrm{d}y).$$

Unless further conditions are imposed, β changes whenever P, φ, X or Y changes, so one could not investigate the form of β by observing the response to alternative ventures Y with known distributions.

Define utility *after* gain, say $\tilde{\psi}(y)$, as the expected utility of the decision maker if he receives an outright gift of y dollars, i.e.

$$\tilde{\psi}(y) = E\varphi(X + y) = \int\varphi(x + y)F_X(\mathrm{d}x). \tag{4.3}$$

Let

$$\psi(y) = \tilde{\psi}(y) - E\varphi(X) = \tilde{\psi}(y) - \tilde{\psi}(0) \tag{4.4}$$

be the utility *of* gain with $\psi(0) = 0$.

Now consider the decision on a new venture Y when Y is independent of the current prospect X:

$$\begin{aligned} E\varphi(X + Y) &= \iint \varphi(x + y)F_{XY}(\mathrm{d}x, \mathrm{d}y) = \int F_Y(\mathrm{d}y) \int \varphi(x + y)F_X(\mathrm{d}x) \\ &= \int \tilde{\psi}(y)F_Y(\mathrm{d}y) = E\tilde{\psi}(Y) \end{aligned} \tag{4.5}$$

Any new venture Y that is independent of X is favorable if the expectation of $\tilde{\psi}(Y)$ is greater than $E\varphi(X)$, i.e. if $E\psi(Y) > 0$. The function $\tilde{\psi}$ does not change as Y changes so long as the alternative ventures are independent of X.

For any y, $\tilde{\psi}(y)$ is an average of possible values of $\varphi(x + y)$. If conditions for differentiating under the integral hold[2], the derivatives of $\tilde{\psi}$ are similar averages of derivatives of φ.

$$\tilde{\psi}^{(n)}(y) = \int \varphi^{(n)}(x + y)F_X(\mathrm{d}x) = E\varphi^{(n)}(X + y). \tag{4.6}$$

Under this assumption, we may note:

(1) If $\varphi^{(n)}$ has the same sign throughout its domain, then $\tilde{\psi}^{(n)}$ has that sign everywhere.

(2) If $\varphi^{(n)}$ has predominantly one sign, the exceptions could be 'averaged out' and $\tilde{\psi}^{(n)}$ could have the predominant sign everywhere. For example, if φ has the Friedman–Savage form (predominantly concave, a convex region), $\tilde{\psi}$ might still be strictly concave.

Averaging preserves the form of some, but not all, commonly considered utility functions. For example:

(3) If $\varphi(x) = a + bx$ then $\tilde{\psi}(y) = a + bEX + by$; if $\varphi(x) = a\mathrm{e}^{bx}$ then $\tilde{\psi}(y) = a(E\mathrm{e}^{bx})\mathrm{e}^{by}$. For $a < 0$, $b < 0$, the latter is the constant absolute risk aversion function, and the coefficient of risk aversion $-b$ is not changed.

---

[2] Sufficient conditions for eq. (4.6) to hold for a particular n and y are
(i) $\varphi^{(m)}(X + y)$ is integrable for $m = 0, 1, \ldots, n - 1$.
(ii) $\varphi^{(m)}(x + y)$ exists for all x in the range of $X(\omega)$ and for $m = 1, 2, \ldots, n$.
(iii) There exists an integrable function, $g(x)$, such that

$$\frac{\varphi^{(m)}(x + y + h) - \varphi^{(m)}(x + y)}{h} \leq g(x)$$

for $m = 0, 1, \ldots, n - 1$ and for all h in an interval about 0.

For $n = 1$, this is a standard theorem. See, for example, ref. [9], p. 67, or ref. [12], p. 217. Under the stated conditions, the theorem can be applied to successive derivatives.

(4) If φ is a polynomial of degree m and X has m moments, then $\tilde{\psi}$ is a polynomial of degree m.

(5) If φ is a decreasing risk aversion function $\varphi(x) = (x + d)^c$ with $0 < c < 1$, then the form of $\tilde{\psi}$ depends on F_X.

Assuming that sufficient ventures that are indepenent of the initial prospect X are available, one can use simple lotteries (for example, as described by Pratt *et al.* [14]) to determine $\tilde{\psi}$. Is there a natural way to then explore φ and F_X? McFadden has indicated the following possibility.

For any given level of wealth, say z, let the experimenter offer a guarantee of z, i.e. if the decision maker realizes $X(\omega)$ from his current prospect, the experimenter will supply $z - X(\omega)$. The experimenter then offers to exchange the guarantee of z for various outright cash gifts and determines a gift y such that the decision maker is indifferent between receiving the gift y or a guarantee of z. This implies that

$$\varphi(z) = \tilde{\psi}(y). \tag{4.7}$$

To determine the distribution function F_X at a selected point x, the experimenter chooses two levels of wealth, $z_1 \neq z_2$, and guarantees the decision maker z_1 if the event $(X \leq x)$ occurs and guarantees z_2 in the event $(X > x)$. He then finds the gift y that is indifferent to this prospect of z_1 or z_2 for the decision maker. Then

$$\tilde{\psi}(y) = F_X(x) \cdot \varphi(z_1) + (1 - F_X(x)) \cdot \varphi(z_2) \tag{4.8}$$

or

$$F_X(x) = \frac{\tilde{\psi}(y) - \varphi(z_2)}{\varphi(z_1) - \varphi(z_2)}. \tag{4.9}$$

4.3. A Possible Decomposition

Since the effect on expected utility of undertaking a new venture can be analyzed more simply if the venture is independent of the current prospect, it seems reasonable sometimes to analyze the effect of a dependent venture Y in two steps. First consider a hypothetical venture that has the same distribution as Y but is independent of X and then consider the difference in the effect of Y and the effect of the hypothetical venture.

Accordingly, for any venture Y, let W be a random variable that is independent of both X and Y and has the same distribution[3] as Y. Now consider

$$E\varphi(X + Y) - E\varphi(X) = [E\varphi(X + \overline{W}) - E\varphi(X)] + [E\varphi(X + W) - E\varphi(X + \overline{W})] + [E\varphi(X + Y) - E\varphi(X + W)], \quad (4.10)$$

where $\overline{W} = EW$.

Call the difference on the left the 'utility of the venture' η. Call the successive terms in square brackets on the right the 'utility of expected gain' η_G, the 'utility of spread' η_S, and the 'utility of dependence' η_D. Note that signs and ratios of these utilities and those that would be obtained for alternative ventures are invariant with respect to positive linear transformations of φ.

A venture is favorable – i.e. undertaking the venture increases expected utility – if $\eta > 0$. In the symbols just introduced[4],

$$\eta = \eta_G + \eta_S + \eta_D, \quad (4.11)$$

with

$$\eta_G = E\varphi(X + \overline{W}) - E\varphi(X) = \psi(\overline{W}) = \psi(\overline{Y}), \quad (4.12)$$

$$\eta_S = E\varphi(X + W) - E\varphi(X + W) = E\psi(W) - \psi(\overline{W}) = E\psi(Y) - \psi(\overline{Y}), \quad (4.13)$$

$$\eta_D = E\varphi(X + Y) - E\varphi(X + W), \quad (4.14)$$

---

[3] This is always possible. If the original universal event Ω does not permit the definition of such a random variable, define $\Omega^* = \Omega \times R$ where R is the real line. Let $\tilde{P}$ be the probability measure on R determined by F_Y and let P^* be the product measure $P \times \tilde{P}$ on Ω^*. Then define X^* by $X^*(\omega, r) = X(\omega)$, Y^* by $Y^*(\omega, r) = Y(\omega)$ and W by $W(\omega, r) = r$.

[4] Using the notions of gain, spread and dependence, there can be six somewhat different decompositions depending on the order in which the components are introduced. One could, for instance, let $\eta = \eta_S^* + \eta_D^* + \eta_G^*$ where

$$\eta_S^* = E\varphi(X + W - \overline{Y}) - E\varphi(X),$$
$$\eta_D^* = E\varphi(X + Y - \overline{Y}) - E\varphi(X + W - \overline{Y}),$$
$$\eta_G^* = E\varphi(X + Y) - E\varphi(X + Y - \overline{Y}).$$

For sufficiently small increments of ventures, such differences between these decompositions become negligible.

where, as defined in section 4.2, $\psi(y) = E\varphi(X + y) - E\varphi(X)$, and $\bar{Y} = \bar{W}$ is the common mean of Y and W.

Utility of expected gain is the increment of expected utility that would accrue to the decision maker if he were given a gift equal to the mean, or (subjective) actuarial value, of the venture. Since $\psi(0) = 0$ and $\psi' > 0$, η_G agrees with the sign of $\bar{Y}$. For a risk averter, $\varphi'' < 0$; hence $\psi'' < 0$ and, by Jensen's inequality, $\eta_S < 0$ (see the second equality of (4.13)). Also, for risk averters, η_D tends to be negative if high values of X and Y occur together more frequently than if X and Y were independent; and η_D tends to be positive if high values of X occur with low values of Y more frequently than if X and Y were independent. This is admittedly a very loose statement; later, it will be made more precise for special cases.

For the present, perhaps the statement about η_D can be given a little plausibility by supposing that X takes either a high value x_1 or a low value x_2 and that Y takes only two values, $y_1 > y_2$. Since W has the same marginal distribution as Y, W is also equal to y_1 or y_2. Let p_{ij} be the probability that $X = x_i$, $Y = y_j$ for $i, j = 1, 2$. Let q_{ij} be the probability that $X = x_i$, $W = y_j$. Then $\Sigma_i p_{ij} = \Sigma_i q_{ij}$ for $j = 1, 2$ and $\Sigma_j p_{ij} = \Sigma_j q_{ij}$ for $i = 1, 2$. Let $\delta = p_{11} - q_{11}$; then $p_{12} - q_{12} = p_{21} - q_{21} = -\delta$ and $p_{22} - q_{22} = \delta$. So

$$\eta_D = \sum_i \sum_j (p_{ij} - q_{ij})\varphi(x_i + y_j) = \delta[\varphi(x_1 + y_1) + \varphi(x_2 + y_2) - \varphi(x_1 + y_2) - \varphi(x_2 + y_1)]. \tag{4.15}$$

If φ is concave, the coefficient of δ is negative since

$$(x_1 + y_2) = \alpha(x_1 + y_1) + (1 - \alpha)(x_2 + y_2),$$
$$(x_2 + y_1) = (1 - \alpha)(x_1 + y_1) + \alpha(x_2 + y_2), \tag{4.16}$$

where

$$\alpha = \frac{x_1 - x_2}{x_1 + y_1 - x_2 - y_2}$$

and, by concavity,

$$\varphi(x_1 + y_2) > \alpha\varphi(x_1 + y_1) + (1 - \alpha)\varphi(x_2 + y_2),$$
$$\varphi(x_2 + y_1) > (1 - \alpha)\varphi(x_1 + y_1) + \alpha\varphi(x_2 + y_2) \tag{4.17}$$

and, adding,

$$\varphi(x_1 + y_2) + \varphi(x_2 + y_1) > \varphi(x_1 + y_1) + \varphi(x_2 + y_2). \quad (4.18)$$

Thus if like values of X and Y are more probable than like values of X and W, δ is positive and η_D negative.

If the venture being considered is an expansion of the decision maker's present business, we expect the venture to be positively correlated with his initial prospect. Thus, if he is a risk averter, we expect that $\eta_D < 0$. Since $\eta_S < 0$ for risk averters this would mean that, to be favorable, the venture would have to offer substantial expected gain so that $\eta_G > 0$ could overbalance η_S, η_D.

On the other hand, for a typical insurance policy, $\bar{Y} < 0$ and $\psi(\bar{Y}) < 0$ (provided the decision maker is not materially more optimistic than the company's tables would justify) so η_D must be sufficiently large to compensate for both $\eta_S < 0$, $\eta_G < 0$. Note that, in this case, Y is positive if a specified loss occurs and negative if it does not occur, so Y is expected to be negatively correlated with X.

Purchase of securities usually involves positive η_G, and η_D could be either positive or negative depending on how contingencies determining yield and/or appreciations of the new securities compare with contingencies determining outcomes under the current prospect. If the favorable contingencies tend to be different we expect $\eta_D > 0$.

We usually think of $\eta_G > 0$ for bets on such events as sporting events or elections. η_G can be positive for both parties if their subjective probabilities of the basic event differ. It will be argued in section 4.6, however, that intelligent betting might involve $\eta_G < 0$, $\eta_D > 0$. By the definition used there, many insurance policies are bets of this type or are combinations of several bets.

For a bet on a random device – cards, dice, etc. – the venture is independent of the current prospect (unless the prospect already includes a bet on the same trial) which implies $\eta_D = 0$. If the gambler in a casino knows the odds, $\eta_G < 0$. His gambling must then be explained by risk preference, which would make $\eta_S > 0$, or by some consideration not included in the theory sketched in section 4.2. Smith [15] suggests modifying the utility function of habitual gamblers to include a positive utility of being in a gambling situation. This is consistent with the views expressed in a popular news magazine [13]. In this paper, the habitual gambler will be neglected and the theoretical framework sketched in section 4.2 retained.

If one can develop approximate relations between the components of a decomposition and properties of the utility function and the random variables, the decomposition may aid in establishing probable reactions to alternative ventures under various circumstances. For this to be realized we need to develop some knowledge of the properties of actual utility functions.

For example, if ψ were a polynomial of degree N, it would follow that

$$\psi(Y) = \psi(\bar{Y}) + \sum_{n=1}^{N} \frac{1}{n!}(Y - \bar{Y})^n \psi^{(n)}(\bar{Y}) \tag{4.19}$$

and

$$\eta_S = E\psi(Y) - \psi(\bar{Y}) = \sum_{n=2}^{N} \frac{1}{n!} \psi^{(n)}(\bar{Y}) E(Y - \bar{Y})^n. \tag{4.20}$$

Thus, to approximate η_S we would have to be able to approximate the first derivatives of ψ at $\bar{Y}$ and the first N moments of Y. If ψ were cubic

$$\eta_S = \tfrac{1}{2}\psi^{(2)}(\bar{Y})\text{Var } Y + \tfrac{1}{6}\psi^{(3)}(\bar{Y})E(Y - \bar{Y})^3. \tag{4.21}$$

If the distribution of Y were approximately symmetric, the last term would be small and the product of the second derivative and the variance would approximate η_S. By similar reasoning, η_D could be approximated for special cases. Realistically, however, one could safely use a polynomial approximation to ψ or φ only if he could be sure that tail probabilities were sufficiently small. If the coefficient of the highest order term of a polynomial representing φ is negative, then $\varphi' < 0$ for sufficiently large wealth; if the coefficient is positive, then φ'' becomes positive and arbitrarily large as wealth increases.

4.4. One-Dimensional Families of Ventures

A family g of possible ventures will be said to be one-dimensional if there exists a nontrivial (not almost surely equal to 0) venture Y, called a base, such that $Z \in g \Rightarrow Z = \alpha Y$ for some real number α. Since only one-dimensional families are considered in this paper, the designation is usually omitted. A family will be called adaptable if α can be any real number.

A one-dimensional family represents a situation in which the gains or losses conditioned on various events will be proportional to an amount that the decision maker elects to stake or invest[5]. Although adaptable families are not often encountered in practice (α is usually confined to a proper subset of R), it turns out that starting with an adaptable family is an effective way to study a variety of nonadaptable one-dimensional families.

Consider a family $g = (\alpha Y)_{\alpha \in R}$, where R is the real line, and let $\eta(\alpha)$ represent the utility of the venture αY, i.e.

$$\eta(\alpha) = E\varphi(X + \alpha Y) - E\varphi(X). \tag{4.22}$$

Thus $\eta(\alpha) > 0 \Leftrightarrow \alpha Y$ is a favorable venture.

We are interested in finding favorable and optimal ventures and will assume that two differentiations under the expectation in eq. (4.22) are valid. Then

$$\eta'(\alpha) = EY\varphi'(X + \alpha Y), \tag{4.23}$$

$$\eta'(0) = EY\varphi'(X). \tag{4.24}$$

If $\eta'(0) > 0$ it follows from an elementary theorem (see, for example, ref. [2], p. 91) that there is a neighborhood of 0 in which the sign of $\eta(\alpha)$ agrees with the sign of α. This implies that there is a $\delta > 0$ such that any $\alpha \in (0, \delta)$ corresponds to a favorable venture αY. Similarly if $\eta'(0) < 0$, $\exists\, \varepsilon > 0$ such that ventures with $\alpha \in (-\varepsilon, 0)$ are favorable.

Note that if we ignore the possibility of nonpayment, two people can always exchange opposite ventures Y, $-Y$. The party that takes Y will receive $Y(\omega)$ from the other if ω is realized and $Y(\omega)$ is positive, and he will pay $-Y(\omega)$ to the other if $Y(\omega)$ is negative for the realized ω. The argument of the preceding paragraph justifies the following proposition which is numbered for future reference.

---

[5] One must be careful not to assume that every venture which can be undertaken in various amounts defines a one-dimensional family. Various sized additions to a factory will not generally yield proportionate returns under all events. A \$10 000 insurance policy on a given property is not an exact multiple of a \$1000 policy. However, bets and purchases of a specific security are usually exact or approximate one-dimensional families.

PROPOSITION 4.1. *Let $\eta(\alpha)$ and $\eta^*(\alpha)$ represent utilities of the ventures $(\alpha Y)_{\alpha \in R}$ for two decision makers. If $\eta'(0)$ and $\eta^{*\prime}(0)$ are of opposite signs, $\exists\, \alpha$ such that the exchange of αY, $-\alpha Y$ is mutually favorable.*

PROOF. Suppose $\eta'(0) > 0$, $\eta^{*\prime}(0) < 0$. By the paragraph cited, $\exists \delta > 0 \ni \alpha \in (0, \delta)$ corresponds to favorable ventures for the first party and $\exists \delta^* > 0 \ni \alpha \in (-\delta^*, 0)$ corresponds to favorable ventures for the second party. Choose $\alpha \in (0, \min \{\delta, \delta^*\})$ and let the first party receive αY and the second party $-\alpha Y$.

Proposition 4.1 does not require any assumption about risk aversion. If we return to the problem of a single decision maker and assume that $\varphi'' < 0$ everywhere, then

$$\eta''(\alpha) = EY^2\varphi''(X + \alpha Y) \tag{4.25}$$

is also everywhere negative. Thus η is strictly concave, η' is decreasing, and optimal and favorable ventures can readily be characterized by

PROPOSITION 4.2. *If $\varphi'' < 0$, then*

(i) *$\eta'(\alpha) = 0$ has at most one solution. If a solution $\hat{\alpha}$ exists then $\hat{\alpha}$ uniquely maximizes $\eta(\alpha)$.*

(ii) *If $\eta'(0) > 0$ [resp. < 0] and $\hat{\alpha}$ exists, $\exists \alpha^* \ni 0 < \hat{\alpha} < \alpha^*$ [$\alpha^* < \hat{\alpha} < 0$] and $\eta(\alpha) > 0 \Leftrightarrow \alpha \in (0, \alpha^*)$ [$(\alpha^*, 0)$].*

(iii) *If $\eta'(0) > 0$ [resp. < 0] and $\eta'(\alpha) = 0$ has no solution, then $\eta(\alpha)$ is monotonic increasing [decreasing] and $\eta(\alpha) > 0 \Leftrightarrow \alpha \in (0, \infty)$ [$(-\infty, 0)$].*

The proof is not given since an easily constructed diagrammatic argument is sufficient.

Clearly one could readily investigate nonadaptable families with the aid of proposition 4.2 by comparing the set of feasible α with $\hat{\alpha}$ and with the intervals of favorable ventures.

In most economic reasoning, the underlying relations are not completely specified and one seeks to conclude something from partial information. Two questions that naturally arise in connection with proposition 4.2 are (1) when does $\hat{\alpha}$ exist? and (2) can partial knowledge be used to indicate the sign of $\eta'(0)$? The following proposition substantially answers (1).

PROPOSITION 4.3. *Suppose $\eta'' < 0$, $\lim_{x \to \infty} \varphi'(x) = 0$, $\exists \alpha \ni EY\varphi'(X + \alpha Y) < \infty$. Then $\eta'(\alpha) = 0$ has a solution $\Leftrightarrow P(Y > 0)$, $P(Y < 0)$ are both positive.*

PROOF. Suppose $P(Y < 0)$. Then the integrand of $\eta'(\alpha) = \int Y\varphi'(X + \alpha Y)$ is nonnegative and $\eta'(\alpha) \geq 0$ with equality only if the integrand is a.s. 0, i.e. if $Y = 0$ a.s. which was excluded at the outset to avoid trivial ventures.

Now suppose $P(Y > 0)$, $P(Y < 0)$ are both positive. Let I_A be the indicator of A and express

$$\begin{aligned}\eta'(\alpha) &= \int I_{[Y>0]}Y\varphi'(X + \alpha Y) + \int I_{[Y<0]}Y\varphi'(X + \alpha Y)\\ &= A_\alpha \qquad\qquad\qquad\qquad\quad + B_\alpha\end{aligned}$$

Clearly A_α is positive and decreasing while B_α is negative and decreasing. It will therefore suffice to show that $\lim_{\alpha\to\infty} A_\alpha = \lim_{\alpha\to-\infty} B_\alpha = 0$. This will show that $\eta'(\alpha)$ is negative for sufficiently large α, positive for sufficiently small α and, by continuity, zero somewhere in between. Let $\{\alpha_n\} \uparrow \infty$, define $Z_n = I_{[Y>0]}Y\varphi'(X + \alpha_n Y)$. Then $\{Z_n\} \downarrow 0$.

Let α_0 be such that $EY\varphi'(X + \alpha_0 Y)$ is integrable. Then $\alpha_n > \alpha_0 \Rightarrow Z_n$ is integrable and $EZ_n \to 0$ by the dominated convergence theorem. Proving $\lim_{\alpha\to-\infty} B_\alpha = 0$ is similar.

Clearly if $\eta'' < 0$ is required to hold only for sufficiently large and sufficiently small α, $\eta'(\alpha) = 0$ still has a solution but multiple solutions are not excluded. Under conditions of proposition 4.3, part (iii) of proposition 4.2 applies only to sure-thing ventures and is therefore uninteresting. Note that bounded φ is sufficient but not necessary for the condition that $\lim_{x\to\infty} \varphi'(x) = 0$.

Turning to determination of the sign of $\eta'(0)$ from partial knowledge, I suspect various approaches will prove useful in different contexts. One example that may sometimes be helpful is developed below. Recall the decomposition of section 4.3 and write

$$\eta(\alpha) = E[\varphi(X + \alpha Y) - \varphi(X)] = \eta_G(\alpha) + \eta_S(\alpha) + \eta_D(\alpha), \qquad (4.26)$$

where

$$\begin{aligned}\eta_G(\alpha) &= E\varphi(X + \alpha\bar{Y}) - E\varphi(X) = \psi(\alpha\bar{Y}),\\ \eta_S(\alpha) &= E\varphi(X + \alpha W) - E\varphi(X + a\bar{Y}) = E\psi(\alpha Y) - \psi(\alpha\bar{Y}),\\ \eta_D(\alpha) &= E\varphi(X + \alpha Y) - E\varphi(X + aW) = E\varphi(X + \alpha Y) - E\psi(\alpha Y)\end{aligned} \qquad (4.27)$$

and, as before, W is a random variable that is independent of X and Y and has the same distribution as Y and $\psi(y) = E\varphi(X + y) - E\varphi(X)$.

Call $\eta'_G(0)$, $\eta'_S(0)$, $\eta'_D(0)$ the initial marginal contributions of gain, spread, and dependence, respectively.

$$\eta'_G(\alpha) = \bar{Y}\psi(\alpha\bar{Y}), \tag{4.28}$$

$$\eta'_G(0) = \bar{Y}\psi'(0), \tag{4.29}$$

$$\eta'_S(\alpha) = EY\psi'(\alpha Y) - \bar{Y}\psi'(\alpha\bar{Y}), \tag{4.30}$$

$$\eta'_S(0) = \psi'(0)(\bar{Y} - \bar{Y}) = 0. \tag{4.31}$$

As has often been observed, for sufficiently small ventures, spread is not important.

Let $\bar{Y}_x$ be the expected value of Y given that $X = x$. Then

$$\eta'_D(\alpha) = EY\varphi'(X + \alpha Y) - EW\psi'(X + \alpha W) \tag{4.32}$$

$$\begin{aligned} \eta'_D(0) &= EY\varphi'(X) - \bar{Y}E\varphi'(X) \\ &= \int F_X(\mathrm{d}x) \int (y - \bar{Y})\varphi'(x) F_{Y|X=x}(\mathrm{d}y) \\ &= \int (\bar{Y}_x - \bar{Y})\varphi'(x) F_X(\mathrm{d}x). \end{aligned} \tag{4.33}$$

Combining eqs. (4.33), (4.31) and (4.29),

$$\eta'(0) = \eta'_G(0) + \eta'_D(0) = \int \bar{Y}_x \varphi'(x) F_X(\mathrm{d}x). \tag{4.34}$$

In cases of partial knowledge, $\bar{Y}_x$ may not be known. If, however, $\bar{Y}_x$ is monotonic, its monotonicity may be known and, with concavity, this is sufficient to determine the sign of the initial contribution of dependence. To avoid triviality in $\bar{Y}_x$ we assume $P(X \neq \bar{X}) > 0$.

PROPOSITION 4.4. *If $\varphi'' < 0$ and $\bar{Y}_x$ is nondecreasing*[6] *[resp. nonincreasing] then $\eta'_D(0) \leq 0$ $[\geq 0]$. If $\bar{Y}_x$ is strictly monotonic, strict inequalities may be substituted in the conclusions.*

PROOF. Let $\bar{Y}_x$ be nondecreasing and not a.s. equal to $\bar{Y}$. Consider the disjoint sets $A_1 = \{x\colon \bar{Y}_x > \bar{Y}\}$, $A_2 = \{x\colon \bar{Y}_x < \bar{Y}\}$. A_1 lies to the right of A_2 and since $\varphi'(x)$ is strictly decreasing $\exists \bar{x} \ni \varphi'(x_1) \leq \varphi'(\bar{x}) \leq \varphi'(x_2)$ for all $x_1 \in A_1$, $x_2 \in A_2$ with strict inequality on at least one side.

---

[6] Since versions of $\bar{Y}_x$ can differ on sets of measure zero, this could be more accurately stated: 'if there is a version of $\bar{Y}_x$ that is nondecreasing and corresponds to a regular conditional probability'. Since $\bar{Y}_x$ enters only under an integral, the choice of a particular version does not affect the equations.

Therefore:

$$\begin{aligned}
&(1)\ \textstyle\int_{A_1}(\bar{Y}_x - \bar{Y})\varphi'(x)F_X(\mathrm{d}x) \le \varphi'(\bar{x})\int_{A_1}(\bar{Y}_x - \bar{Y})F_X(\mathrm{d}x),\\
&(2)\ \textstyle\int_{A_2}(\bar{Y}_x - \bar{Y})\varphi'(x)F_X(\mathrm{d}x) \le \varphi'(\bar{x})\int_{A_2}(\bar{Y}_x - \bar{Y})F_X(\mathrm{d}x),\\
&(3)\ \eta'_D(0) = \textstyle\int_{A_1}(\bar{Y}_x - \bar{Y})\varphi'(x)F_X(\mathrm{d}x) + \int_{A_2}(\bar{Y}_x - \bar{Y})\varphi'(x)F_X(\mathrm{d}x)\\
&\qquad \le \varphi'(\bar{x})\textstyle\int_{A_1\cup A_2}(\bar{Y}_x - \bar{Y})F_X(\mathrm{d}x) = 0.
\end{aligned}$$

If $\bar{Y}_x$ is strictly increasing, the strict inequality must hold in (1) or (2) and therefore in (3). Adjustments for the nonincreasing cases are obvious.

Thus if sign $\bar{Y}$ is known, sign $\eta'_G(0)$ is known (4.29) and if, in addition, $\bar{Y}_x$ is monotonic of known direction, then sign $\eta'_D(0)$ is known. If these signs agree, sign $\eta'(0)$ is known (4.34). Otherwise approximations of magnitude are needed. Hopefully, satisfactory approximations can be developed as we learn more about utilities, probabilities and the array of contingencies faced by people whose behavior we hope to understand.

Of major interest in the analysis of any static model are the equations of comparative statics expressing the response of an individual or market to changes in circumstances. Equations indicating marginal responses to changes in wealth and to price of a venture are briefly noted below. Responses to changes in beliefs (probabilities) and to other changes in ventures are important but beyond the scope of this paper.

Write $X = \bar{X} + U$, $Y = Z - h$. A cash gift changes $\bar{X}$ but not other data of the decision problem, so adjustment to a change in $\bar{X}$ will be called a wealth response. h is a scalar which may be regarded as the price of the venture Z. Of course, if $Z^* = Z + k$ then purchasing Z^* for $h + k$ is the same venture as purchasing Z for h, so an initial Z can be changed by an additive constant at the convenience of the investigator.

If U and Z are regarded as fixed, then

$$\begin{aligned}
\eta(\alpha, \bar{X}, h) &= E\varphi(\bar{X} + U + \alpha Z - \alpha h) - E\varphi(\bar{X} + U)\\
&= E\varphi(X + \alpha Y) - E\varphi(X). \qquad (4.35)
\end{aligned}$$

We shall mostly use the latter notation letting the decompositions of X, Y be understood.

$$\begin{aligned}
\partial\eta/\partial\alpha &= EY\varphi'(X + \alpha Y) = EZ\varphi'(X + \alpha Y) - hE\varphi'(X + \alpha Y),\\
\partial^2\eta/\partial\alpha^2 &= EY^2\varphi''(X + \alpha Y) = -\Delta \quad \text{where } \varphi'' < 0 \Rightarrow \Delta > 0,\\
\partial^2\eta/\partial\alpha\partial\bar{X} &= EY\varphi''(X + \alpha Y),\\
\partial^2\eta/\partial\alpha\partial h &= -\alpha EY\varphi''(X + \alpha Y) - E\varphi'(X + \alpha Y) = -E\varphi'(X + \alpha Y)\\
&\quad - \alpha(\partial^2\eta/\partial\alpha\partial\bar{X}). \qquad (4.36)
\end{aligned}$$

Letting $\hat{\alpha}(\overline{X}, h)$ be the solution to $\partial\eta/\partial\alpha = 0$, the wealth and price responses are

$$\frac{\partial\hat{\alpha}}{\partial\overline{X}} = -\frac{\partial^2\eta}{\partial\alpha\partial\overline{X}}\bigg/\frac{\partial^2\eta}{\partial\alpha^2} = \frac{1}{\Delta}EY\varphi''(X + \hat{\alpha}Y)$$

and

$$\frac{\partial\hat{\alpha}}{\partial h} = -\frac{\partial^2\eta}{\partial\alpha\partial h}\bigg/\frac{\partial^2\eta}{\partial\alpha^2} = -\frac{1}{\Delta}E\varphi'(X + \hat{\alpha}Y) - \hat{\alpha}\frac{\partial\hat{\alpha}}{\partial\overline{X}}. \tag{4.37}$$

As with Slutsky equations for other models, it is natural to interpret the second term of the price response as a wealth effect and the negativity of the first term seems plausible.

4.5. Bets

For an event A with $0 < PA < 1$, call the indicator I_A of A a unit claim on A. The venture $I_A - h_A$ is a purchase of a unit claim on A at price h_A. The venture $\alpha(I_A - h_A)$ is a purchase of α unit claims (or an α-claim) on A at price h_A.

For A as above, B the complement of A, $y > 0$; call $I_A - yI_B$ a unit bet on A. y will be called the market odds on A and PA/PB the subjective odds. If market odds and subjective odds are equal, expected gain from the venture is zero. Clearly a unit bet on A is the same venture as the purchase of an α-claim on A with $\alpha = 1 + y$ and price $h_A = y/(1 + y)$ so a discussion of such ventures can be conducted in terms of either claims or bets. We shall proceed with the latter.

The utility of the venture αY where Y is a unit bet on A is

$$\eta(\alpha) = E\varphi(X + \alpha Y) - E\varphi(X) = \textstyle\int_A\varphi(X + \alpha) + \int_B\varphi(X - \alpha y) - \int\varphi(X) \tag{4.38}$$

and

$$\begin{aligned}\eta'(\alpha) &= \textstyle\int_A\varphi'(X + \alpha) - y\int_B\varphi'(X - \alpha y) \\ &= (PA)E_A\varphi'(X + \alpha) - y(PB)E_B\varphi'(X - \alpha y),\end{aligned} \tag{4.39}$$

where, for any non-null event C and any integrable random variable Z,

$$E_C Z = \frac{1}{PC}\int_C Z.$$

Thus $E_C Z$ is the average value of Z on C. Note that αY for $\alpha < 0$ is a bet on B.

Propositions 4.1 and 4.2 illustrate how a number of facts about favorable regions for α and favorable exchanges are related to the sign of $\eta'(0)$. For a family of bets,

$$\eta'(0) = (PA)E_A\varphi'(X) - y(PB)E_B\varphi'(X). \tag{4.40}$$

Thus

$$\eta'(0) \gtreqless 0 \Leftrightarrow \frac{PA}{PB}\cdot\frac{E_A\varphi'(X)}{E_B\varphi'(X)} \gtreqless y. \tag{4.41}$$

$E_A\varphi'(X)$ is the average marginal utility of a gain in A. $(PA)E_A\varphi'(X)$ is the initial marginal utility of a unit claim on A. Call $E_A\varphi'(X)/E_B\varphi'(X)$ the relative need in A and $PA/PB \cdot E_A\varphi'(X)/E_B\varphi'(X)$ the critical odds[7].

If X is independent of A, the relative need in A is unity and a risk averter will prefer a positive ($\alpha > 0$) bet on A if the subjective odds PA/PB exceed the market odds y. With dependence, the relative need in A will tend to exceed unity if X tends to be lower on A. Two instances in which we can give more exact expressions of this tendency are furnished by Hanoch and Levy's results on stochastic dominance.

PROPOSITION 4.5. *If* $\varphi'' < 0$, *then*

(i) $F_{X|A} \geq F_{X|B}$ *with strict inequality for some* $x \Rightarrow E_A\varphi'(X) > E_B\varphi'(X)$.

(ii) $\varphi''' > 0$ *and*

$$\int_{-\infty}^{x} F_{X|A} \geq \int_{-\infty}^{x} F_{X|B}$$

for all $x \in R$ *with strict inequality for some* $x \Rightarrow E_A\varphi'(X) > E_B\varphi'(X)$.

---

[7] The factors determining critical odds are closely related to the decomposition introduced in preceding sections. For a family of bets on A, $\eta_G'(0) = (PA - yPB)E\varphi'(X)$ and is thus positive if subjective odds are greater than market odds.

$$\eta_D'(0) = (1 + y)(PA)(PB)(E_A\varphi'(X) - E_B\varphi'(X))$$

and is positive if the relative need in A exceeds unity.

PROOF. (i) follows from theorem 1 of Hanoch and Levy [11, p. 337] and (ii) from theorem 2 [11, p. 338]. See also Hadar and Russell [10].

For two individuals, proposition 4.1 and eq. (4.41) justify

PROPOSITION 4.6. *If the critical odds of two decision makers for bets on an event A differ, there exist mutually favorable exchanges of bets on A.*

PROOF. Suppose

$$\frac{PA}{PB}\cdot\frac{E_A\varphi'(X)}{E_B\varphi'(X)} > \frac{P^*A}{P^*B}\cdot\frac{E_A^*\varphi^{*\prime}(X^*)}{E_B^*\varphi^{*\prime}(X^*)}$$

where starred symbols pertain to the second decision maker. Choose y strictly between the two critical odds. By (4.41) $\eta'(0) > 0$, $\eta^{*\prime}(0) < 0$ so by proposition 4.1 there exists an interval $(0, \delta) \ni \alpha \in (0, \delta) \Rightarrow \alpha(I_A - yI_B)$ is favorable for the first decision maker and $-\alpha(I_A - yI_B)$ is favorable for the second.

For bets that are independent of the current prospect, the specialization of proposition 4.6 has been discussed by Smith [15] and by Aumann [4].

4.6. *To Bet or Not to Bet*

In the present model two people may be led to exchange bets on A if one has greater relative need in A and/or believes A is more probable. Ordinary conversations suggest that different opinions on future events are common[8] and it is not hard to think of people who must surely

---

[8] As a hurried and inexpensive check on divergent opinions I handed a short questionnaire to early arrivals at an economics seminar on April 24, 1972. Responses to the question:

'What are your personal probabilities of the following events?

(*a*) Richard Nixon will be reelected president next November.
(*b*) Edward Kennedy will be the nominee of the Democratic party.
(*c*) The Balance of Trade deficit of the United States, for the fiscal 1972, will exceed five billion dollars.
(*d*) North Vietnam, the Viet Cong, the Saigon government, and the United States will agree to a permanent cease-fire before August 1, 1972.
(*e*) The legislative maintenance appropriation to the University of Minnesota for

have widely different relative needs in particular events. With the multitude of possible people–event combinations in any large community, it would seem at first glance that there must be many potential mutually favorable bets.

Why is more betting by the general public (as opposed to habitual gamblers) not observed? Is there a defect or incompleteness in the theory? Lack of knowledge? Social or institutional barriers? Transactions costs? Various possibilities have been noted (see, for example, refs. [4] and [15]) and are briefly discussed below, not with the idea of suggesting a conclusion but to suggest that this is a worthwhile area for research. Knowing with confidence what does explain this divergence between an implication of an admittedly crude theory and common observation should improve our grasp of decision making under uncertainty.

4.6.1. Transactions costs

Although it costs no money for two private parties to agree to a bet, it does take some time and requires some internal calculating. Substantial negotiations may be required if the basic event is complicated and it may cost something to observe. Guaranteeing performance by each party may be a problem and might involve putting up stakes thus forgoing cash balances.

1973–75 will be less than $150 million (comparable figure for 1971–73 is $162 million).'

were

	Respondent number											
Question	1	2	3	4	5	6	7	8	9	10	11	Range
a	0.75	0.35	0.6	0.95	0.48	0.75	0.55	0.6	0.8	0.95	0.4	0.35–0.95
b	0.10	0.2	0.2	0.3	0.35	0.20	0.25	0.2	0.3	0.10	0.08	0.08–0.35
c	0.50	0.1	0.2	0.7	0.55	0.50	0.10	0.3	0.5	1.0	0.15	0.1 –1.0
d	0.10	0.05	0.1	0.2	0.25	0.30	0	0.1	0.1	0	0.5	0 –0.5
e	0.05	0.1	0.1	0.1	0.40	0.05	0	0.1	0.3	0	0.01	0 –0.4

4.6.2. Moral considerations

Some people's training includes classifying gambling as evil. For others, the reflection that goods and services as usually conceived are not increased by betting may suggest that prospective gains, at least in the long run, are illusory. Time spent in arranging bets may therefore be regarded as socially wasteful. Some may also feel that in nearly every bet one party is better informed and taking advantage of the other party. Someone who might not have personal moral objections to betting may still feel that others will disapprove, or that winning will be accompanied by resentment on the loser's part.

4.6.3. 'Shaky' prior distributions

If I consider betting with a well informed person, his willingness to offer what may at first seem very attractive odds may give me second thoughts. My subjective probability of the event conditional on his offer may be much lower than my original subjective probability of the event. I may usually avoid betting with well informed people on the ground that the chances that my subjective probability given the offer will differ from the market odds sufficiently to overcome transaction costs are small.

If my subjective probability of the betting event is 'firm', i.e. largely independent of occurrence or nonoccurrence of other events about which I might learn before the decision is to be made, I might proceed despite the different subjective probability of a betting partner whose opinions I respect. To be in this position could imply that I feel I have as good information as anyone and have analyzed it reasonably well.

In analogous business situations, this consideration would seem to place a premium on investing in one's own business or at least in familiar areas.

Instances of business behavior analogous to revising one's probability on learning of an informed person's willingness to make an opposing bet might be second thoughts on locating a plant after learning a respected competitor has rejected the location under consideration or reconsidering a purchase if the owner seems overanxious to sell.

Daniel McFadden has called my attention to the fact that this is also similar to Akerlof's 'lemon principle' [1]. The fact that a car is offered on a used car lot is itself evidence of mechanical difficulty.

4.6.4. Disutility of losing

It seems possible that, for many people, losing a bet has disutility beyond whatever assets have to be paid to the winner. This is probably particularly true if the fact of losing will become known. Perhaps, for some, there is a counterbalancing direct utility of winning. It would be good to know something of these matters. Smith [15] and *Newsweek* magazine [13] both invoked direct utility of being in a gambling situation as a major explanation of the behavior of habitual gamblers. Perhaps there is a counterpart to this for habitual nongamblers.

4.6.5. Better future or alternative opportunities

A bet might look favorable in terms of a decision maker's current commitments, but be rejected because he visualizes the future possibility of alternatives (for example, investments, loans, increased savings or a bet at more favorable odds) that look better. This raises complicated questions about the precise definition of one's current prospect. It would seem that possible future opportunities must be considered as part of the current prospect as long as the decision maker assigns them sufficient personal probability that they may influence his current decisions. Thus contingencies that involve no contracts, negotiations or stated plans may still be important parts of the current prospect.

Note that in an exchange economy in which people could costlessly bet on any event, had complete information, and could avoid inhibiting disputes over division of increments of expected utility (possibly by organizing markets), one would expect exchanges to take place until, for any nontrivial event (everyone agrees $0 < PA < 1$), everyone would have the same critical odds. Such stability would also be a necessary condition for a Pareto optimum based on expected utility.

To reach the stability indicated above it would not be necessary to exchange bets if the economy offered sufficient alternative ways for each individual to transfer wealth among events[9]. If the alternative ways were superior in some way one would expect, upon investigation, to find little or no betting, but few if any potential mutually favorable bets.

Thus, searching for opportunities for mutually favorable bets in an

---

[9] Arrow [3, chapter 4] has analyzed a dynamic model of exchange in which elimination of some markets does not affect the achievable sets for individuals nor the equilibrium position.

actual economy may be interpreted as looking for imperfections in opportunities to change the way in which one's wealth depends on environmental contingencies. The following are hypothetical examples of kinds of situations in which intelligent betting might seem to have some merit. As noted in earlier discussion, if there is a favorable betting possibility, it does not necessarily follow that exchanging bets is the best way to improve the prospects of the persons affected. Other ways of altering prospects may be suggested by studying the circumstances, and these may have advantages. Possibilities of this sort are noted in examples III and IV below.

4.7. Example I: a Specialty Crop

Suppose a hypothetical fruit – call it a plone – is grown in a small geographic area and is very sensitive to rain during the ten-day harvest period, a measurable rain almost completely ruining the crop. As harvest approaches, growers are in a precarious position, having used most of their liquid assets to plant and cultivate. Plones can be stored at moderate cost and each year there are warehouses with one-fourth to one-half of a crop in storage as harvest approaches. When the dates of harvest can safely be forecast, suppose the growers bet on rain and the warehousemen bet against rain during the harvest period. If it rains, the growers will lose their crop but collect their bets. The warehousemen will pay out of the increased price of plones in storage. If no rain, the growers will pay out of the crop and the warehousemen will recover storage costs by winning bets.

If growers sought security through the usual crop insurance, someone would have to verify the preharvest condition of the crop, and the company and growers would have to negotiate their estimates of any losses incurred. This would add greatly to the cost. A bet should be much less expensive. All that would have to be verified would be rain or no rain at a weather station in the growing region. If the region were small enough, rain at the weather station should be a good indicator of probable damage. A grower who lost his crop even though there were no rain at the station would lose doubly, crop and bet, but if this compound event has sufficiently small probability, covering it may not be worth the loading added to an insurance premium.

4.8. Example II: a Merchant

A merchant has put most of his liquid assets into inventories of seasonal items. If retail sales in his region hold reasonably well, he is almost sure he can sell enough to avoid severe financial strain. On the other hand, if there is a regional slump, there isn't much he can do. Suppose he bets that retail sales in his area will be less than 80% of normal. This clearly reduces the uncertainty in his prospect.

There may not be a natural second party to take this bet at actuarial value, but, assuming a regional slump is his main hazard, he might well be able to afford to offer sufficiently good market odds that people who hold conservative securities (and thus have current prospects largely independent of regional sales) would be induced to risk some of their funds.

If the merchant tried to insure his own sales, both an incentive problem and a problem of verification would immediately arise. The merchant could make his coverage more precise by placing several bets – one that regional sales would be 80% or less, one that they would be 70% or less, etc.

4.9. Example III: a Patient

A handicapped person contemplates an operation which, if successful, will greatly increase his earning power. If it fails, his physical circumstances are about as before. Suppose he bets the operation fails. If he loses the bet, he pays out of increased earnings. If he wins the bet and it covers the cost of the operation, he has not lost financially. The doctor and the hospital might be good prospects to take his bet. Any effect on their incentives would certainly be in the right direction. The same final prospect could be achieved if the doctor and/or hospital charged more for successful operations.

4.10. Example IV: Fixed Money Income

Someone on social security or other fixed money income might bet that the consumer price index will rise at least x% in the next ten years.

Although the bet might be better than doing nothing, a person who can shift his assets might be able to do better by other devices. If there are some for whom some combination of bets is the best alternative, the government might be a good second party, especially if the government claims to be pursuing effective price stabilization.

The first two examples suggest that insurance may sometimes be obtained more economically by making payment conditional on a general easily observed event rather than a special event that relates directly to the personal affairs of the insured. No-fault insurance seems to substitute a somewhat more general event, occurrence of specified damage in a vehicular accident, for a still more specific event that also includes specification of legal liability.

James C. Hickman has suggested that pools created by companies to compensate a firm experiencing a strike (and similar pools among labor unions) are essentially combinations of bets of the type illustrated, and that formal reinsurance arrangements[10] have similar qualities.

Thus the theory does not suggest a new phenomenon, but the general theoretical conditions under which favorable exchanges exist suggest further study to see if there are ways to help people find unexploited opportunities to advantageously rearrange their prospects.

References

[1] Akerlof, G. A. The market for 'lemons': quantitative uncertainty and the market mechanism. *Quarterly Journal of Economics*, **84** (1970), 488–500.

[2] Apostol, T. M. *Mathematical Analysis*. Addison-Wesley: Reading, Mass. (1957).

[3] Arrow, K. J. *Essays in the Theory of Risk-Bearing*. Markham: Chicago (1971).

[4] Aumann, R. J. *Subjectively mixed strategies*. Technical Report No. 45, Institute for Mathematical Studies in the Social Sciences, Stanford University (August 6, 1971).

[5] Bohman, G. A mutual reinsurance scheme. *The ASTIN Bulletin*, **6**, Part 2 (December 1971).

[6] Borch, K. Reciprocal reinsurance treaties. *The ASTIN Bulletin*, **1**, Part 4 (December 1960).

[7] Borch, K. Reciprocal reinsurance treaties seen as a two person cooperative game. *Skandinavaisk Aktuarietidshift* (1960).

[8] Collins, R. A non-proportional (spread loss) reinsurance pool. *Transactions 19th International Congress of Actuaries* (1972).

[9] Cramér, H. *Mathematical Methods of Statistics*. Princeton University Press: Princeton, N. J. (1946).

[10] See refs. [5]–[7] and [8].

[10] Hadar, J. and Russell, W. R. Rules for ordering uncertain prospects. *Journal of Economic Theory,* **3** (1971), 288–305.
[11] Hanoch, G. and Levy, H. Efficiency analysis of choices involving risk. *Review of Economic Studies,* **36** (1969), 335–346.
[12] McShane, E. J. *Integration.* Princeton University Press: Princeton, N. J. (1944).
[13] *Newsweek* (April 10, 1972), 46.
[14] Pratt, J. W., Raiffa, H. and Schlaifer, R. *Introduction to Statistical Decision Theory.* McGraw-Hill, New York (1965).
[15] Smith, V. L. Economic theory of wager markets. *Western Economic Journal,* **9** (1971), 242–255.

COMMENTS

On 'Ventures, bets and initial prospects'

James C. Hickman

Professor Hildreth's key idea, like most good ideas, is very appealing once it is stated. In this paper he has developed the notion that decisions about embarking on uncertain new ventures may be viewed as involving a comparison between the expected utility of current uncertain prospects and the expected utility of these prospects modified by the decision to enter the new venture. He has used this model, which differs from the conventional model which usually involves the assumption that the current prospects are certain, to arrive at an interesting decomposition of the decision analysis and some fresh insights into the conditions under which mutually advantageous bets exist. Professor Hildreth is probably correct in stating that this paper raises more questions than it settles. Yet one cannot go away from reading this paper without a feeling that his model is more useful than the conventional one.

The paper seems to culminate in proposition 4.6. In this proposition we see a marvelous summarizing criterion for the existence of a mutually advantageous bet. The criterion captures key aspects of the current prospects and the probability and utility assessments of the two protagonists in the bet.

A general discussion of this paper rather naturally will center on proposition 4.6, for it seems to indicate that in the real world there exist a vast number of potentially mutually advantageous betting situations. In section 4.6 Professor Hildreth also turns to the fascinating question as to why the apparently vast opportunities for mutually advantageous bets are not exploited.

In joining this discussion, I would like to suggest that observations 4.6.1 (transaction costs), 4.6.2 (moral considerations) and 4.6.3 ('shaky' prior distributions) on the reasons for a relative lack of betting are inexorably intertwined. Transaction costs are often significant because the cost of the information needed to formulate a coherent probability distribution is not negligible. The classification of betting as evil has a pragmatic basis that is also related to the high cost of information. Moral, political and business leaders, concerned with the efficient organization of society, long ago perceived the relative high cost of obtaining the information necessary to formulate coherent probability distributions. Crudely stated, the time spent building a relative frequency table on the behavior of a 'wheel of fortune' at the county fair is time spent away from weeding the corn.

In our economy we need to continually strive to identify those areas where the possibility of mutually advantageous bets is great enough to justify the creation of an information system. In the world of risk and insurance this is called the problem of defining an insurable risk. The operational definition is continually changing with the nature of the economic uncertainty that the society and natural environment create and the cost of insurance statistical and rating systems.

In the world of finance we have developed highly organized securities markets to facilitate the marshalling and allocation of capital and to make possible mutually advantageous bets. For many of these markets the federal government, through such agencies as the SEC, attempts to create a common base of information for the formulation of priors. In addition, the accounting profession, with its elaborate set of principles, rules and opinions, has taken as its primary obligation the creation of 'comparable' financial data to facilitate the assessment of the personal probabilities of investors.

The agricultural commodity markets in the United States depend on the statistical systems of the Department of Agriculture for the flow of information on which the shifting probability assessments concerning supply and demand are made. Comparison with other countries seems to indicate that an ongoing market, involving mutually advantageous bets on the prices of agricultural products, is hard to organize without a flow of relatively cheap and reliable information. The individual grain speculator could not duplicate the US Crop Reporting Service.

The existence of television rating and credit rating bureaus indicates that not all information systems needed to make firm probability assignments are public ventures. The changing nature of economic uncertainty periodically creates an opportunity for a statistical entrepreneur to fill a need by creating an information system.

Buried into the criterion of proposition 4.6 is an alpha (scale) parameter for each of the participants in the bet. These scale parameters may be very small. If a bet is mutually advantageous only over very small scale parameters, the expected gain may not match the inevitable transaction and information costs.

Let us illustrate this matching difficulty with a real problem. Owners with property astride the San Andreas fault live with a small probability of large property losses. Similarly, those who own property near the coast of the Gulf of Mexico also live with the threat of large property losses. Recently the probability of losses due to hurricanes in the Gulf has appeared to be higher than that of earthquake losses in California. Property owners in the upper Midwest occasionally suffer losses from floods caused by rapid snow run-off and unseasonably early spring rains. It would seem that, since these three uncertain and unfortunate events are reasonably independent, some sort of mutually advantageous betting arrangement (insurance pool) might be organized. Ideally, such a pool should not encourage uneconomic ventures such as major construction projects on the sand dunes along the Gulf Coast. Yet appealing as the idea is, the matching of the bets is hard to arrange. Large bets are needed if a real economic service is to be provided. However, markedly different probabilities of loss are involved and some of the probabilities are a bit 'shaky'. The details of negotiating mutually advantageous bets for such catastrophes on a scale to truly stabilize results would seem to be an enormously important but profoundly difficult undertaking.

On some facets of betting

Daniel McFadden

C4.2.1. Introduction

A standard proposition of the theory of choice under uncertainty is that two individuals whose personal probabilities of a future event differ can make a mutually advantageous wager. On the other hand, empirical observation suggests that widespread betting is absent on events where individuals' personal probabilities apparently differ widely. Professor Hildreth's interesting paper suggests several possible explanations of this inconsistency. Extending this analysis, we look for answers in the nature of beliefs, the structure of markets for wagers and the impact of market form on beliefs. In each case, we must ask whether the postulated phenomenon is likely to be prevalent in reality, and whether it is sufficient to imply the observed paucity of wagers.

C4.2.2. The nature of beliefs

Professor Hildreth has suggested that when individuals consider wagers against the background of the 'grand lottery of life', they may not view as independent the events determining the outcomes of the 'grand' lottery and the wager. We first ask whether it is likely that personal probabilities would tend to display this non-independence; in particular, more likely than 'objective' probabilities determined by relative frequencies. An examination of human psychology suggests an affirmative answer. Chance jolts the harmony of conscious belief; relief from this dissonance is gained by imposing an order over chaos, weaving a fabric of cause and effect, out of the jumbled coincidences of random events.

> It is so much easier to assume than to prove; it is so much less painful to believe than to doubt; there is such a charm in the repose of prejudice, when no discordant voice jars upon the harmony of beliefs. . . .
>
> W. E. H. Lecky, *A History of Rationalism* (1900).

> Nothing is so easy as to deceive one's self; for what we wish, we readily believe.
>
> Demosthenes, *Third Olynthiac* (348 B.C.).

The mind accepts and emphasizes those coincidences which reaffirm the perceived order of the universe, ignores and forgets inconsistent data.

L. Festinger (*A Theory of Cognitive Dissonance*, pp. 162–176) has carried out a study of subjects given an opportunity to accept a series of wagers involving a complex random event, and has examined the willingness of the subjects in the course of play to accept information dissonant with their beliefs about the random event. He concludes that 'the interaction between the amount of dissonance which exists and the expectation concerning some particular source of new information in determining whether or not a person will expose himself to, or avoid, this source of information is [clearly consistent with the theory of dissonance reduction]'. In an experimental test of the von Neumann–Morgenstern axioms, D. Davidson and P. Suppes (*Decision Making*, p. 53) report that 'Winning or losing several times in a row made subjects sanguine or pessimistic and tended to produce altered responses to the same offers'.

Thus, the evidence is persuasive that personal probabilities will tend to distort the independence properties of 'objective' probabilities, implying correlations between events which are in fact independent. A simple model of personal probability determination with selective memory gives a final illustration of this point. Suppose two events E_1 and E_2 yield favorable outcomes to the individual in repetitive play, and that these events are in fact independent, each occurring with probability one-half. Suppose the individual computes personal probabilities from observed relative frequencies, remembering coincidences of favorable or unfavorable outcomes perfectly but forgetting a proportion θ of the observations when a coincidence does not occur. Then the probability limit of the individual's personal probability of the joint event (E_1, E_2) as the number of repetitions goes to infinity equals $1/(4 - 2\theta)$, greater than the objective probability $\frac{1}{4}$. Note in this example that we not only have non-independence, but also that the personal correlation between events corresponds to the individual view that luck occurs in runs, so that favorable results tend to go together. This last observation has some further implications, which we shall return to later.

We next ask whether pervasiveness of non-independence in personal probabilities of events is itself sufficient to explain the paucity of wagers. We employ Professor Hildreth's notation, and for concreteness assume further that the underlying Bernoulli utility indicator exhibits constant risk aversion and that the personal probability for the current prospect

X and new venture Y is multivariate normal with means (μ_X, μ_Y) and covariance matrix

$$\begin{bmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

Then, the expected utility of the current prospect plus a share s in the new venture is a monotone increasing function of

$$\lambda(s) = K + (\mu_Y - q + \alpha\rho_{XY}\sigma_X\sigma_Y)s - \alpha\sigma_Y^2 s^2/2, \qquad \text{(C4.1)}$$

where α is the degree of absolute risk aversion, $K = \mu_X - \alpha\sigma_X^2/2$, and q is the price of a unit share in the new venture (assumed zero by Professor Hildreth).

Even if the new venture is actuarially favorable $(\mu_Y - q > 0)$, the individual believing $\rho_{XY} > 0$ may find it undesirable to acquire a positive share; only if

$$\mu_Y - \alpha\rho_{XY}\sigma_X\sigma_Y - q > 0 \qquad \text{(C4.2)}$$

will a positive share be chosen. This argument would seem to support Professor Hildreth's conclusions. Note, however, that two individuals A and B with the preference structure above and differing personal probabilities satisfying

$$\mu_Y^A - \alpha^A\rho_{XY}^A\sigma_X^A\sigma_Y^A > \mu_Y^B - \alpha^B\rho_{XY}^B\sigma_X^B\sigma_Y^B \qquad \text{(C4.3)}$$

can find a price q between these quantities at which it is mutually advantageous for B to sell a share of the new venture to A. Thus, we see that non-independence alone is not sufficient to rule out widespread betting.

The psychological argument we made earlier implied more than non-independence of personal probabilities, however; it implied an 'irrational' belief that the probabilities of events depend on the desirability of outcomes, with 'luck' running in 'streaks'. One might incorporate this phenomenon into the example above by postulating that the parameters of the personal probability distribution $(\mu_Y, \sigma_Y, \rho_{XY})$ depend on the individual's decision variable, the net share purchase s. If, in particular, $\rho_{XY} > 0$ when $s > 0$ and $\rho_{XY} < 0$ when $s < 0$, implying the outcome of the wager is likely to be good when the outcome of the current prospect is good, and *vice versa*, no matter which way the wager is laid, then individuals with differing personal probabilities may find no grounds for a mutually advantageous wager. This explanation is of course inconsistent with the Savage axioms, and seems to have the same behavioral implications as a pure 'distaste for gambling', the mirror image of the phenomenon claimed by V. Smith to be necessary to explain compulsive gambling. It is worth noting that the effect we have postulated is 'rational' in the sense that it can result from rational preference maximization over acts, and in the sense that one cannot engage the individual in a series of wagers that would result in his taking a sure loss.

C4.2.3. The structure of markets for wagers

The paucity of wager markets could result from the presence of high organizational costs, transactions costs or redundancy. It is plausible that the costs of searching for potential traders and enforcing contracts, particularly time costs, are a significant deterrent to the placing of small wagers. Since risk aversion lowers the desirability of large wagers where transactions costs are relatively unimportant, the combination of effects may be sufficient to explain the lack of markets. A second possible explanation is that most wager markets are in fact redundant; the individual can achieve any desired risk position through the operation of a few well-organized markets such as securities markets. This phenomenon has been noted in papers by K. Arrow and by P. Diamond on the allocation of risk-bearing showing that generally a system with N commodities and S states of nature needs only $(N - 1)(S - 1)$ markets, of which $S - 1$ are wager markets, instead of the maximum possible number of barter markets, $NS(NS - 1)/2$, including $N^2S(S - 1)/2$ wager markets. In a study of the existence of equilibrium under uncertainty, R. Radner points out that wagers can be made only on information that will be common to the participants, reducing further the number of wager markets that can form. We conclude that the absence of widespread wager markets may be the result of redundancy or transactions costs rather than individual aversion to betting.

C4.2.4. Market effects on beliefs

Thus far, we have considered only the possibility that beliefs are affected by actions via a psychological mechanism of selective memory. There is the additional possibility that the events on which an individual might wager could be affected by the actions of his opponent; the problem of *moral risk*. The presence of such an effect will introduce a dependence of the expected payoff of a new venture on the position held by the oponent. This can have the effect of eliminating the possibility of a mutually advantageous wager; the argument is the same as in the paragraphs following.

We next examine the role of the market itself in providing information and influencing beliefs. G. Akerlof has pointed out the *lemon principle*, which states that in the presence of uncertainty about the quality of a commodity unit, the fact that it is offered in a market may be taken as information on its quality. The usual example of operation of this principle is in the used car market, where the fact that a vehicle is in the market suggests that it may be below average quality for vehicles of the same identifiable type; i.e. the seller of the vehicle may have information, withheld from the potential buyer, that the car is a 'lemon'. Applied to a market for wagers, this principle suggests that a potential buyer of a lottery ticket may suspect that the seller holds inside information unavailable to himself which indicates the yield of the ticket will be low, and takes the fact that the ticket is being offered in the market as evidence supporting this suspicion. Symmetrically, a potential seller may suspect that an individual soliciting a wager has inside information. It is clear that the presence of such suspicions will inhibit the trading of wagers. In terms of eq. (C4.1) expressing the desirability of a net share s in the new venture, the expected return μ_Y will be considered a function of s, with $\mu_Y(1) < \mu_Y(-1)$ for the reason above. If, by contrast to eq. (C4.3),

$$\mu_Y^A(1) - \alpha^A \rho_{XY}^A \sigma_X^A \sigma_Y^A < \mu_Y^B(-1) - \alpha^B \rho_{XY}^B \sigma_X^B \sigma_Y^B,$$
$$\mu_Y^B(1) - \alpha^B \rho_{XY}^B \sigma_X^B \sigma_Y^B < \mu_Y^A(-1) - \alpha^A \rho_{XY}^A \sigma_X^A \sigma_Y^A,$$

then no mutually advantageous wager is possible.

The same psychological phenomenon as discussed earlier may tend to reinforce belief in the lemon principle. Actions leading to losses suggest bad judgment, generating dissonance, which can be reduced psychologically by attributing 'inside information' or 'unfair advantage' to the opponent. Further, there is a 'Gresham's law' aspect to the lemon principle; one can show that it will tend to drive out of the market a disproportionate number of 'honest' lottery tickets. Thus, the lemon principle becomes a 'self-fulfilling prophecy'. A strong argument can hence be made that the lemon principle will operate to inhibit many wager markets.

C4.2.5. Summary

This discussion has pointed out the following possible explanations for the paucity of wager markets in the presence of differing personal probabilities:

(1) The psychology of cognitive processes suggests that individuals will tend to believe in 'runs of luck', with an effect similar to that caused by a pure 'distaste for gambling'.

(2) Transactions costs, of consequence for small transactions, combined with the inhibiting effect of risk aversion on large transactions, may prevent wager markets from forming.

(3) Many wager markets may fail to form because they are redundant.

(4) Moral risks may inhibit the offering of wagers on some events.

(5) The lemon principle may operate to indicate the presence of inside information to be used to the disadvantage of potential traders, inhibiting formation of a market.

CHAPTER 5

STOCHASTIC DOMINANCE IN CHOICE UNDER UNCERTAINTY

Josef Hadar and William Russell

5.1. Introduction

When a decision maker finds himself in an uncertain environment which he is able and willing to view in probabilistic terms, then making a decision means choosing the preferred probability distribution from among those that are feasible[1]. Recently, it has been shown that, when decision makers follow the rules of expected utility maximization, equivalence theorems of the following type hold:

For any two uncertain prospects (probability distributions) P and P', prospect P stochastically dominates P' if and only if the expected utility under P is at least as high as under P' for all individuals (utility functions) in some appropriate set.

The exact meaning of 'stochastic dominance', and a more precise formulation of the theorems in question, will be given below. What we wish to point out in these introductory remarks is the following. When one is confronted with the problem of determining which of two uncertain prospects a typical decision maker will choose, one may, by virtue of the above equivalence theorems, take one of two approaches: (a) one may try to determine which of the two prospects yields a higher expected utility, or (b) one may try to establish the existence of stochastic dominance between the two prospects. Of course, if the utility function

[1] In some cases, the set of all feasible distributions may consist of a family of distributions generated by a parameter which is under the control of the decision maker. In that case, the decision maker's choice takes the form of setting the parameter equal to that value which is associated with the preferred distribution.

of the decision maker is specified only in general terms (such as monotonicity and concavity), the problem may not have a determinate answer. However, if it does yield an unambiguous result by one approach, then it should in principle be possible to prove the same result by means of the other approach as well. In practice, the choice of approach is likely to be governed by its tractability in the context of the specific problem under investigation. But aside from such factors as analytical facility, it should be pointed out that the stochastic dominance approach is distinct in that it directly points to certain inherent properties of the prospects under consideration, and thereby it exposes certain features of the nature of preferences for uncertain prospects. In some sense, the stochastic dominance approach tells us what it is that makes the chosen prospect preferred over the others.

In this paper we do two things. First, we present some basic theorems and properties of distributions satisfying stochastic dominance. These should be useful in many applications of this approach to problems under uncertainty. Second, we present specific problems of choice under uncertainty, and show how the stochastic dominance approach may produce interesting theorems.

5.2. *Basic Definitions and Properties*

We start off by defining three types of stochastic dominance: first, second and third-degree stochastic dominance, denoted respectively by FSD, SSD and TSD. In what follows, R is the common domain of the functions under consideration; unless stated otherwise, it is assumed to be unbounded. The distribution functions are denoted by capital Roman letters.

For notational convenience, we define the following integrals: if $\mathrm{F}(x)$ is a distribution function, then

$$\hat{\mathrm{F}}(x) = \int_{-\infty}^{x} \mathrm{F}(t)\mathrm{d}t,$$

and

$$\hat{\hat{\mathrm{F}}}(x) = \int_{-\infty}^{x} \hat{\mathrm{F}}(t)\mathrm{d}t.$$

DEFINITION 5.1.

(a) $G(x) \underset{\sim}{\textcircled{F}} F(x)$ if and only if $G(x) \le F(x)$ for all $x \in R$.

(b) $G(x) \underset{\sim}{\textcircled{S}} F(x)$ if and only if $\hat{G}(x) \le \hat{F}(x)$ for all $x \in R$.

(c) $G(x) \underset{\sim}{\textcircled{T}} F(x)$ if and only if $\hat{G}(\infty) \le \hat{F}(\infty)$, and $\hat{\hat{G}}(x) \le \hat{\hat{F}}(x)$ for all $x \in R$.

In the above, the symbol $\underset{\sim}{\textcircled{F}}$ means 'is at least as large in the sense of FSD as', the symbol $\underset{\sim}{\textcircled{S}}$ means 'is at least as large in the sense of SSD as', and the symbol $\underset{\sim}{\textcircled{T}}$ means 'is at least as large in the sense of TSD as'. If any of the above inequalities holds strictly for at least one $x \in R$, then G is larger than F in the sense of the respective degree of dominance. When strict dominance holds, it is denoted by either $\textcircled{F}$, $\textcircled{S}$ or $\textcircled{T}$, as the case may be.

It may be noted that the three types of dominance represent specifications of decreasing strength; that is, FSD $\Rightarrow$ SSD $\Rightarrow$ TSD. Each defines a partial ordering on the set of all probability distributions such that the set of distributions that can be ordered by FSD is a subset of the set that can be ordered by SSD, and the latter set is a subset of the set that can be ordered by TSD.

5.2.1. Restrictions on moments

The relative strength of the FSD condition is reflected in certain restrictions on the moments of the distributions that are implied by this condition. It can be shown that if G $\textcircled{F}$ F, then all the odd moments of G are larger than those of F, and if the domain of G and F is the non-negative half-line, or a subset thereof, then all the moments of G are larger than those of F.

In the case of SSD, only the means are restricted, the restriction being that the mean of the dominating distribution is at least as large as that of the other distribution[2]. If the means are equal, then the dominating distribution has a smaller variance[3].

In the case of TSD, a restriction on the means (similar to the restriction implied by SSD) is incorporated into the definition itself, as

---

[2] The implications of FSD and SSD with respect to the moments of the distributions are immediate corollaries of theorems 5.4 and 5.5 stated below. See corollaries 1 and 2 in ref. [1].

[3] Theorem 3 in ref. [1].

indicated by the single-integral inequality. This is necessary, since the double-integral inequality appearing in the definition implies nothing about the relative magnitudes of the means[4]. However, if the means are equal, then the variance of the dominating distribution cannot exceed that of the other distribution[5].

5.2.2. Dominance by construction

It is often the case that one random variable is related to another random variable by an exact formula, that is, one variable is obtained from another by some transformation. Some such transformations can create situations of stochastic dominance between the respective distributions. Here we shall give two examples, one for the case of FSD, the other for SSD.

THEOREM 5.1. *Let X denote a random variable with a finite mean, and let $Y = \theta(X)$, where the function θ (not necessarily monotonic) has the property $\theta(x) \geq x$ for all $x \in R$. If* F *is the distribution function of X, and* G *is the distribution function of Y, then* G Ⓕ F.

PROOF. Let $x_y = \sup \{x \mid x = \theta^{-1}(y)\}$. Then $G(y) \leq F(x_y)$. But $\theta(x) \geq x$ implies $x_y \leq y$; hence $G(y) \leq F(y)$ for all $y \in R$.

It is obvious that if $\theta(x) > x$ for some $x \in R$, then G Ⓕ F. A special case of theorem 5.1 is $Y = a + bX$, $a \geq 0$, $b \geq 1$.

THEOREM 5.2. *Let X denote a random variable with a finite mean $\bar{x}$, and let $Y = a + bX$, where $0 < b < 1$, and $a \geq (1 - b)\bar{x}$. If* F *is the distribution function of X, and* G *is the distribution function of Y, then* G Ⓢ F.

This result has been proved in theorem 4 in ref. [1].

5.2.3. An invariance property

Suppose that one random variable dominates (in a sense to be made precise below) another random variable, and suppose that we construct two new random variables by multiplying the original variables by a positive constant, and adding another random variable to them. The question then is whether the newly constructed random variables will

---

[4] Without such a restriction on the means the desired theorem may not hold.

[5] See ref. [6].

satisfy the same dominance relation as the original ones. The answer is in the affirmative, as shown in the next theorem for the three types of dominance.

THEOREM 5.3. *Let X^1 and X^2 be two random variables with distribution functions* F^1 *and* F^2*, respectively, and let Y^1 and Y^2 be two identically distributed random variables with common distribution function* F. *Let* H^1 *be the conditional distribution function of X^1, given Y^1, and* H^2 *the conditional distribution function of X^2, given Y^2. Finally, let* G^1 *and* G^2 *be the distribution functions of the random variables $aX^1 + bY^1$ and $aX^2 + bY^2$, respectively, where $a > 0$ and $b \geq 0$. Then the following three statements are true:*

(a) *If* $\mathrm{H}^1 \ (\underset{\sim}{\mathrm{F}})\ \mathrm{H}^2$, *then* $\mathrm{G}^1 \ (\underset{\sim}{\mathrm{F}})\ \mathrm{G}^2$.

(b) *If* $\mathrm{H}^1 \ (\underset{\sim}{\mathrm{S}})\ \mathrm{H}^2$, *then* $\mathrm{G}^1 \ (\underset{\sim}{\mathrm{S}})\ \mathrm{G}^2$.

(c) *If* $\mathrm{H}^1 \ (\underset{\sim}{\mathrm{T}})\ \mathrm{H}^2$, *then* $\mathrm{G}^1 \ (\underset{\sim}{\mathrm{T}})\ \mathrm{G}^2$.

In the above, $\mathrm{H}^1 \ (\underset{\sim}{\mathrm{F}})\ \mathrm{H}^2$ means $\mathrm{H}^1(x \mid y) \leq \mathrm{H}^2(x \mid y)$ for all $x \in R$ and all $y \in R$, $\mathrm{H}^1 \ (\underset{\sim}{\mathrm{S}})\ \mathrm{H}^2$ means

$$\int_{-\infty}^{x} [\mathrm{H}^1(t \mid y) - \mathrm{H}^2(t \mid y)]\mathrm{d}t \leq 0$$

for all $x \in R$ and all $y \in R$, and $\mathrm{H}^1 \ (\underset{\sim}{\mathrm{T}})\ \mathrm{H}^2$ means

$$\int_{-\infty}^{\infty} [\mathrm{H}^1(t \mid y) - \mathrm{H}^2(t \mid y)]\mathrm{d}t \leq 0$$

for all $y \in R$, and

$$\int_{-\infty}^{x} \int_{-\infty}^{s} [\mathrm{H}^1(t \mid y) - \mathrm{H}^2(t \mid y)]\mathrm{d}t\mathrm{d}s \leq 0$$

for all $x \in R$ and all $y \in R$. The proofs of parts (a) and (b) are analogous to the proofs of theorem 5 in ref. [1] except that in the present theorem we employ conditional distributions. While part (c) has not been proved, the proof is omitted here since it follows along the same lines as those of parts (a) and (b).

5.2.4. Multivariate distributions

The case of multivariate distributions is not nearly as developed as that of single-variate distributions for the obvious reason that, except for some rather special cases, the complexity inherent in models with multivariate distributions makes it quite difficult to obtain neat theorems. In this paper we make a modest beginning by extending the analysis to situations in which multivariate distributions satisfy the FSD condition.

The definition of FSD for multivariate distributions is essentially the same as that given in definition 5.1 with appropriate changes in the dimensionality of the functions involved. The joint distribution functions are denoted by $\mathrm{H}(x)$ and $\hat{\mathrm{H}}(x)$, where x is an n-dimensional vector with components $x^1, x^2, \ldots, x^n$, and R^n is the common domain.

DEFINITION 5.1′.

$\mathrm{H}(x) \underset{\sim}{\textcircled{F}} \hat{\mathrm{H}}(x)$ if and only if $\mathrm{H}(x) \leq \hat{\mathrm{H}}(x)$ for all $x \in R^n$.

$\mathrm{H}(x) \textcircled{F} \hat{\mathrm{H}}(x)$ if and only if $\mathrm{H}(x) \leq \hat{\mathrm{H}}(x)$ for all $x \in R^n$, the strict inequality holding for at least one $x \in R^n$.

Clearly, definition 5.1 is a special case of definition 5.1′. An obvious property of joint distributions satisfying the FSD condition is that all the respective 'marginal' distributions of dimensions 1 through $n - 1$ satisfy a similar dominance relation. That is, if H_s and $\hat{\mathrm{H}}_s$, $s = 1, 2, \ldots, n - 1$, are s-dimensional distribution functions of H and $\hat{\mathrm{H}}$, respectively, each containing as arguments variables having the same indices, then $\mathrm{H} \underset{\sim}{\textcircled{F}} \hat{\mathrm{H}}$ implies $\mathrm{H}_s \underset{\sim}{\textcircled{F}} \hat{\mathrm{H}}_s$.

The FSD condition also has certain implications with respect to the moments of the joint distribution functions; these will be indicated in Section 5.3 below.

5.3. Stochastic Dominance and Preference

The application of the stochastic dominance conditions to problems of choice under uncertainty is made possible by virtue of certain equivalence theorems stated below. Implicit in these theorems is the assumption that the decision maker's choice is governed by expected utility maximization. Each theorem applies to a particular set of utility functions; these sets are defined below.

DEFINITION 5.2.

(a) U_1 is the set of all bounded and strictly increasing functions with a continuous first derivative everywhere in R.

(b) U_2 is the set of all bounded and strictly increasing functions with continuous derivatives of order one and two, the second derivative being nonpositive everywhere in R.

(c) U_3 is the set of all bounded and strictly increasing functions with continuous derivatives of order one, two and three, the second derivative being non-positive, and the third derivative being non-negative everywhere in R.[6]

We shall use the symbols $\underset{\sim}{\textcircled{P}}$ and $\textcircled{P}$ to denote 'is at least as preferred as' and 'is preferred to', respectively. In theorems 5.4–5.6 the distribution functions are one-dimensional.

THEOREM 5.4. G $\textcircled{F}$ F *if and only if* G $\textcircled{P}$ F *for all utility functions in* U_1.

For proofs of this theorem see refs. [1]–[3] and [5]. This theorem applies essentially to all expected utility maximizers, regardless of their attitude toward risk. The next theorem is restricted to risk averters.

THEOREM 5.5. G $\underset{\sim}{\textcircled{S}}$ F *if and only if* G $\underset{\sim}{\textcircled{P}}$ F *for all utility functions in* U_2.

For proofs see refs. [1] and [2]. The last theorem in this group, which uses the TSD condition, is motivated by the hypothesis of decreasing absolute risk aversion. A necessary condition for that hypothesis is that marginal utility be a convex function; hence utility functions satisfying this condition are members of U_3. However, convexity of marginal utility is not a sufficient condition for decreasing absolute risk aversion, so that if the analysis is confined to only that subset of U_3 which exhibits decreasing absolute risk aversion, it should be possible to use a condition that is weaker than TSD.

THEOREM 5.6. G $\underset{\sim}{\textcircled{T}}$ F *if and only if* G $\underset{\sim}{\textcircled{P}}$ F *for all utility functions in* U_3.

For a proof see ref. [6].

It may be pointed out that because of the monotonicity of the utility function, stochastic dominance between uncertain prospects carries over to the distributions of utility. Let $\theta_F^j(u)$ denote the distribution function of the jth individual's utility when the distribution function of the

---

6 These sets can be enlarged inasmuch as in some of the proofs of theorems 5.4 and 5.5 the differentiability assumption is not used.

uncertain prospect X is given by F; and a similar notation for the distributition G. And let D stand for either FSD, SSD or TSD. Then G Ⓓ F implies θ_G^j Ⓓ θ_F^j for all j. Thus, if a set of uncertain prospects is completely ordered by one of the stochastic dominance relations, then for each individual j the set of distributions of utility induced by the set of uncertain prospects is also completely ordered by the same relation.

For problems with multivariate distributions we have, at this time, only a sufficient condition for preference between uncertain prospects. Because of the complexity encountered in these problems, it may be worthwhile to look first at the special case of bivariate distributions. We shall consider two joint distribution functions $H(x^1, x^2)$ and $\hat{H}(x^1, x^2)$ with associated marginal distributions F^1, F^2, $\hat{F}^1$ and $\hat{F}^2$. The associated densities will be denoted by the respective lower case letters. The utility function is denoted by $u = \phi(x^1, x^2)$, and its partial derivatives by ϕ_1, ϕ_2 and ϕ_{12}. As in the one-dimensional case, the utility function is strictly increasing and bounded. A symbol such as $\bar{u}_H$ denotes expected utility when the joint distribution of the uncertain prospects is given by H.

THEOREM 5.7. *Assume that $\phi_{12} \leq 0$ for all $x \in R^2$. Then* H Ⓕ $\hat{\text{H}}$ *implies* H Ⓟ $\hat{\text{H}}$.

PROOF. Since a distribution is preferred only if its expected utility is higher, we prove the theorem by showing that the difference between the expected utilities (assumed to be finite) is non-negative. By definition

$$\bar{u}_H - \bar{u}_{\hat{H}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x^1, x^2)[h(x^1, x^2) - \hat{h}(x^1, x^2)]\,dx^1 dx^2, \quad (5.1)$$

where h and $\hat{h}$ are the densities of H and $\hat{\text{H}}$, respectively.

We now carry out several integrations by parts. First, integrating the inner integral in eq. (5.1) with respect to x^1 yields

$$\phi(\infty, x^2)[f^2(x^2) - \hat{f}^2(x^2)] - \int_{-\infty}^{\infty} \phi_1(x^1, x^2) \int_{-\infty}^{x^1} [h(s, x^2) - \hat{h}(s, x^2)]\,ds\,dx^1. \quad (5.2)$$

Integrating the first expression in eq. (5.2) with respect to x^2 yields

$$-\int_{-\infty}^{\infty} \phi_2(\infty, x^2)[F^2(x^2) - \hat{F}^2(x^2)]\,dx^2, \quad (5.3)$$

and integrating the second expression in eq. (5.2) with respect to x^2 gives

$$- \int_{-\infty}^{\infty} \phi_1(x^1, \infty)[F^1(x^1) - \hat{F}^1(x^1)]dx^1$$
$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_{12}(x^1, x^2)[H(x^1, x^2) - \hat{H}(x^1, x^2)]dx^1dx^2, \quad (5.4)$$

so that

$$\bar{u}_H - \bar{u}_{\hat{H}} = - \int_{-\infty}^{\infty} \phi_1(x^1, \infty)[F^1(x^1) - \hat{F}^1(x^1)]dx^1 - \int_{-\infty}^{\infty} \phi_2(\infty, x^2)[F^2(x^2)$$
$$- \hat{F}^2(x^2)]dx^2 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_{12}(x^1, x^2)[H(x^1, x^2) - \hat{H}(x^1, x^2)]dx^1dx^2. \quad (5.5)$$

The first two terms in eq. (5.5) are non-negative by the monotonicity of the utility function, and the dominance relations between the respective marginal distributions (which are implied by the FSD assumption postulated in the theorem). The last term is non-negative by virtue of the nonpositivity of the cross-partial derivative, and the FSD assumption.

If the utility function is separable, i.e. $\phi_{12} = 0$ for all $x \in R^2$, then the properties of the joint distributions (except for the marginal distributions) play no role in determining preference between the two distributions. In that case, the distribution H is preferred because each of its marginal distributions dominates the respective marginal distribution of $\hat{H}$. Since the existence of FSD between two marginal distributions corresponds roughly to a shift in the probability distribution, the first two terms in eq. (5.5) may be said to reflect a *shift effect.* When there exists FSD between two joint distributions, on the other hand, there occurs a redistribution of the probability mass not only along the 'margins', but possibly over the entire domain. Therefore, the last term in eq. (5.5) represents the *redistribution effect.* In the special case in which the marginal distributions are equal, that is, $F^1 = \hat{F}^1$, and $F^2 = \hat{F}^2$, preference for H is due entirely to the redistribution effect.

For the case of identical marginal distributions one can immediately obtain some implications about the joint moments. Let us call the expected value of the function $\phi(x^1, x^2) = (x^1)^q(x^2)^r$ an *odd* moment if and only if both the integers q and r are odd. Since in that case we have $\phi_{12} = qr(x^1)^{q-1}(x^2)^{r-1} \geq 0$, it follows from eq. (5.5) that all the odd moments of H are at most as large as those of $\hat{H}$. If both random variables are non-negative, then all the joint moments of H are at most as large as those of $\hat{H}$.

When $\phi(x^1, x^2) = x^1x^2$, its expected value is the product moment, and $\phi_{12} = 1$. From eq. (5.5) we then see that the product moment of H is less than that of $\hat{\mathrm{H}}$. But when the marginal distributions (and hence the means) are identical, the difference between the product moments is equal to the difference between the covariances, and therefore the covariance of H is less than that of $\hat{\mathrm{H}}$.

This last implication helps to explain the preference for H when the cross-partial derivative is negative. This latter condition means that an increase in one variable of the utility function decreases the marginal utility of the other. Then it is desirable that the two variables should move in opposite directions. For example, suppose that x^1 increases while x^2 decreases. Since this increases the marginal utility of x^1, the increase in utility due to the increase in x^1 is larger than it would be if x^2 did not decrease. Furthermore, the increase in x^1 diminishes the marginal utility of x^2; hence the loss in utility due to the decrease in x^2 is less than it would be if x^1 did not increase. Thus, H is preferred since the smaller its covariance, the greater is the likelihood that the two variables will move in opposite directions. By the same argument, it also follows that if the cross-partial derivative of the utility function is positive (still assuming identical marginals), then $\hat{\mathrm{H}}$ is the preferred distribution.

When there are n variables in the utility function, preference depends on the signs of various cross-partial derivatives up to order n. In order to state the general equation for the n-variable case, we shall use notation such as H_i for a marginal distribution of H, H_{ij} for a joint distribution of the variables x^i and x^j, and so on. Then the generalized version of eq. (5.5) can be written as

$$
\begin{aligned}
\bar{u}_{\mathrm{H}} - \bar{u}_{\hat{\mathrm{H}}} = & -\sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi_i[\mathrm{H}_i(x^i) - \hat{\mathrm{H}}_i(x^i)]\mathrm{d}x^i \\
& + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_{ij}[\mathrm{H}_{ij}(x^i, x^j) - \hat{\mathrm{H}}_{ij}(x^i, x^j)]\mathrm{d}x^i\mathrm{d}x^j \\
& - \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_{ijk}[\mathrm{H}_{ijk}(x^i, x^j, x^k) \\
& \qquad - \hat{\mathrm{H}}_{ijk}(x^i, x^j, x^k)]\mathrm{d}x^i\mathrm{d}x^j\mathrm{d}x^k \\
& \quad \vdots \\
& + (-1)^n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_{12\cdots n}[\mathrm{H}(x) - \hat{\mathrm{H}}(x)]\mathrm{d}x^1\mathrm{d}x^2 \ldots \mathrm{d}x^n. \qquad (5.5')
\end{aligned}
$$

It may be noted that the number of terms involving a partial derivative of order r is equal to the number of possible combinations of n objects taken r at a time. We can now generalize theorem 5.7 as follows:

THEOREM 5.8. *Suppose that the utility function satisfies, in addition to monotonicity, the following conditions:* $\phi_{ij} \leq 0$ *for all* $i, j, i \neq j$, $\phi_{ijk} \geq 0$ *for all* $i, j, k, i \neq j, j \neq k, i \neq k, \ldots (-1)^{n+1}\phi_{12\cdots n} \geq 0$. *Then* H Ⓕ Ĥ *implies* H Ⓟ Ĥ.

5.4. Applications

One area in which the stochastic dominance approach has already produced several results is portfolio choice, and in particular the problem of optimality of diversification. In the presentation of these results we restrict ourselves to the set of risk averters (definition 5.2). In that case, the utility function is not only monotonic and bounded, but also concave, and therefore its domain is bounded from below. For convenience, the random variables X^i are taken to be non-negative. A portfolio will be defined by the random variable $P(k) = \sum_{i=1}^{n} k^i X^i$, where $k^i \geq 0$, $\sum_{i=1}^{n} k^i = 1$, and k.denotes the vector of the mixture coefficients k^i.

A portfolio is said to be:

completely diversified if and only if $k^i \neq 0$, $i = 1, 2, \ldots, n$;
partially diversified if and only if $k^i = 0$ for at least one i, and at most $(n - 2)$ i;
specialized if and only if $k^i = 1$ for some i.

We begin with the special case of portfolios composed of two uncertain prospects X^1 and X^2. In that case we write $P(k) = kX^1 + (1 - k)X^2$, where the joint distribution of X^1 and X^2 is denoted by H, and its marginal distributions by F^1 and F^2. First we show that if the marginal distributions are identical, then *any* diversified portfolio is at least as preferred as the specialized portfolios for *every* risk averter. This we have already proved in theorem 8 in ref. [1] for the case of independently distributed prospects. It turns out that this result generalizes to interdependent prospects without any additional restrictions. In view of theorem 5.5, the result is proved by showing that the distribution of the diversified portfolio is at least as large as that of the specialized portfolios in the sense of SSD.

THEOREM 5.9. *Suppose the distribution function of the two-prospect portfolio P(k) is given by* G(x; k). *If* $F^1(x) = F^2(x) = F(x)$ *for all* $x \in R$, *then* G(x; k) Ⓢ F(x) *for all k satisfying* $0 < k < 1$.

PROOF. The distribution function of the portfolio is given by

$$G(z; k) = \frac{1}{1-k} \int_0^z H_2\left(\frac{t}{k}, \frac{z-t}{1-k}\right) dt, \tag{5.6}$$

where $H_2 = \partial H/\partial(x^2)$. Integrating eq. (5.6) yields

$$\int_0^x G(z; k) dz = \int_0^x H\left(\frac{t}{k}, \frac{x-t}{1-k}\right) dt. \tag{5.7}$$

Defining a new variable $Y = T/k$, we can write

$$\int_0^x G(z; k) dz = \underbrace{k \int_0^x H\left(y, \frac{x-ky}{1-k}\right) dy}_{I_1} + \underbrace{k \int_x^{x/k} H\left(y, \frac{x-ky}{1-k}\right) dy}_{I_2}. \tag{5.8}$$

Changing variables in I_2 by defining $T = (X - kY)/1 - k$, we get

$$I_2 = (1-k) \int_0^x H\left[\frac{x-(1-k)t}{k}, t\right] dt. \tag{5.9}$$

Clearly, $I_1 \leq k\int_0^x F(y)dy$, and $I_2 \leq (1-k)\int_0^x F(t)dt$ (from the definition of joint and marginal distributions). Therefore

$$\int_0^x G(z; k) dz \leq k\int_0^x F(y)dy + (1-k)\int_0^x F(t)dt = \int_0^x F(t)dt \quad \text{for all } x \text{ and } k, \tag{5.10}$$

which was to be proved.

Further results can be obtained under the stronger assumption that the joint distribution function H is symmetric; that is, $H(x^1, x^2) = H(x^2, x^1)$ for all $x \in R^2$. Under these conditions, the set of all portfolios can be completely ordered by the mixture coefficient k as shown in the next theorem.

THEOREM 5.10. *Let* $G(x; k)$ *denote the distribution function of the two-prospect portfolio* $P(k)$, *and assume that the joint distribution function* H *is symmetric in its two arguments. Then for any two* k *and* k', $G(x; k)$ Ⓢ $G(x; k')$ *if and only if* $(\frac{1}{2} - k)^2 < (\frac{1}{2} - k')^2$.

PROOF. The proof consists of showing that, for each positive x, the function $G^*(x; k) = \int_0^x G(z; k)dz$ is a strictly convex and symmetric function of k, attaining its unique minimum at $k = \frac{1}{2}$. From eq. (5.7) we have

$$G^*(x; k) = \int_0^x H\left(\frac{t}{k}, \frac{x - t}{1 - k}\right)dt. \tag{5.11}$$

Changing variables in eq. (5.11) with the relation $Y = x - T$, one sees immediately that $G^*(x; k) = G^*(x; 1 - k)$ for all x by the symmetry of H, so that $G^*(x; k)$ is symmetric around $k = \frac{1}{2}$. Strict convexity is established by evaluating the second derivative of G^* with respect to k. Differentiating G^* twice, and using suitable changes of variables, one can show that

$$G^*_{kk}(z; k) = \frac{1}{(1 - k)^3} \int_0^{z/k} h\left(t, \frac{z - kt}{1 - k}\right)(z - t)^2 dt > 0 \tag{5.12}$$

for all z and k, where h is the joint density function.

Thus, we have shown that, under the conditions postulated, all portfolios can be completely ordered, and the ordering has the property that the closer k is to $\frac{1}{2}$, the better the portfolio.

The value of the mixture coefficients can also be related to the variance of the portfolio. In order to derive this relationship, we consider the general case of n-prospect portfolios, and define a distance function of the vector k by

$$D(k) = \sum_{i=1}^{n} \left(\frac{1}{n} - k^i\right)^2 = \sum_{i=1}^{n} (k^i)^2 - \frac{1}{n}, \tag{5.13}$$

where the second equation follows from the fact that the k^i add up to unity. Now, if the joint distribution function is symmetric, the variances of all prospects are equal, and the covariances between all pairs of prospects are equal. We denote the common variance by v, and the common covariance by c. If the variance of the portfolio with coefficient vector k is denoted by $v(k)$, then we have

$$v(k) = (v - c) \sum_{i=1}^{n} (k^i)^2 + c. \tag{5.14}$$

Then, given any two portfolios (constructed from some common set of n prospects) with coefficient vectors k and k', the following equivalence relation holds:

$$v(k) < v(k') \Leftrightarrow D(k) < D(k'). \tag{5.15}$$

But for $n = 2$ we can invoke theorem 5.10 which shows that

$$D(k) < D(k') \Leftrightarrow \mathrm{G}(x; k) \circledS \mathrm{G}(x; k'). \tag{5.16}$$

Combining (5.15) and (5.16) we get

$$v(k) < v(k') \Leftrightarrow \mathrm{G}(x; k) \circledS \mathrm{G}(x; k'). \tag{5.17}$$

The equivalence relation in eq. (5.17) shows that all two-prospect portfolios constructed from prospects whose joint distribution function is symmetric can be completely ordered either by the portfolio variance or by the SSD condition. This case, therefore, represents another instance in which the mean-variance method yields a correct ordering of portfolios[7]. At the same time it should be noted that the result in eq. (5.17) does in general not hold for $n > 2$. The reason for this is that when $n > 2$, the equivalence relation in eq. (5.16) may no longer hold.

Other results presented above also require stronger restrictions if their validity is to be extended to situations in which $n > 2$. For example, identity of marginal distributions is no longer sufficient to establish optimality of complete diversification. But that result will hold if all joint distribution functions of dimension $n - 1$ are equal. Then it can be shown that for any partially diversified portfolio there exists at least one completely diversified portfolio whose distribution dominates that of the partially diversified portfolio in the sense of SSD. This result is the subject of the next theorem.

---

[7] The other known instances are: (1) a quadratic utility function, and (2) portfolios whose distributions belong to a special class of two-parameter families of distributions. See ref. [2].

THEOREM 5.11. *Let* H *denote the joint distribution function of the* n *non-negative random variables* X^i, *and* H^{-s} *the joint distribution function of the* $n-1$ *variables* X^i, $i = 1, 2, \ldots, n$, $i \neq s$. *Consider the* $(n-1)$-*prospect portfolio*

$$P(k) = \sum_{i=1}^{n-1} k^i X^i,$$

where $k = (k^1, k^2, \ldots, k^{n-1})$. *If* $\mathrm{H}^{-s} = \mathrm{H}^{-t}$ *for* $s, t = 1, 2, \ldots, n$, *then there exists a vector of mixture coefficients* k' *with* n *components such that the distribution of the portfolio* $P(k')$ *dominates that of* $P(k)$ *in the sense of SSD.*

PROOF. Denote $P(k)$ by the random variable V, and consider

$$W = \sum_{i=1}^{n-2} k^i X^i + k^{n-1} X^n.$$

From the identity of the joint distribution functions of dimension $n-1$ it is clear that V and W are identically distributed. Then it follows from theorem 5.9 that the distribution of Z, where $Z = qV + (1-q)W$, $0 < q < 1$, is at least as large as that of V in the sense of SSD. But

$$Z = \sum_{i=1}^{n-2} k^i X^i + qk^{n-1} X^{n-1} + (1-q)k^{n-1} X^n$$

is an n-prospect portfolio $P(c)$ where $c^i = k^i$, $i = 1, 2, \ldots, n-2$, $c^{n-1} = qk^{n-1}$, and $c^n = (1-q)k^{n-1}$, so that the conclusion of the theorem holds for $k' = c$.

The above result shows that equality of joint distributions of dimension $n-1$ is sufficient for optimality of complete diversification; so long as for any partially diversified portfolio there exists a completely diversified portfolio which dominates in the sense of SSD, all risk averters will prefer a completely diversified portfolio. When the joint distribution function H satisfies the stronger condition of symmetry then, as in the case of $n = 2$, equal diversification is the best strategy. And while it is in general not true (except when $n = 2$) that any given portfolio can be improved by merely changing the mixture coefficients in a way which reduces the value of the distance function $D(k)$ (as defined in eq. (5.13)), there exist ways of changing the mixture coefficients which will produce superior portfolios. One such method is the following: every mixture coefficient is moved one position forward, and the first coefficient is moved to the last position. Thus, if the initial portfolio is $Y = \sum_{i=1}^{n} k^i X^i$, the rearranged portfolio is $Y' = \sum_{i=1}^{n-1} k^{i+1} X^i + k^1 X^n$, where we call Y' a rearrangement of Y. Then any mixture of Y and Y' is superior to the original portfolio Y as proved below.

THEOREM 5.12. *Let $P(k) = \sum_{i=1}^{n} k^i X^i$ be a portfolio of the n non-negative random variables X^i whose joint distribution function is given by* H. *If* H *is symmetric in its arguments, and $k^i \neq 1/n$ for at least one i, then there exists a vector of mixture coefficients k' such that the distribution of $P(k')$ is at least as large as that of $P(k)$ in the sense of SSD.*

PROOF. Let $Y = \sum_{i=1}^{n} k^i X^i$ and let Y' be a rearrangement of Y. Since Y and Y' are identically distributed (by the symmetry of H), it follows from theorem 5.9 that the distribution of Z, where $Z = cY + (1 - c)Y'$, $0 < c < 1$, is at least as large as that of Y in the sense of SSD. But Z is the portfolio $P(\hat{k}) = \sum_{i=1}^{n} \hat{k}^i X^i$, where $\hat{k}^i = ck^i + (1 - c)k^{i+1}$, $i = 1, 2, \ldots, n - 1$, and $\hat{k}^n = ck^n + (1 - c)k^1$. Therefore the conclusion of the theorem holds for $k' = \hat{k}$.

This method of reshuffling the mixture coefficients can be applied repeatedly with the result that at each stage there emerges a portfolio that dominates its predecessor in the sense of SSD. It follows also from eq. (5.16) that after each reshuffling of the mixture coefficients the distance function $D(k)$ is reduced. It is clear that no further improvement will be possible when $D(k)$ reaches zero. This will be the case if and only if $k^i = 1/n$ for all i. Hence the following conclusion.

THEOREM 5.13. *If the joint distribution function* H *is symmetric in its arguments, then complete and equal diversification is optimal.*

One common feature found in all the above theorems on diversification is the equality of the marginal distributions of the uncertain prospects. When this condition fails to hold, diversification is much less likely to be optimal. Below we present an example of two-prospect portfolios in which the prospects need not be identically distributed, but have equal means. To prove the desired result, we shall make use of a theorem by Samuelson according to which each risk averter will choose a diversified portfolio so long as all prospects have equal means, and are independently distributed[8]. Then it can be shown that diversification is optimal even if the prospects are not independent, provided the joint distribution function dominates the product of its marginal distributions in the sense of FSD.

THEOREM 5.14. *Let* $H(x^1, x^2)$ *denote the joint distribution function of two non-negative random variables with marginal distributions* F^1 *and* F^2. *Assume that the two variables have the same mean. Then* H Ⓕ F^1F^2 *implies that diversification is optimal.*

PROOF. The utility function is defined by $\phi(x^1, x^2) = \psi[kx^1 + (1-k)x^2]$, so that $\phi_{12} = k(1-k)\psi'' < 0$ by the assumption of risk aversion. Then it follows from theorem 5.7 that H Ⓟ F^1F^2. But from Samuelson's theorem we know that if the prospects are independent, every risk averter chooses a diversified portfolio. Consequently, since each risk averter prefers H to F^1F^2, he will choose a diversified portfolio when H prevails[9].

5.5. Conclusion

The equivalence between stochastic dominance and preference implies that problems of choice under uncertainty can be analyzed by either one of two methods: (i) comparing the expected utilities of the prospects under consideration, or (ii) testing for the existence of stochastic dominance. The stochastic dominance approach differs from the expected utility approach in that it focuses directly on the probability distributions which are, of course, the objects of the decision maker's choice. This has the advantage of providing an insight into the nature of the decision maker's preferences.

In the present paper, the applications of the stochastic dominance approach are taken from the area of portfolio selection, but choice problems in the context of the theory of the firm and the consumer lend themselves equally well to this approach. This paper also makes an initial attempt to extend the analysis of choice under uncertainty to cases of multivariate utility functions. Such an extension is essential for

[8] Corollary II in ref. [4].
[9] For a different approach to this problem see theorem IV in ref. [4].

problems in which prospects are characterized by a vector of goods or attributes which may or may not enter into the utility function as a linear combination. Such situations occur in the theory of the firm, consumer behavior, welfare economics and others. Because of the complexity of problems with multivariate distributions, strong results are more difficult to obtain. At the present time the whole area of uncertainty with multivariate distributions is still largely unexplored, and much more research needs to be done to determine the scope of this theory.

References

[1] Hadar, J. and Russell, W. R. Stochastic dominance and diversification. *Journal of Economic Theory,* **3** (1971), 288–305.
[2] Hanoch, G. and Levy, H. The efficiency analysis of choices involving risk. *Review of Economic Studies,* **36** (1969), 335–346.
[3] Quirk, J. P. and Saposnik, R. Admissibility and measurable utility functions. *Review of Economic Studies,* **29** (1962), 140–146.
[4] Samuelson, P. A. General proof that diversification pays. *Journal of Financial and Quantitative Analysis,* **2** (1967), 1–13.
[5] Schneeweiss, H. Note on two dominance principles in decision theory. *Unternehmensforschung,* **12** (1968), 213–216.
[6] Whitmore, G. A. Third-degree stochastic dominance. *American Economic Review,* **60** (1970), 457–459.

COMMENTS

*On 'Stochastic dominance in choice under uncertainty'**

Koichi Hamada

The Hadar–Russell paper provides a systematic exposition of the stochastic dominance approach, and at the same time generalizes the

* I owe much to Professor S. Wu for helpful correspondence.

authors' previous work [1][1] to include the case where there are more than two random variables which may be interdependent.

In order to discuss the significance of their contribution, let us compare their results about the diversification of portfolio with those of Samuelson [4]. Samuelson has proved that the symmetry of distribution, which includes as a special case the multivariate distribution of independent random variables with identical marginal distribution, implies the equal holding of every security as the optimal portfolio mix. On the other hand, Hadar and Russell generalize the argument to the case where the distribution is not symmetric but has identical joint (marginal) distributions if one excludes one variable. Theorem 5.11 shows that a completely diversified portfolio dominates a partially diversified portfolio.

Therefore, the significance of this generalization depends on the equality of $(n - 1)$ dimensional joint distribution. Let us take the example of a multivariate normal distribution. The equality of these joint distributions implies the equality of all variances and the equality of all covariances. This means that the distribution is symmetric.

If we do not specify the parametric form of a distribution, we can generate quite a wide class of distributions illustrated by the following device for a two-dimensional distribution. Let us take initially a symmetric bivariate density function $h(x, y)$. Take small positive numbers ε and δ, and construct a new density function $h^*(x, y)$ as follows. Let $a < a + \delta < b < b + \delta < c$, and $d < d + \delta < e < e + \delta < f$. Define

$$h^*(x, y) = h(x, y) + \varepsilon \quad \text{for } a \leq x < a + \delta,\ f \leq y < f + \delta,\quad b \leq x < b + \delta,\ d \leq y < d + \delta,\quad c \leq x < c + \delta,\ e \leq y < e + \delta.$$

$$h^*(x, y) = h(x, y) - \varepsilon \quad \text{for } a \leq x < a + \delta,\ e \leq y < e + \delta,\quad b \leq x < b + \delta,\ f \leq y < f + \delta,\quad c \leq x < c + \delta,\ d \leq y < d + \delta.$$

and

$$h^*(x, y) = h(x, y) \qquad \text{otherwise.}$$

[1] The references here are the same as those in the Hadar–Russell paper.

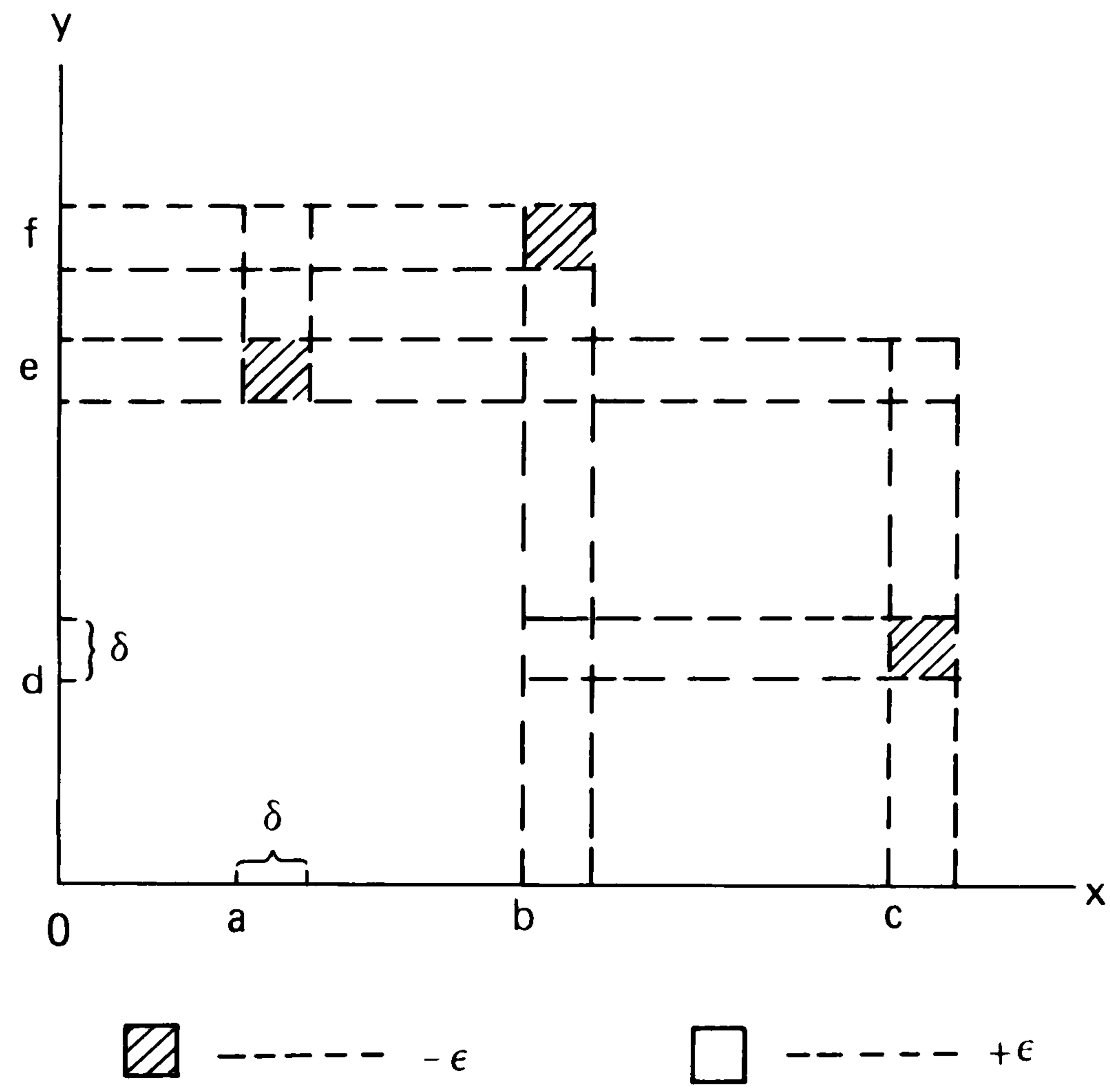

Fig. 5.1. *h* to *h**.

(Of course, it is assumed that ε is taken small enough to keep $h^*(x, y)$ non-negative.) This method is illustrated in fig. 5.1. By repeating this kind of 'marginal preserving transformation', we can reach a wide variety of density functions.

Incidentally, this transformation can be also used to visualize how to construct an example of two distributions which have identical marginal distributions but one of which dominates the other in the sense of first degree dominance. The reader will easily see in fig. 5.2

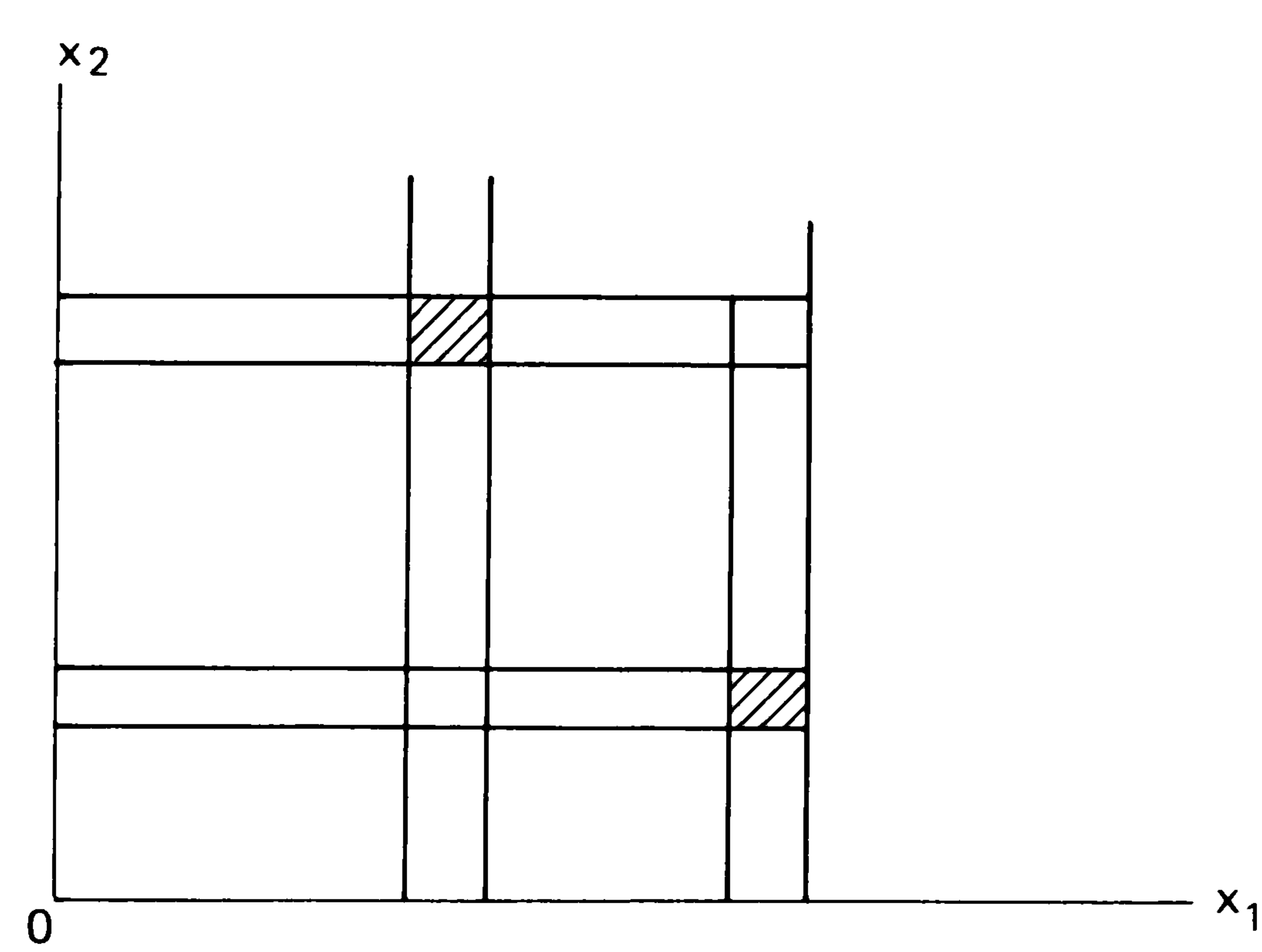

Fig. 5.2. H to Ĥ.

why covariance of H is less than that of $\hat{H}$, as argued in the discussion after theorem 5.7.

Thus, the generalization allows many more cases to be covered by the stochastic dominance approach. Theoretically, this indicates a significant advance. However, from the practical or empirical point of view, it still will be very hard to find cases where random variables have identical $(n - 1)$-dimensional joint distributions. The generality realized by the stochastic dominance approach so far is not without cost.

In this respect, we should not yet dismiss the mean-variance approach too lightly, even though its lack of generality has led us to the expected utility approach and the stochastic dominance approach. Since the authors have already compared the characteristics of the stochastic dominance approach with those of the expected utility approach, this might not be the place to defend the remaining value of the mean-variance approach. However, the recent work of Steve Ross[2] seems to enhance the practicality of the mean-variance approach. He has found that when all optimal portfolios are greatly diversified and when returns of securities are roughly independently distributed, the mean-variance approach can be used as an approximation. What is important is that he has found concrete bounds on the errors in using the mean-variance approach.

Finally, we cannot help but doubt the operationality of asking when it pays to diversify completely. Indeed, it is an interesting intellectual exercise. However, does it lead to any operationally testable hypotheses even under ideal conditions, emphasized by Samuelson, the originator of the question?

Note that the above critical remark is not directed at the stochastic dominance approach. It should be taken as an encouragement for the authors to remove the remaining obstacles to make the stochastic dominance approach more operational as well as to make it applicable to a wider range of problems.

It is always good to have many ways of looking at the same thing. The lucid exposition of the stochastic dominance approach by Hadar and Russell convinces us of the capacity of the new microscope with the name of stochastic dominance approach.

---

[2] S. A. Ross. *Portfolio and capital market theory with arbitrary preferences and distribution: the general validity of mean-variance approach in large markets.* Working Paper No. 12-72. Wharton School of Finance and Commerce, University of Pennsylvania.

On 'Stochastic dominance in choice under uncertainty'

Gerald L. Nordquist

Hadar and Russell's paper attempts both a summary and extension of recent work on stochastic dominance as a method for ordering preferences among risky prospects. The first six theorems restate and in minor respects modify propositions appearing in Quirk and Saposnik, Hadar and Russell, Hanoch and Levy, and Whitmore. Theorems 5.7 and 5.8 take a couple of definite albeit limited steps towards generalization of dominance to include multivariate distributions. The results here give sufficient conditions for the class of distribution functions ordered by dominance of the first degree (FSD). Theorems 5.9–5.14 develop further the authors' application of stochastic dominance to the problem of portfolio choice. I offer two points on the content of the paper and a general observation on the value of stochastic dominance as an approach to the economics of uncertainty.

From the standpoint of uncertainty theory, the most interesting part of the paper is to be found in the two theorems on multivariate distributions. Although by their own admission the authors do not get very far in establishing the relationship between stochastic dominance and preference in the n-variate case, they do provide some general directions and incidentally alert us to some of the obstacles to be encountered. We see immediately that monotonicity of the utility function is not sufficient to insure preference in the case of FSD; now we must take into account and properly sign all cross-partial derivatives up to order n. This is not quite as bad as it seems at first glance because of the reversibility of the order of partial differentiation and the fact that integration by parts yields non-repeating indices of differentiation (i.e. after a variable is integrated out, it stays out). Even so, the number of surviving terms is still $(2^n - n - 1)$, which obviously grows very fast with n.

The multivariate case grows even more complex in passing from FSD to dominance of the second and higher degrees. One serious question relates to the signs of the marginal distributions in the expansion. (The reference is to eq. (5.13).) Although it is surely true, as the authors suggest, that dominance on the joint distributions implies dominance on the respective marginals of various orders, it is not so obvious that this must be so in the cases of SSD and TSD. In any event, it is abundantly clear that the price of generalizing this theory is very steep.

The last section of the Hadar–Russell paper is devoted to showing how stochastic dominance can be applied to the problem of portfolio choices. In view of the current state of portfolio theory one is tempted to say that there is less here than meets the eye. The argument is that portfolios of various mixtures can be ordered by second-degree stochastic dominance. Thus, we find that given equal marginal distributions, a two-prospect portfolio dominates a specialized portfolio (theorem 5.9); a more equally diversified two-prospect portfolio dominates one that is less equally diversified assuming a symmetric joint distribution (theorem 5.10); complete diversification dominates partial diversification in the case of $n > 2$ so long as all the joint distributions of $n - 1$ variables are equal (theorem 5.11); and complete and equal diversification dominates complete and unequal diversification, again assuming a symmetric joint distribution of the n random variables.

These results, although interesting, are easily seen to be implications of the main theorem on SSD and certain well known propositions in portfolio theory. In particular there are the theorems by Samuelson which assert the optimality of diversification under similar if not identical conditions. The point is that given the class of symmetric, concave utility functions, if (1) greater diversification is preferred to less, and (2) a preference ordering implies and is implied by SSD, must it not follow that greater diversification will dominate less diversification in the sense of SSD? On the other hand, the Hadar and Russell theorems do provide some additional insights which should not be overlooked. For one thing they show once again the limited scope of mean-variance ranking. They also show that independence need not be invoked to establish the proposition that complete diversification dominates partial diversification.

My final comment is a judgment on the stochastic dominance approach to the economics of uncertainty. It is impossible to deny that it has produced several propositions and corollaries which help a great deal to fill in some awkward gaps in our understanding of the expected utility hypothesis. So far as they go, the main dominance theorems reveal with considerable precision and generality the tradeoff between different attitudes towards risk and required restrictions on the set of admissible prospects. Rival propositions involving moments, spreads and other wrinkles on probability densities are shown to be special cases of stochastic dominance. Of course we must temper our praise with the knowledge that these theorems still give us only partial rankings of risky prospects. In this connection it is particularly disappointing to find that dominance of the third degree does not take us much beyond SSD; TSD still requires that the mean of a dominant distribution be not less than the mean of any that is dominated by it.

The other side of the coin is that stochastic dominance is not likely to generate much practical interest largely because of its limited intuitive appeal. Imagine if you will the reaction of the typical banker who is advised that portfolio A is better than portfolio B in the sense of SSD! Knowledgeable observers of business practice will tell you that there is enough trouble with the notions of variance and mean-preserving spreads.

CHAPTER 6

CONSUMPTION AND PORTFOLIO CHOICES WITH TRANSACTION COSTS*

Robin Mukherjee and Edward Zabel

6.1. Introduction

The problem of the individual's consumption and portfolio choices over time has been the focus of recent studies by a number of authors, for example, Levhari and Srinivasan [10], Hakansson [6–8], Leland [9], Yaari [17], Samuelson [16], Merton [11, 12], Hahn [5] and Fama [3]. Apart from Yaari, in one way or another these studies have offered exposition, simplification of proof or generalization of Phelps' 1962 paper on consumer behavior [14].

An attempt is made here to extend these results by examining the impact of transaction costs on optimal consumption and portfolio decisions. We are able to show that these costs considerably modify available results and greatly increase the difficulty of analyzing the consumer choice problem. The major reason is that now not only wealth but also the composition of wealth becomes important in the decision making process.

A very brief outline of Phelps' paper and later generalizations will help to provide the setting of the present study. Phelps determines the individual's allocation over time of an initial wealth and a given income stream, between periodic consumption and investment in a single risky asset, which maximizes the expected utility of lifetime consumption. One direction of generalization is to extend the number of portfolio opportunities available to the individual. Hakansson, for example, permits investment in an arbitrary number of risky assets, some of

* The second author gratefully acknowledges the support of the National Science Foundation under Grant GS-30513 to the University of Rochester.

which may be sold short and also allows borrowing and lending [6, 7]. Samuelson [16] and Merton [11, 12], with Merton using continuous, rather than discrete, time analysis, have treated a similar problem. Another direction is to introduce an end-of-lifetime function when lifetime is uncertain. Yaari [17] and Hakansson [8] have considered this possibility.

The character of these studies is strongly influenced by two major assumptions through which the essential simplicity of Phelps' model is retained. One is that the utility of each period's consumption is independent of past and future consumption. Thus, lifetime utility is the sum of periodic utilities, each a function only of a single-period's consumption plus, possibly, an additional term giving the utility of a bequest. The second is that conversion of assets from one form to another is costless, which implies that the composition of wealth is irrelevant in decision making and, moreover, that in the presence of riskless investment opportunities the individual has no incentive to hold cash as an asset.

Authors who attempt to derive explicit properties of optimal behavior also tend to assume particular one-period utility functions which afford additional analytic convenience. For example, if the utility function exhibits constant absolute risk aversion, the optimal mix of risky assets depends only on the one-period utility function and the joint probability distribution of returns to risky assets. Consequently, the composition of risky assets is determined independently of the amount of wealth or consumption, a property which has been called *separation* or *partial separation*. Additionally, the composition of risky assets is a short-run, or myopic, problem in that it does not depend on any past or future events. When relative risk aversion is constant, the optimal mix of risky *and* riskless assets is independent of wealth, exhibiting *complete separation*, and the composition of assets problem is also short-run. Given limited wealth, however, in both cases the consumption decision remains as a long-run problem since today's optimal consumption depends on the distribution of wealth over consumption in future periods. For a more complete discussion of these points the reader is referred to various papers mentioned previously [6, 7, 9, 11, 12, 16].

In the present paper, to keep the exposition reasonably manageable, and to focus on the role of transaction costs, we introduce a number of simplifications, including some mentioned above. We assume a constant relative risk averse one-period utility function and confine explicit attention to a two-period horizon. We only briefly suggest the form of the n-period outcomes if the horizon is extended. Since it is now necessary to examine portfolio choices in detail, we limit portfolio opportunities to two assets. One is a riskless asset used for consumption transactions, which we refer to as cash, and the other is a risky asset with a random return. Wealth is taken to be the sum of cash and the risky asset at the beginning of a period. Income is assumed to be zero, or included in initial wealth.

In this framework we consider two types of transaction cost. The first is a proportional cost and the second a fixed, or lump sum, cost for purchases or sales of the risky asset. With both costs the optimal portfolio mix is sensitive to the composition of wealth, and it is also apparent that, independently of the form of the utility function, portfolio choice is now a long-run problem since the frequency and magnitude of transactions, and thus the lifetime average expected transaction cost per period, depend on the length of the horizon. A distinguishing feature is that with the first the portfolio mix is independent of wealth, as in the work of Hakansson and others, while with the second the mix directly depends on the magnitude of wealth. Other features of the outcomes are given as the argument develops.

The case of proportional transaction cost has been developed at length elsewhere by the second author [18] and we only summarize that paper here to facilitate comparisons. Consequently, major attention is given to examining the impact of fixed transaction cost. The plan of the paper is the following. In section 6.2 we present the model of consumer choice and outline behavior when transaction cost is proportional. In sections 6.3 and 6.4 we analyze the portfolio and consumption problems in the presence of fixed transaction cost. Section 6.5 is a concluding section.

6.2. A Model of Consumption and Portfolio Choice

The specific model may be described as follows. Given $U(c_1, c_2)$ as utility from consumption in the two periods, then $U(c_1, c_2) = u(c_1) + \alpha u(c_2)$ where $u(c)$ is the one-period utility function and α is the individual's utility discount factor. As a particular utility function we choose

$u(c) = c^\lambda, 0 < \lambda < 1$, where $(1 - \lambda)$ is the constant measure of relative risk aversion. The reader may easily extend the analysis to situations where $u(c) = -c^\lambda, \lambda < 0$ and $u(c) = \ln c$, the remaining members of the family of utility functions with constant relative risk aversion. Similarly, it would not be difficult to consider more general cases such as the HARA (hyperbolic absolute risk aversion) family of utility functions which, among others, includes the functions with constant relative risk aversion [12]. We omit the details here.

The individual wishes to maximize the expectation $EU(c_1, c_2)$ subject to non-negativity of consumption, each asset, and wealth. Initial wealth is $w = x + y$ where x is cash and y gives the initial amount of the risky asset valued at its current price. Assuming decisions are made at the beginning of a period, the current transaction cost is $b|Y - y|$ where b, which lies in the interval (0, 1), is the cost of buying or selling one dollar's worth of risky asset and Y is the value of the risky asset after the current transaction. The random value of a dollar's worth of risky asset at the end of a period is given by $0 \leq a \leq \beta \leq A \leq \infty$ where β has some continuous distribution function $\Phi(\beta)$ with mean value $\bar{\beta} > 1$, $a < 1$ and $A > 1$.

Using the definition of wealth to obtain cash as the difference between wealth and risky asset holding, we may now derive next period's initial assets and wealth. Obviously, asset holding becomes βY while cash is given by $[w - y - c - b\,|Y - y| - (Y - y)]$ or $[w - c - Y - b\,|Y - y|]$ where, subsequently as well, we omit the subscript on today's consumption. Thus, wealth next period is the amount $[w - c - Y - b\,|Y - y| + \beta Y]$ or $[w - c + (\beta - 1)Y - b\,|Y - y|]$. Since borrowing or lending is not allowed, we require next period's cash to be non-negative, which clearly implies the non-negativity of next period's wealth.

In the two-period problem the consumer obviously consumes the cash value of his wealth in the second period, $[w - c + (\beta - 1)Y - b\,|Y - y| - b\beta Y]$ or $[w - c + (\beta(1 - b) - 1)Y - b\,|Y - y|]$. Thus, defining $f_2(w, y)$ as the maximum expected utility over the two periods, the functional equation representing the problem is given by

$$f_2(w, y) = \max_{c,\,Y} \left\{c^\lambda + \alpha \int_a^A [w - c + (\beta(1 - b) - 1)Y - b\,|Y - y|]^\lambda d\Phi(\beta)\right\}, \quad (6.1)$$

where $c \geq 0$, $Y \geq 0$ and $[w - c - Y - b\,|Y - y|] \geq 0$.

In analyzing eq. (6.1), it is useful to identify $(w - c)$ as the consumer's investment fund, the portion of wealth available for distribution among cash, risky asset and transaction costs. From the consumption bound, $(w - c)$ is positive if w and y are positive. In fact, $(w - c)$ will be positive even if y equals zero since an examination of marginal utilities easily shows that the optimal consumption plan requires positive consumption in both periods which implies $(w - c) > 0$. Now, multiply and divide the integrand in eq. (6.1) by $(w - c)^{\lambda}$, let $V = Y/(w - c)$ and $v = y/(w - c)$, to obtain

$$f_2(w, y) = \max_{c, V} \{c^{\lambda} + \alpha(w - c)^{\lambda} \int_a^A [1 + (\beta(1 - b) - 1)V - b|V - v|]^{\lambda} d\Phi(\beta)\}. \quad (6.2)$$

With the appropriate modification of constraints, eq. (6.2) is equivalent to eq. (6.1).

Clearly, for each dollar of the investment fund, V is the share invested in risky asset, $b|V - v|$ is the transaction cost share and $(1 - V - b|V - v|)$ is the cash share. Thus, taking the investment fund as given, the portfolio problem consists of the distribution of a dollar, given v as the initial risky asset share of a dollar of investment funds. While v may exceed one, for example, if $(w - c) = by$, the portfolio shares must be non-negative and sum to one after the portfolio choices. To see that cash share is non-negative, note that the cash constraint $(w - c - Y) \geq b|Y - y|$ becomes $(1 - V) \geq b|V - v|$ by the definitions of V and v.

The cash share constraint may be sharpened by taking into account bounds on v and the magnitude of initial cash $(w - y)$ relative to current consumption c. In addition to its non-negativity, $v = y/(w - c)$ has an upper bound determined by maximum consumption $(w - by)$ which gives $0 \leq v \leq 1/b$. Now if initial cash is less than consumption ($w - y < c$ or $v > 1$), a reduction in the risky asset holding is required and, after rearrangement, the cash constraint becomes $V \leq (1 - bv)/(1 - b)$. On the other hand, if initial cash is sufficient ($w - y \geq c$ or $v \leq 1$), it is possible to change the risky asset holding in either direction with the proviso that enough cash remains to finance consumption. Obviously, if the amount of the risky asset is not increased, sufficient cash remains available. An increase, however, is limited by the cash constraint which, in this event, becomes $V \leq (1 + bv)/(1 + b)$. In examining eq. (6.2), then, apart from the definitions of V and v, and the non-negativities of V and c, the relevant constraints are $V \leq (1 - bv)/(1 - b)$ if $1 < v \leq 1/b$ and $V \leq (1 + bv)/(1 + b)$ if $0 \leq v \leq 1$.

To determine portfolio decisions requires examining properties of the integral in eq. (6.2) which we now write as

$$H(V, v) = \int_a^A [1 + (\beta(1 - b) - 1)V - b\,|V - v|]^\lambda \mathrm{d}\Phi(\beta). \qquad (6.3)$$

The function $H(V, v)$ then gives the expected utility of investing one dollar when the stock share after a decision is V and the initial share is v. To shorten the exposition we only explicitly consider the situation when $0 \leq v \leq 1$ and confine attention here, mainly, to an intuitive and diagrammatic explanation of behavior. As noted, proofs are given elsewhere [18]. In this case, V may exceed or fall short of v and the problem now consists of comparing $H(v, v)$, the expected utility if the stock share is unchanged, and $H(V, v)$ for $V \neq v$. Whenever $H(V, v)$ exceeds $H(v, v)$ for some V, we then want to find V which maximizes the expected utility. In this comparison a number of facts about $H(V, v)$ are useful. An immediate observation from eq. (6.3) is that $H(V, v) < H(V, V)$ when $V \neq v$, or in other words, if the risky asset share is already at V, a change to V from some other initial share v must decrease expected utility. Also, $H(v, v)$ is concave in v and $H(V, v)$ is concave in V for fixed v and $V \neq v$.

Another important observation is that $H(V, v)$ is not differentiable in V at the value $V = v$ but does have a right- and a left-hand derivative. If we let $H_V^+(v, v)$ and $H_V^-(v, v)$ respectively represent these derivatives, we then need to recall that an interior optimum at $V = v$ now requires $H_V^+(v, v) \leq 0$ and $H_V^-(v, v) \geq 0$. The reader who wishes to write out these derivatives explicitly will also easily note that $H_V^- \geq H_V^+$ for all $0 < v < 1$.

One final general fact is needed. After a rather lengthy derivation, it is possible to show that the derivatives $H_V^+(v, v)$ and $H_V^-(v, v)$ change sign once at most as v increases, possibly changing from positive to negative but not in the reverse direction. This sign pattern prevents reversals in behavior as v increases. That is, if initially H_V^+ is positive, which means $H_V^- > 0$, it is optimal to increase the risky asset holding, but once H_V^+ becomes negative it remains negative and the optimal risky asset share is unchanged or diminished. Similarly, if $H_V^- > 0$ while $H_V^+ < 0$, the risky asset share is unchanged, but if H_V^- becomes negative it also remains negative, and thereafter it is optimal to decrease this share. These properties of $H(V, v)$ may be used to prove the following theorem.

THEOREM 6.1. *If* $0 \leq v \leq 1$, *the optimal risky asset share* $V^*(v)$ *is uniquely determined with properties* $V^*(v) \geq v$ *if* $0 \leq v \leq v^0$, $V^*(v) = v$ *if* $v^0 \leq v \leq v^{00}$, *and* $V^*(v) \leq v$ *if* $v^0 \leq v \leq 1$, *where* $0 \leq v^0 \leq v^{00} \leq 1$.

The properties of $H(V, v)$ listed above, and their consequence, theorem 6.1, are independent of particular values of parameters. To obtain further characteristics of behavior, for example, to specify whether the intervals for $V^*(v)$ are distinct, requires more exact specification of parameter values. Again, to shorten the exposition we only consider a special case, a more general treatment being given in the cited paper [18]. In particular we assume $\bar{\beta}(1-b) > (1+b)$ and $E[\beta^{\lambda-1}(\beta-1)] < 0$ which ensure $0 < v^0 < v^{00} < 1$ and consider properties of $V^*(v)$ outside the interval $[v^0, v^{00}]$. Now when $0 \leq v < v^0$, both $H_V^+(v, v) > 0$ and $H_V^-(v, v) > 0$ and it is optimal to increase the risky asset share until $H_V(V^*(v), v) = 0$. Similarly, since $H_V^+(v, v)$ and $H_V^-(v, v)$ are negative if $v^{00} < v < 1$, the risky asset share is decreased until $H_V(V^*(v), v) = 0$. Differentiation of these equations with respect to v provides the properties of $V^*(v)$ given in the following theorem.

THEOREM 6.2. *If* $0 \leq v \leq 1$ *and the conditions* $\bar{\beta}(1-b) > (1+b)$ *and* $E[\beta^{\lambda-1}(\beta-1)] < 0$ *are satisfied, then*

$$\frac{\mathrm{d}V^*(v)}{\mathrm{d}v} = \frac{bV^*}{1+bv} < \frac{b}{1+b} \qquad \text{if } 0 \leq v < v^0, \tag{6.4}$$

$$\frac{\mathrm{d}V^*(v)}{\mathrm{d}v} = \frac{-bV^*}{1-bv} > \frac{-b}{1-b} \qquad \text{if } v^{00} < v < 1. \tag{6.5}$$

The inequality in eq. (6.4) implies that the cash share varies directly with v and the inequality in eq. (6.5) that it varies inversely with v. Both $bV^*/(1+bv)$ and $-bV^*/(1-bv)$ are constant, implying that the optimal risky asset and cash shares are linear in v outside the interval $[v^0, v^{00}]$.

In giving an explanation of theorem 6.2, it is useful to consider the amount $(1 + bv)$ in eq. (6.4), which we may interpret as the 'real' value of a dollar in investment fund. In other words, in an interval in which it is optimal to increase the risky asset share, the gain per dollar of having initial risky asset on hand equals the unit transaction cost times the initial risky asset share of an investment dollar. Now, since $V^*(v)$ is linear, the equality in eq. (6.4) then states that the optimal risky asset and cash shares are constant proportions of a real dollar of the investment fund. A similar interpretation applies in the interval $v^{00} < v < 1$. This proportionality of asset holdings to a real dollar of wealth is analogous to previous results mentioned earlier. However, in the interval $[v^0, v^{00}]$, while $dV^*/dv = 1$, a real dollar changes in value from $(1 + bv)$ to $(1 - bv)$ and $V^*(v)$ is then not proportional to this variable. Eq. (6.5) and the appropriate subsequent parts of the theorem also indicate properties of $V^*(v)$ when $1 \leq v < 1/b$.

Figure 6.1 gives an example of the optimal risky asset shares over the entire inverval $0 \leq v \leq 1/b$ when $\bar{\beta}(1-b) > (1+b)$ and $E[\beta^{\lambda-1}(\beta-1)] < 0$. The upper line traces maximum V while the heavy line gives the path of optimal V. The dotted line illustrates the variation in $V^*(v)$ with an increase in the unit transaction cost b. Omitting the details of the computations, the intuitive explanation is clear. In the interval $0 \leq v \leq 1$, where risky asset holding may move in either direction, a rise in transaction costs increases the consumer's inertia by enlarging the interval in which the risky asset share is unchanged. When risky asset holding must be decreased, in the interval $1 < v \leq 1/b$, the effect is to cause this share to decrease even more.

Returning to eq. (6.2) and its constraints, we now need to examine the terms in brackets on the right-hand side of the equation. After substituting the optimal share $V^*(v)$ we rewrite these terms as

$$G(c,w,y) = c^\lambda + \alpha(w-c)^\lambda \int_a^A [1 + (\beta(1-b)-1)V^* - b\,|V^* - v|]^\lambda \mathrm{d}\Phi(\beta), \qquad (6.6)$$

where $G(c, w, y)$ is the expected utility over the two periods given the optimal risky asset share of any investment fund and optimal behavior in the second period. Among other parameters, the marginal expected utility with respect to consumption will depend on the phase of the risky asset share. The relationship between consumption and these phases is seen from fig. 6.1 and the definition of v. For example, if $c < w - (y/v^0)$, then $V^*(v) > v$. Similarly $w - (y/v^0) < c < w - (y/v^{00})$ implies $V^*(v) = v$, and finally $V^*(v) < v$ if $c > w - (y/v^{00})$. Of course consumption must also satisfy $c \leq w - by$, an original bound now implied by the modified constraints.

If we examine the partial derivatives of $G(c, w, y)$, we discover that while $V^*(v)$ is not differentiable at $c = w - (y/v^0)$ or $c = w - (y/v^{00})$, nevertheless, $G(c, w, y)$ is differentiable in c in the entire interval

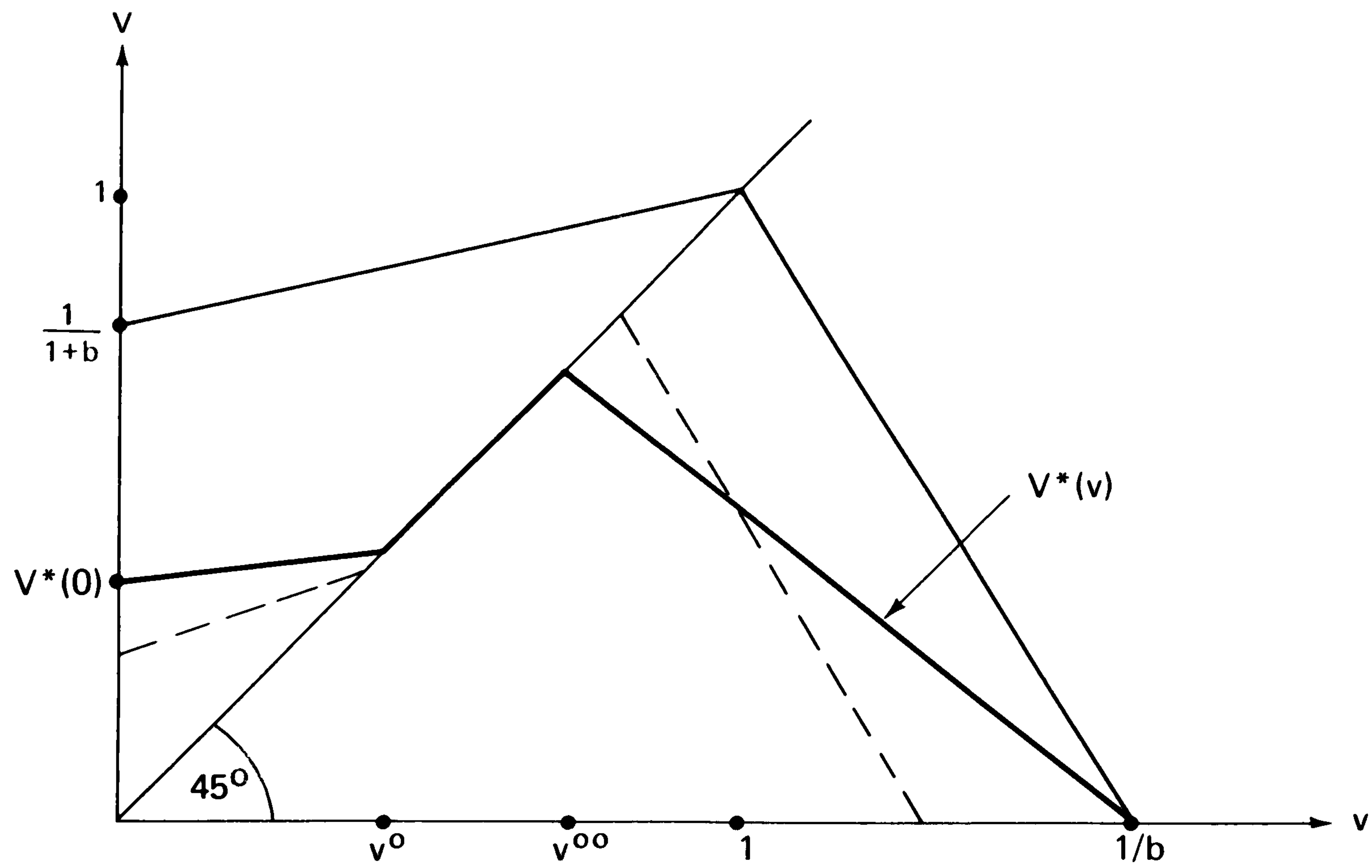

Fig. 6.1. Optimal stock share $V^*(v)$ when $\bar{\beta}(1-b) > (1+b)$ and $E[\beta^{\lambda-1}(\beta-1)] < 0$.

$[0, w - by]$. Moreover, for boundary values we obtain $G_c(0, w, y) = \infty$ and $G_c(w - by, y, y) = -\infty$. Thus, for each w and y, optimal consumption is interior to the constraints $0 \leq c \leq w - by$ and also uniquely determined since $G(c, w, y)$ is a concave function in c. Now, in specifying circumstances in which optimal consumption lies in one interval or another, the composition of wealth is obviously important. In fact, corresponding to v^0 and v^{00}, there exest two critical ratios of y/w, say q^0 and q^{00}, which denote the phases of optimal consumption.

Using this information, and more detailed characteristics of derivatives, the next theorem specifies properties of optimal consumption $c^*(w, y)$ when it lies in the first interval.

THEOREM 6.3. *If* $y/w < q^0$ *then* $c^*(w, y) < w - (y/v^0)$ *and has the properties*

$$\frac{\partial c^*}{\partial y} = b\frac{\partial c^*}{\partial w} = \frac{bc^*}{w + by}. \tag{6.7}$$

Similarly,

$$\frac{\partial Y^*}{\partial y} = b\frac{\partial Y^*}{\partial w} = \frac{bY^*}{w + by}, \tag{6.8}$$

where $Y^*(w, y)$ *is the optimal amount of the risky asset.*

From eq. (6.7) it is seen that the marginal propensity to consume is less than one since, obviously, $c^* < w + by$. To understand the relationships in eqs. (6.7) and (6.8) between partial derivatives with respect to y and w, note that in an interval where it is optimal to increase the risky asset holding, an increase in y of one dollar is equivalent to an increase in wealth of b dollars, or, for example, $\partial c^*/\partial y = b\,\partial c^*/\partial w$.

In the absence of transaction costs, previous studies [7, 8, 11, 12, 16] have shown that consumption and asset holdings are proportional to wealth if relative risk aversion is constant. The analogous result here is given in the final theorem of this section.

THEOREM 6.4. *When $c^* < w - (y/v^0)$, the ratios $c^*/(w + by)$ and $Y^*/(w + by)$ are constant, which implies that optimal consumption and asset holdings are both linear and linear homogeneous in wealth and the initial risky asset.*

PROOF. Differentiating $c^*/(w + by)$ in turn with respect to w and y, using eq. (6.7), gives the constancy of this ratio in both variables. Thus, optimal consumption equals a constant times $(w + by)$, and it is linear and linear homogeneous in wealth and the initial risky asset. An analogous argument applies to cash and the risky asset.

To complete the comparison with earlier results, we may interpret $(w + by)$ as 'real' wealth in an interval in which the risky asset holding is increased. Consequently, theorems 6.3 and 6.4 show that consumption and asset holdings are proportional to real wealth in the initial interval.

Since derivations are similar, we sketch characteristics of behavior in other intervals without further justification. In the final interval, where $V^*(v) < v$, consumption and asset holdings, again, are linear and linear homogeneous in wealth and the initial risky asset and proportional to real wealth $(w - by)$. In the second interval, with $V^*(v) = v$, consumption and risky asset holdings are linear homogeneous but neither linear in their arguments nor proportional to real wealth. Both decision variables change directly with wealth and, as the initial risky asset becomes larger, increase at first and then eventually decrease.

6.3. *Fixed Transaction Cost – Portfolio Choice*

In this section we modify the model to allow fixed, instead of proportional, transaction cost. We now let $K(\cdot)$ represent transaction cost for the purchase or sale of an asset where $K(\cdot) = K$ if the argument is non-zero and $K(\cdot) = 0$ otherwise. For notational convenience we also let $Z = Y - y$ and $\Pi = \beta Y$. The constraints are unchanged with the exception that the cash constraint is replaced by a net cash constraint which is developed as follows.

As in the preceding model, the optimal consumption policy in the second period is to consume the cash value of wealth available in that period. In defining the cash value of wealth, we assume that at the end of the horizon the individual must incur the fixed cost K to settle his estate whenever his risky asset holding is positive. One implication of this assumption is that the second period's net asset value $[\beta Y - K(\Pi)]$ may be negative. As we show later, a consequence of a possible negative net asset value is that the consumer now has some incentive to reduce the risky asset level to zero. Naturally, an alternative assumption is to allow the individual to dispose of the risky asset without transaction cost whenever $\beta Y < K(\Pi)$. However, this assumption violates the spirit of the model, introducing variable rather than fixed transaction cost, and also creates serious difficulties in analysis. We discuss this alternative assumption in more detail in the appendix.

Since we wish to assume that the individual remains solvent, we then require that the cash after the sale of assets – i.e. the net cash at the end of the second period – be non-negative. Thus, we assume

$$[w - Y - c - K(Z) + \beta Y - K(\Pi)] \geq 0 \qquad \text{for all } \beta,$$

or

$$[w - Y - c - K(Z) + aY - K(\Pi)] \geq 0$$

since a is the lowest possible value of β.[1]

Again defining $f_2(w, y)$ as the maximum expected utility over the two periods, we obtain

[1] For all $Y \leq K/a$, this net cash constraint implies that the cash in the beginning of the second period is non-negative. This becomes clear if we rewrite our cash constraint as $[w - Y - c - K(Z)] \geq [K(\Pi) - aY] \geq 0$. If, on the other hand, $Y > K/a$, second period's initial cash might become negative. However, a would presumably be very small, possibly zero, and K large enough so that K/a would be greater than all admissible Y. In any case, hereafter we assume this to be true.

$$f_2(w, y) = \max_{c,\,Y} \{c^\lambda + \alpha \int_a^A [w - c - K(Z) + (\beta - 1)Y - K(\Pi)]^\lambda \mathrm{d}\Phi(\beta)\} \quad (6.9)$$

subject to $c \geq 0$, $Y \geq 0$ and the net cash constraint

$$[w - Y - c - K(Z) + aY - K(\Pi)] \geq 0. \quad (6.10)$$

To make the problem general enough to allow the individual the choice of both increasing and decreasing asset levels, depending on the composition of wealth, we also assume $w > 2K$. As the analysis progresses, the reader may easily deduce the outcomes when $w \leq 2K$.

6.3.1. *Portfolio choice*

Now, given the level of consumption in the first period, it is clear that the portfolio policy depends on the second term of the maximand in eq. (6.9) which gives the expected utility of consumption in the second period, and which we now write as

$$H(Y, y, w - c) = \int_a^A [w - c - K(Z) + (\beta - 1)Y - K(\Pi)]^\lambda \mathrm{d}\Phi(\beta). \quad (6.11)$$

Depending on initial and final assets, eq. (6.11) assumes various forms. In particular, if the asset level is changed,

$$H(Y, y, w - c) = \int_a^A [w - c - 2K + (\beta - 1)Y]^\lambda \mathrm{d}\Phi(\beta) \qquad \text{if } Y \neq y \text{ and } Y \neq 0, \quad (6.12)$$

$$H(0, y, w - c) = [w - c - K]^\lambda \qquad \text{if } Y \neq y \neq 0 \text{ and } Y = 0. \quad (6.13)$$

If the asset level is unchanged,

$$H(y, y, w - c) = \int_a^A [w - c - K + (\beta - 1)y]^\lambda \mathrm{d}\Phi(\beta) \qquad \text{if } y \neq 0, \quad (6.14)$$

$$H(0, 0, w - c) = (w - c)^\lambda \qquad \text{if } y = 0. \quad (6.15)$$

Before examining the expected utility in eq. (6.11), we develop more explicit limitations on portfolio choice imposed by the given level of consumption and parameter values. Choice is limited by two circumstances: the availability of cash on hand to finance current consumption and the availability of assets to finance changes in the portfolio.

Now suppose $y \neq 0$ and consider the possibility of choosing $Y = y$. Then from eq. (6.10), net cash becomes $[w - c - K - (1 - a)y]$ whose sign depends on the given c and parameter values. If it is negative, then, rearranging,

$$(w - y) - (K - ay) < c. \tag{6.16}$$

Thus, given that $(K - ay)$ is needed to guarantee enough cash to settle the estate in the second period, the inequality (6.16) states that there is insufficient cash to finance current consumption and hence the choice $Y = y$ is not possible; in fact, in this case, we must have $Y < y$. Again assuming (6.16) to be true, then for $Y < y$ and $Y \neq 0$, net cash is given by $[w - c - 2K - (1 - a)Y]$ which is non-negative if and only if $(w - c - 2K) \geq (1 - a)Y$. For $Y \neq 0$ to be feasible, this latter inequality requires that

$$(w - c - 2K) > 0, \tag{6.17}$$

since otherwise the net cash constraint is violated. Inequality (6.17) is thus the condition which guarantees that wealth after consumption is large enough to finance two transaction charges.

Next, suppose $y \neq 0$ and $[w - c - K - (1 - a)y] \geq 0$, or

$$(w - y) - (K - ay) \geq c. \tag{6.18}$$

Here, $Y = y$ is obviously feasible. Also, for $Y = 0$, the net cash constraint requires $(w - K) \geq c$ which is seen to be satisfied since (6.18) implies $(w - c - K) \geq (1 - a)y > 0$. Moreover, if $Y \neq y$ and $Y \neq 0$, the net cash becomes $[w - c - 2K - (1 - a)Y]$. By the preceding argument, feasibility of $Y \neq y$ and $Y \neq 0$ again requires that inequality (6.17) be satisfied. Finally, in the situation $y = 0$, inequality (6.17) is, once more, the requirement for the feasibility of $Y \neq 0$.

From this discussion it is apparent that portfolio choice possibilities depend upon the signs of the parameter combinations $[(w - y) - (K - ay) - c]$ and $(w - c - 2K)$. In terms of Y^*, the optimal level of the risky asset, we may summarize the extent of these restrictions as follows.

For $y \neq 0$:

$$\begin{array}{lll}
\text{(a) } Y^* = 0 & \text{if } (w-y)-(K-ay) < c \text{ and } (w-c-2K) \leq 0. \\
\text{(b) } Y^* < y & \text{if } (w-y)-(K-ay) < c \text{ and } (w-c-2K) > 0. \\
\text{(c) } Y^* = 0 \text{ or } Y^* = y & \text{if } (w-y)-(K-ay) \geq c \text{ and } (w-c-2K) \leq 0. \\
\text{(d) } Y^* \gtreqless y & \text{if } (w-y)-(K-ay) \geq c \text{ and } (w-c-2K) > 0.
\end{array} \tag{6.19}$$

For $y = 0$:

$$\begin{aligned} &\text{(a)}\ \ Y^* = 0 \qquad \text{if } (w - c - 2K) \leq 0. \\ &\text{(b)}\ \ Y^* \geq 0 \qquad \text{if } (w - c - 2K) > 0. \end{aligned} \tag{6.20}$$

While optimal behavior is fully determined in (6.19a) and (6.20a), some degree of choice remains in all other situations, varying from two particular asset levels in (6.19c) to the possibility of increasing, decreasing, or not changing the asset level in (6.19d). The task of the next section is to give a complete description of optimal portfolio behavior, taking into account the restrictions in (6.19) and (6.20) and, of course, the non-negativity and net cash constraints. As we show there, it will again be useful to consider the allocation of a dollar of appropriate investment funds.

6.3.2. *Derivation of the portfolio policy*

In this section we present two major categories of outcomes. The first indicates that, independently of parameter values and the given consumption, the optimal portfolio always assumes the same qualitative form. Here we obtain results analogous to theorem 6.1, applying when the transaction cost is proportional. In the second category we derive more detailed characteristics of behavior, paralleling the outcomes in theorem 6.2. However, since we now have to contend with various combinations of parameters, as indicated in (6.19) and (6.20), we limit detailed analysis to one particular case only.

We now suppose (6.19d) applies, i.e. $y \neq 0$, $(w - y) - (K - ay) \geq c$ and $(w - c - 2K) > 0$. As noted, this is a general case in the sense that the risky asset may be increased, decreased, or unchanged. First, rewrite the expected utility expressions in eqs. (6.12) and (6.14) as

$$H(Y, y, w - c) = (w - c - 2K)^{\lambda} \int_a^A [1 + (\beta - 1)V]^{\lambda} \mathrm{d}\Phi(\beta), \tag{6.21}$$

$$H(y, y, w - c) = (w - c - K)^{\lambda} \int_a^A [1 + (\beta - 1)v]^{\lambda} \mathrm{d}\Phi(\beta), \tag{6.22}$$

where

$$V = Y/(w - c - 2K) \qquad \text{and} \qquad v = y/(w - c - K). \tag{6.23}$$

Obviously the divisions performed in eq. (6.23) are valid given the characterization of the present case.

Now, we may interpret $(w - c - 2K)$ as the net investment fund divided between risky asset and cash if $Y \neq y$ and $Y \neq 0$. In other words, it is the investment fund $(w - c)$ less the fixed charge incurred now and the fixed cost due in the next period. Similarly, $(w - c - K)$ is the net investment fund to be divided between cash and risky asset when $Y = y \neq 0$. Therefore, V represents the share of risky asset and $(1 - V)$ the cash share in an investment dollar when $Y \neq y$ and $Y \neq 0$. A similar interpretation applies to v; it is the initial and final share in an investment dollar when $Y = y \neq 0$.

For $Y \neq y$, $Y \neq 0$, the net cash constraint is $[w - c - 2K - (1-a)Y] \geq 0$, from which it follows that $(1 - V + aV) \geq 0$, or

$$0 < V \leq 1/(1 - a). \tag{6.24}$$

Consequently, V may exceed one and hence the cash share $(1 - V)$ may become negative reflecting the fact that for each dollar of the net investment fund there is a guaranteed cash return of aV in the second period. Similarly, if $Y = y \neq 0$, the net cash constraint would be $[w - c - K - (1-a)y] \geq 0$, which gives

$$0 < v \leq 1/(1 - a). \tag{6.25}$$

Our task then is to find the optimal share $V^*(v)$ over V satisfying (6.24) and also $V = 0$, for each v satisfying (6.25).

Now examining eqs. (6.21) and (6.22) for given initial values, the factors $(w - c - 2K)^\lambda$ and $(w - c - K)^\lambda$ are constant terms and the integrals involving V and v are the same function, say $G(\cdot)$. For future reference notice that $G(\cdot)$ is a strictly concave function and that $G(0) = 1$.

Differentiating $G(V)$,

$$G'(V) = \lambda \int_a^A \{[1 + (\beta - 1)V]^{\lambda - 1}(\beta - 1)\} \mathrm{d}\Phi(\beta), \tag{6.26}$$

which gives $G'(0) = \lambda(\bar{\beta} - 1) > 0$ and $G'[1/(1 - a)] = \lambda[1/(1 - a)]^{\lambda - 1} E[(\beta - a)^{\lambda - 1}(\beta - 1)]$ whose sign depends only on the distributions of β and the risk aversion index $(1 - \lambda)$. If it is non-negative, $G(V)$ attains a boundary maximum at $V = 1/(1 - a)$, while $G'[(1/(1 - a)] < 0$ assures an interior maximum.

With this basic information we are now in a position to prove a theorem about the form of the optimal portfolio policy. In stating and proving the theorem we identify $\bar{V}$ as the unique V maximizing eq. (6.21) subject to (6.24), let $V^0 = 0$ or $V^0 = \bar{V}$, depending on which of the two risky asset shares provides the larger expected utility, and let

$\bar{Y} = \bar{V} \cdot (w - c - 2K)$ and $Y^0 = V^0 \cdot (w - c - 2K)$. Also, we use the notation (v^0, V^0, v^{00}), with $0 \leq v^0 \leq V^0 \leq v^{00} \leq 1/(1 - a)$, to specify the following policy: if $v < v^0$ change the risky asset share to V^0, if $v^0 \leq v \leq v^{00}$ do not change the risky asset level, and if $v^{00} \leq v$ again change the share to V^0.

THEOREM 6.5. *Assuming* $y \neq 0$, $(w-y)-(K-ay) \geq c$ *and* $(w-c-2K)>0$, *the optimal risky asset share* $V^*(v)$ *is uniquely determined with form* (v^0, V^0, v^{00}). *Analogous results hold for the optimal cash share* $[1 - V^*(v)]$.

PROOF. We prove the theorem by considering all possible cases which may arise, given the conditions of the theorem. While a simpler proof may be devised, the method used here has the compensating virtue of directly specifying the computation of the policy parameters v^0, V^0 and v^{00} in all circumstances, information needed later in any event.

First, we note that for a given investment fund $(w - c)$, the graphs of $H(y, y, w - c)$ and $H(Y, y, w - c)$ are obtained from the graph of $G(\cdot)$. In fact, the graph of $H(y, y, w - c)$ is just a blown up or scaled down version of that of $G(v)$ depending on whether $(w - c - K)^\lambda \gtrless 1$. The same is true of $H(Y, y, w - c)$, depending on whether $(w - c - 2K)^\lambda \gtrless 1$. However, since $(w - c - 2K)^\lambda < (w - c - K)^\lambda$, it is clear that the graph of $H(Y, y, w - c)$ will always lie below that of $H(y, y, w - c)$.

Figure 6.2 now shows the case in which $G'[1/(1-a)]<0$; $H(0, y, w-c) < H(\bar{Y}, y, w - c)$, or $(w - c - K)^\lambda < (w - c - 2K)^\lambda G(\bar{V})$; and

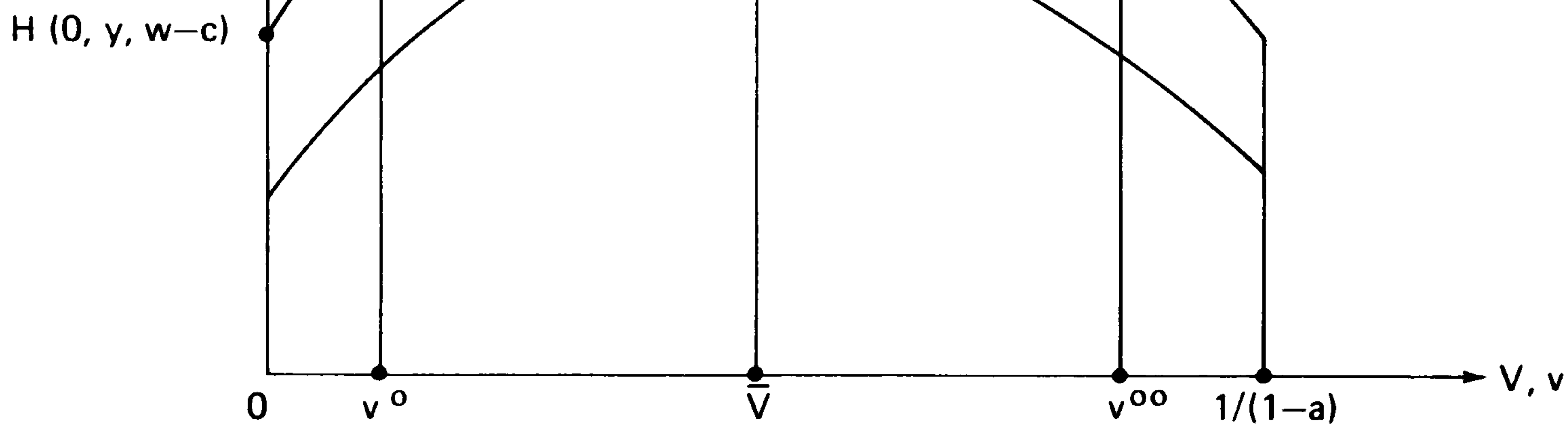

Fig. 6.2. Derivation of $V^*(v)$ when $G'[1/(1-a)] < 0$, $(w-c-K)^\lambda < (w-c-2K)^\lambda G(\bar{V})$ and $(w-c-2K)^\lambda G(\bar{V}) > (w-c-K)^\lambda G[1/(1-a)]$.

$H(\bar{Y}, y, w - c) > H[w - c - K/(1 - a), w - c - K/(1 - a), w - c]$ or $(w - c - 2K)^{\lambda}G(\bar{V}) > (w - c - K)^{\lambda}G[1/(1 - a)]$. As seen in the figure, $\bar{V}$ maximizes $H(Y, y, w - c)$ and the two points v^0 and v^{00} satisfy

$$(w - c - 2K)^{\lambda}G(\bar{V}) = (w - c - K)^{\lambda}G(v^0) \text{ and } G'(v^0) > 0, \qquad (6.27)$$

$$(w - c - 2K)^{\lambda}G(\bar{V}) = (w - c - K)^{\lambda}G(v^{00}) \text{ and } G'(v^{00}) < 0. \qquad (6.28)$$

Now, given $H(0, y, w - c) < H(\bar{Y}, y, w - c)$, the optimal risky asset share $V^*(v)$ has the form (v^0, V^0, v^{00}) since when $v < v^0$, expected utility is increased by changing v to $\bar{V} = V^0$; when $v^0 \leq v \leq v^{00}$ expected utility is maximized by setting $V^*(v) = v$ and, finally, when $v^{00} < v$ it is again optimal to change the risky asset share to $\bar{V}$. The case we have analyzed is one in which $0 < v^0 < V^0 < v^{00} < 1/(1 - a)$.

Consider other possibilities. Suppose $G'[1/(1 - a)] < 0, (w - c - K)^{\lambda} < (w - c - 2K)^{\lambda}G(\bar{V})$ but $(w - c - 2K)^{\lambda}G(\bar{V}) < (w - c - K)^{\lambda}G[1/(1 - a)]$. By appropriately modifying fig. 6.2 it is easily seen that the form of optimal behavior is (v^0, V^0, v^{00}) with $0 < v^0 < V^0 = \bar{V} < v^{00} = 1/(1 - a)$ where v^0 again satisfies eq. (6.27). Next, suppose $G'[1/(1 - a)] < 0, (w - c - K)^{\lambda} > (w - c - 2K)^{\lambda}G(\bar{V})$ and $(w - c - K)^{\lambda} > (w - c - K)^{\lambda}G[1/(1 - a)]$. It now follows that the optimal policy is (v^0, V^0, v^{00}) where $v^0 = V^0 = 0$ and v^{00} satisfies

$$(w - c - K)^{\lambda} = (w - c - K)^{\lambda}G(v^{00}), \qquad v^{00} \neq 0 \qquad (6.29)$$

or, $G(v^{00}) = 1$, $v^{00} \neq 0$. The remaining possibilities consist of one more case in which $G'[1/(1 - a)] < 0$ and two in which $G'[1/(1 - a)] \geq 0$, all yielding optimal policies of the form (v^0, V^0, v^{00}). Omitting these details, the proof of the theorem is now complete. As noted earlier, the proof not only gives the optimality of the policy (v^0, V^0, v^{00}) but also specifies the computation of the policy parameters.

Our next task is to derive the optimal portfolio policy in terms of absolute quantities rather than amounts per investment dollar. This translation is not completely apparent because of the different investment funds used in defining v and V.

THEOREM 6.6. *Given the conditions of theorem 6.5, the optimal risky asset amount $Y^*(y)$ is uniquely determined with form (y^0, Y^0, y^{00}) where $Y^0 = V^0 \cdot (w - c - 2K)$, $y^0 = v^0 \cdot (w - c - K)$ and $y^{00} = v^{00} \cdot (w - c - K)$.*

PROOF. The proof consists of showing that $y^0 \leq Y^0 \leq y^{00}$, for then it is clear that $Y^*(y)$ is described by the policy (y^0, Y^0, y^{00}). We provide a proof only for the case described in fig. 6.2, omitting similar derivations for other combinations. This case is one in which $y^0 < Y^0 = \bar{Y} < y^{00}$. Obviously $\bar{Y} < y^{00}$; to show that $\bar{Y} > y^0$, we have

$$\int_a^A [w-c-K+(\beta-1)y^0]^\lambda \mathrm{d}\Phi(\beta) = \int_a^A [w-c-2K+(\beta-1)\bar{Y}]^\lambda \mathrm{d}\Phi(\beta), \quad (6.30)$$

which implies that

$$\int_a^A [w-c-K+(\beta-1)y^0]^\lambda \mathrm{d}\Phi(\beta) < \int_a^A [w-c-K+(\beta-1)\bar{Y}]^\lambda \mathrm{d}\Phi(\beta). \quad (6.31)$$

Now since, for $v \leq v^0$, or $y \leq y^0$, fig. 6.2 implies

$$\int_a^A [w-c-K+(\beta-1)y^0]^\lambda \mathrm{d}\Phi(\beta) \geq \int_a^A [w-c-K+(\beta-1)y]^\lambda \mathrm{d}\Phi(\beta), \quad (6.32)$$

we must have $\bar{Y} > y^0$, completing the proof of the theorem[2].

It is now reasonably straightforward to show that the policies (v^0, V^0, v^{00}) and (y^0, Y^0, y^{00}) are optimal in all other situations listed in eqs. (6.19) and (6.20) and to specify values of policy parameters. We leave the verification of these outcomes to the reader and turn to a final exercise in this section. One implicit result of the previous analysis is that, since the factors $(w - c - 2K)$ and $(w - c - K)$ enter directly in the derivation of optimal behavior, the optimal portfolio mix is not independent of initial wealth. In other words, while $\bar{V}$ depends only on the sign of $G'[1/(1 - a)]$, the parameters of the policies (v^0, V^0, v^{00}) and (y^0, Y^0, y^{00}) depend on the investment fund $(w - c)$ and the fixed cost K. Again considering the case described in fig. 6.2, we briefly examine the sensitivity of policy parameters to changes in $M = (w - c)$ and K.

Clearly, since $\bar{Y} = \bar{V}\cdot(w - c - 2K)$, $\partial\bar{Y}/\partial M > 0$ and $\partial\bar{Y}/\partial K < 0$. Now differentiating eqs. (6.27) and (6.28) with respect to M, using these equations again, and collecting terms,

$$\frac{\partial v^0}{\partial M} = \frac{\lambda K(M-2K)^{\lambda-1}G(\bar{V})}{(M-K)^{\lambda+1}G'(v^0)} > 0, \quad \frac{\partial v^{00}}{\partial M} = \frac{\lambda K(M-2K)^{\lambda-1}G(\bar{V})}{(M-K)^{\lambda+1}G'(v^{00})} < 0. \quad (6.33)$$

[2] A modification of the proof of theorem 6.6 also yields the intuitive result that if the initial risky asset holding is small enough, it is never optimal to decrease its amount. That is, if $y < K$, then $Y^*(y) \geq y$.

Finally, differentiating with respect to K we also obtain

$$\partial v^0/\partial K < 0 \qquad \text{and} \qquad \partial v^{00}/\partial K > 0. \tag{6.34}$$

These results have an intuitive interpretation, hinging on the relative cost of changing asset levels. That is, the inequalities in (6.33) show that the larger is wealth net of current consumption, relative to the fixed cost, the smaller is the interval in which asset holdings remain unchanged. The reverse behavior is observed in (6.34) when the fixed cost rises relative to the investment fund. Analogous results apply in terms of y^0 and y^{00}.

We also need to examine events which may lead to large jumps in policy parameters. An examination of eq. (6.19) shows, for example, that as M becomes sufficiently small, we will shift from case (6.19d) to cases (6.19c), (6.19b) or (6.19a) with, possibly, large changes in one or more of the parameters v^0, V^0 and v^{00}. We consider this possibility in more detail in the next section.

6.4. Fixed Transaction Cost – the Consumption Policy

To determine the consumption decision we examine the maximand in eq. (6.9) which, after substituting the optimal asset level, becomes

$$g(w, y, c) = c^\lambda + \alpha \int_a^A [w-c-K(Z) + (\beta-1)Y^* - K(\Pi)]^\lambda \mathrm{d}\Phi(\beta). \tag{6.35}$$

The expression in eq. (6.35) gives the expected utility over the two periods if portfolio choice and behavior in the second period are optimal. The problem is to maximize eq. (6.35) with respect to c satisfying

$$[w - c - Y^* - K(Z) + aY^* - K(\Pi)] \geq 0 \qquad \text{and} \qquad c \geq 0. \tag{6.36}$$

In situations in which $Y^* = 0$, we see from (6.36) that the maximum value of c is $(w - K)$ if $y \neq 0$ and w if $y = 0$.

6.4.1. Nature of the consumption choice problem

The first task here is to examine how variations in c cause changes in the optimal portfolio policy. Now, suppose $y \neq 0, w - y - (K - ay) > 0$ and $(w - 2K) > 0$, which satisfy the conditions of theorem 6.5 for small c, and suppose also that the conditions in fig. 6.2 are met. In this situation, by the previous analysis, $v < v^0 < v^{00}$ for $c = 0$, and as c increases, v^0 decreases to zero but v increases beyond bounds. Hence there exists a c^0 such that, for all $c \leq c^0$, the optimal portfolio policy is to choose $Y^* = \bar{Y}$. Moreover, as c increases, v^{00} increases to $1/(1 - a)$ or the v^{00} satisfying eq. (6.29), depending on whether $(w - c - K)^\lambda$ is less than or exceeds $(w - c - K)^\lambda G[1/(1 - a)]$. Consequently, there exists a c^{00} such that $Y^* = \bar{Y}$ or $Y^* = 0$ for all $c \geq c^{00}$ and $Y^* = y$ for $c^0 \leq c \leq c^{00}$. We also note that, in this case, v^{00} attains its limit and v^0 equals zero at $c = w - 2K$ and at the value c^{00}, $Y^* = \bar{Y}$ or $Y^* = 0$, depending on whether v intersects v^{00} before or after it reaches its maximum amount. Fig. 6.3 illustrates the identification of c^0 and c^{00}.

In fig. 6.3 the points c^0 and c^{00} are marked at the intersections of v with v^0 and v^{00}. The cases of eq. (6.19) are also clearly indicated in the diagram. For example, $v < 1/(1 - a)$ and $c < w - 2K$ mean case (6.19d) prevails and $Y^0 = \bar{Y}$. As c increases, we enter case (6.19c) where $v < 1/(1 - a)$ but $c \geq w - 2K$, which here implies $Y^0 = 0$ since v exceeds maximum v^{00}. In this diagram, then, $w - 2K$ is also a critical

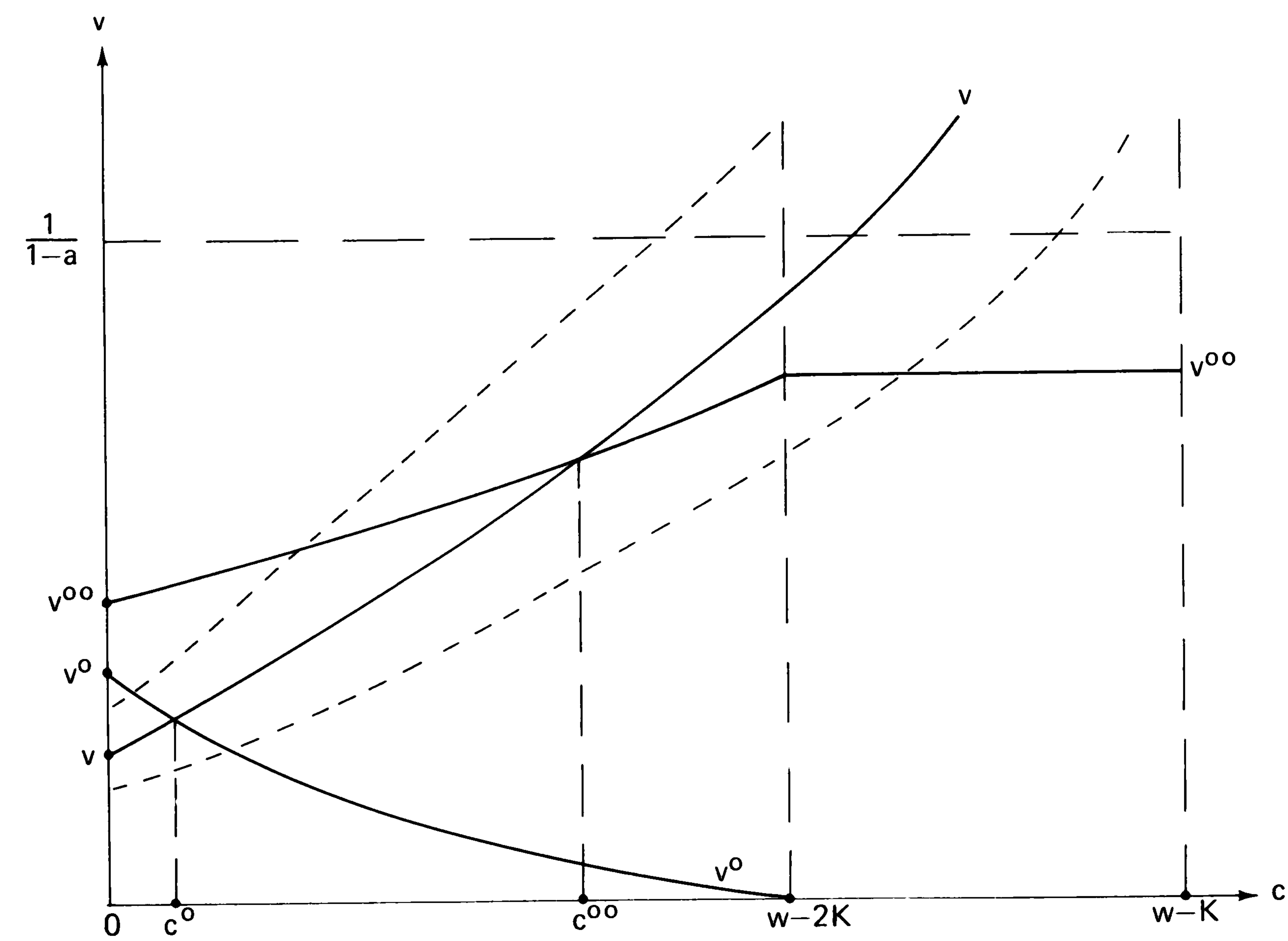

Fig. 6.3. Derivation of c^0 and c^{00} when the conditions in fig. 6.2 are satisfied and $(w-c-K)^\lambda > (w-c-K)^\lambda G[1/(1-a)]$.

point at which Y^* jumps from $\bar{Y}$ to 0. Finally, as c increases further, we enter the realm of (6.19a) where $v > 1/(1-a)$ and $c > w - 2K$ and, again, $Y^0 = 0$. Case (6.19b) does not arise since the combination $v > 1/(1-a)$ and $c < w - 2K$ is precluded in this situation. The dotted lines in fig. 6.3 indicate other possible positions of the v curve for other values of y, keeping w and K fixed. The reader may easily deduce the implications of these curves.

Omitting details, it is also possible to identify corresponding critical levels of c for all other combinations of parameters. Thus, the admissible values of c may always be subdivided into intervals such that the phase of the optimal portfolio policy is identified.

We now investigate conditions under which optimal consumption would lie in one or the other intervals. It will be seen that this is a difficult task, and we have not succeeded in obtaining a full characterization of such conditions. For this purpose, we examine the derivatives of eq. (6.35) with respect to c for the three alternative choices for Y^*. Thus

$$g_c(w,y,c) = \lambda c^{\lambda-1} - \alpha\lambda(w-c-2K)^{\lambda-1}G(\bar{V}) \qquad \text{if } Y^* = \bar{Y} \neq y, \tag{6.37}$$

$$g_c(w,y,c) = \lambda c^{\lambda-1} - \alpha\lambda(w-c-K)^{\lambda-1}\int_a^A [1+(\beta-1)v]^{\lambda-1}\mathrm{d}\Phi(\beta) \qquad \text{if } Y^* = y \neq 0, \tag{6.38}$$

and

$$g_c(w,y,c) = \lambda c^{\lambda-1} - \alpha\lambda(w-c-K(y))^{\lambda-1} \qquad \text{if } Y^* = 0. \tag{6.39}$$

Now consider again the situation described in fig. 6.3. Here, for $0 \leq c \leq c^0$, $g_c(w,y,c)$ is given by eq. (6.37), whereas for $c^0 < c < c^{00}$ the relevant derivative is eq. (6.38). It is then clear that there is a discontinuity in $g_c(w,y,c)$ at $c = c^0$, but the left and right-hand derivatives g_c^+ and g_c^- exist and are available in eqs. (6.37) and (6.38) respectively. Again, at $c = c^{00}$ there is a discontinuity with the left-hand derivative given by eq. (6.38) and the right-hand derivative by eq. (6.37). A discontinuity in derivatives arises at one other value of c. As noted, in explaining fig. 6.3, $c = w - 2K$ is a critical point at which Y^* jumps from $\bar{Y}$ to 0. This jump leads to a third discontinuity[3].

Thus, even though $g_{cc}(w,y,c)$ is negative at all points interior to subintervals, $g_c(w,y,w-K) = -\infty$, and $g_c(w,y,0) = +\infty$, a possibility

[3] It is possible to show that, at $c = c^0$ and $c = w - 2K$, the right-hand derivative exceeds the left-hand derivative. However, at $c = c^{00}$, the direction of inequality, in comparing these derivatives, appears to be indeterminate.

remains of multiple local maxima of $g(w, y, c)$. In other words, we might have $g_c^-(w, y, c^0) < 0$ but $g_c^+(w, y, c^0) > 0$, for example. In that case there would be one local maximum interior to the first subinterval and at least one more for $c > c^0$. Consequently, a direct comparison between the different local maxima of $g(w, y, c)$ is needed to determine the global maximum.

In itself, the existence of discontinuities in derivatives is not particularly troublesome in determining optimal consumption for any given set of parameters. What is a matter of concern is that the possible need to examine various local maxima compounds the difficulty in identifying combinations of parameters which lead to optimal consumption lying in a particular interval. As noted earlier, in the case of proportional transaction cost, this identification was easily accomplished by determining critical ratios of initial risky asset to wealth. Here, however, the previous discussion indicates clearly that the level of initial risky asset, as well as its ratio with wealth, is important in determining optimal consumption. In other words, for example, in the case of proportional transaction cost, optimal consumption lies in the initial interval whenever $y/w < q^0$, where q^0 is an easily determined constant. In the case of fixed transaction cost, however, there would exist a real valued transformation, say, $q(y, y/w)$ and a real number q^0 such that optimal consumption lies in the initial interval whenever $q(y, y/w) < q^0$. As yet, partly because of discontinuities in derivatives, we have not succeeded in specifying properties of the transformation $q(y, y/w)$.

Keeping this unresolved problem in mind, in the next section we consider further properties of optimal consumption.

6.4.2. Properties of optimal consumption

First, it is easily seen that $c^*(w, y)$ is interior to the interval $[0, w - K]$ since $g_c(w, y, c) \to -\infty$ as $c \to w - K$ and $g_c(w, y, c) \to \infty$ as $c \to 0$.

We now examine properties of $c^*(w, y)$ when it lies in one interval or another. For this purpose, define

$$B = [\alpha G(\bar{V})]^{1/(\lambda-1)}/\{1 + [\alpha G(\bar{V})]^{1/(\lambda-1)}\}.$$

THEOREM 6.7. *If $c^*(w, y)$ lies in either of the intervals where the optimal portfolio policy is to change the initial asset to $Y^* = \bar{Y}$, then*

$$c^* = B \cdot (w - 2K), \tag{6.40}$$

$$Y^* = (1 - B) \cdot \bar{V} \cdot (w - 2K), \tag{6.41}$$

$$0 < \frac{\partial c^*}{\partial w} = \frac{c^*}{w - 2K} < 1. \tag{6.42}$$

PROOF. Under the condition of the theorem, $c^*(w, y)$ must satisfy

$$g_c(w, y, c^*) = \lambda c^{*\lambda - 1} - \alpha\lambda(w - c^* - 2K)^{\lambda - 1} G(\bar{V}) = 0.$$

Hence, solving for c^*, we obtain $c^* = B \cdot (w - 2K)$. Now since $Y^* = \bar{V} \cdot (w - c^* - 2K)$, we have $Y^* = (1 - B) \cdot \bar{V} \cdot (w - 2K)$. Differentiating eq. (6.40), and using eq. (6.40) again, $\partial c^*/\partial w = c^*/(w - 2K) > 0$. Finally, since $Y^* = \bar{Y}$ implies $c < w - 2K$, we find that $c^*/(w - 2K) < 1$, or that the marginal propensity to consume wealth is positive but less than one, which completes the proof.

Equation (6.40) also states that optimal consumption is proportional to real wealth $(w - 2K)$. In other words, since $Y^* \neq 0$ and $Y^* \neq y$, the consumer incurs a cost of $2K$ consisting of a transaction cost payment now and in the next period; hence the real worth of wealth is $(w - 2K)$ and optimal consumption is proportional to real wealth. A similar interpretation applies to investment in the risky asset and cash. It is also obvious from eq. (6.40) that $\partial c^*/\partial y = 0$. That is, in intervals where the optimal portfolio choice is to change the initial asset to a non-zero level, optimal consumption does not depend directly on the initial composition of wealth. However, of course, as we argued previously, these intervals of consumption do depend on y.

We next turn to other intervals of optimal consumption and state the corresponding results there without giving explicit proofs. If $c^*(w, y)$ lies in a final interval implying a change in the risky asset level to zero, it has properties analogous to those listed in theorem 6.7 with real wealth being $(w - K)$. In the interval of no change in the risky asset level, we obtain that $c^*(w, y)$ varies directly with wealth, the marginal propensity to consume wealth is again positive and less than one, $c^*(w, y)$ varies directly with the initial risky asset at first and then inversely as that asset level increases. Finally, we are able to show that, in this interval, $c^*(w, y)$ is linear homogeneous in the initial risky asset and re. wealth which, again, is $(w - K)$.

6.5. Conclusions

Our results clearly indicate that the introduction of transaction costs changes the character of the individual's consumption and portfolio choices over time, modifying available research in the direction of greater realism. The two types of cost we have studied would appear to capture the major features of transactions costs. The analysis of proportional transaction cost shows the importance of the composition of wealth in decision making and suggests that portfolio, as well as consumption, choice is a long-run decision. However, proportional transaction cost does not remedy other limitations of the Phelps type models discussed earlier. Specifically, another characteristic of these models is that the consumer usually invests in a wide range of portfolio opportunities, often, in all opportunities if selling short and borrowing is allowed. Similarly, the consumer tends to be indifferent between investing in individual portfolio opportunities and investing in a suitably chosen mutual fund consisting of a combination of assets. These same features would be observed in the presence of a proportional transaction cost since then the total transaction cost depends only on the volume of trading and not on the number of transactions.

Though our analysis has not as yet proceeded far enough, the introduction of fixed transaction cost should help to explain why portfolios tend to have only a limited number of assets and why consumers may prefer investing in mutual funds rather than in individual assets. With a fixed payment per transaction, it seems clear that, even with the availability of an arbitrary number of possible asset choices, the number of assets in an optimal portfolio would be sensitive to the magnitude of the fixed charge. Moreover, the choice between investing in individual assets and in mutual funds would now seem to hinge critically on comparative fixed payments. Dependent on the relative magnitudes of these payments, there would seem to be some tendency for the risk averse consumer to prefer mutual funds. In general the form of these possible outcomes seems to be reasonably consistent with observed individual behavior and suggests the potential usefulness of examining the force of fixed, as well as proportional, transaction cost in the theory of the consumer.

As a final comment we briefly discuss the possibility of extending the analysis to multi-period horizons. In the case of fixed transaction cost, a critical need is to obtain properties of the transformation $q(y, y/w)$.

Otherwise, it is not possible to specify sufficient properties of the expectation $f_2(w, y)$ to continue an induction. Moreover, even given properties of $q(y, y/w)$, earlier papers introducing fixed costs suggest that additional assumptions about probability distributions would be needed in multiperiod models [1, 2, 4, 13, 15].

In the case of proportional transaction cost, however, it seems clear that without additional assumptions an induction to an arbitrary finite horizon would be reasonably straightforward, yielding the anticipated outcome that optimal asset holding as well as consumption is now a long-run decision. In particular, since the average expected transaction cost per period should decrease as the horizon lengthens, the character of results should resemble a pattern given in fig. 6.1. Recalling the definitions of the dotted and heavy lines, the conjecture is that in a multiperiod setting the heavy line would represent the optimal risky asset share $V_n^*(v)$ in the first period of an n-period horizon and the dotted line the share $V_{n-1}^*(v)$ in the first period of an $(n - 1)$-period horizon (with the proviso that the horizontal intercept of the dotted line should be $1/b$, which is constant in this interpretation). Corresponding results should apply to optimal consumption and risky asset holding.

References

[1] Chitre, V. *A dynamic programming model of demand for money*. University of Toronto (mimeographed).

[2] Eppen, G. D. and Fama, E. F. Cash balance and simple dynamic portfolio problems with proportional costs. *International Economic Review* (June 1969), 119–133.

[3] Fama, E. F. Multiperiod consumption – investment decisions. *The American Economic Review* (March 1970), 163–174.

[4] Girgis, N. M. Optimal cash balance levels. *Management Science* (November 1968), 130–140.

[5] Hahn, F. H. Savings and uncertainty. *The Review of Economic Studies* (January 1970), 21–24.

[6] Hakansson, N. H. *Optimal investment and consumption strategies for a class of utility functions*. PhD dissertation, University of California at Los Angeles (1966).

[7] Hakansson, N. H. Optimal investment and consumption strategies under risk for a class of utility functions. *Econometrica* (September 1970), 587–607.

[8] Hakansson, N. H. Optimal investment and consumption strategies under risk, an uncertain lifetime, and insurance. *International Economic Review* (October 1969), 443–466.

[9] Leland, H. E. *Dynamic portfolio theory*. PhD dissertation, Harvard University (May 1968).

[10] Levhari, D. and Srinivasan, T. N. Optimal savings under uncertainty. *The Review of Economic Studies* (April 1969), 153–163.

[11] Merton, R. C. Lifetime portfolio selection under uncertainty: the continuous time case. *The Review of Economics and Statistics* (August 1969), 247–257.
[12] Merton, R. C. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory* (December 1971), 373–413.
[13] Neave, E. The stochastic cash balance problem with fixed costs for increases and decreases. *Management Science* (March 1970), 472–490.
[14] Phelps, E. S. The accumulation of risky capital: a sequential utility analysis. *Econometrica* (January 1962), 122–136.
[15] Roberts, S. M. A stochastic cash balance inventory model with non-zero fixed and proportional costs and proportional opportunity and penalty costs. *Working Paper No. 13*, Special Studies Paper – Division of Research and Statistics, Federal Reserve Board, Washington DC (January 1971).
[16] Samuelson, P. A. Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics* (August 1969), 239–246.
[17] Yaari, M. E. Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies* (April 1965), 137–150.
[18] Zabel, E. Consumer choice, portfolio decisions and transaction costs. *Econometrica* (March 1973), 321–336.

APPENDIX

An alternative formulation of the fixed transaction cost model

In this appendix we shall describe an alternative formulation of the fixed transaction cost model and suggest the difficulties arising in its analysis. The basic description of the problem is the same as in the text except that we now allow the consumer to dispose of the risky asset without transaction cost whenever $\beta Y < K$. However, we still do not permit borrowing and lending and hence require that cash holding in the beginning of the second period be non-negative.

If we define $L_t(w, y)$ as the maximum expected utility over a t-period horizon, then in the two-period case

$$L_2(w, y) = \underset{c,\,Y}{\text{Max}}\,\{c^\lambda + \alpha \int_a^A L_1[w - Y - c - K(Z); \beta Y]\,d\Phi(\beta) \tag{A6.1}$$

subject to $c, Y \geq 0$ and $[w - c - Y - k(Z)] \geq 0$.

Again, given that the individual will consume the cash value of his wealth in the second period, we may derive an expression for $L_1[w - Y - c - K(Z); \beta Y]$:

$$L_1[w - Y - c - K(Z); \beta Y] = \begin{cases} [w-Y-c-K(Z) + \beta Y-K]^{\lambda} & \text{if } \beta Y > K \\ [w-Y-c-K(Z)]^{\lambda} & \text{if } \beta Y \leq K. \end{cases} \tag{A6.2}$$

Now using eq. (A6.2) we can write eq. (A6.1) as

$$\begin{aligned} L_2(w,y) = \operatorname*{Max}_{c,Y} \Big\{ c^{\lambda} + \alpha \int_a^{K/Y} [w-Y-c-K(Z)]^{\lambda} \mathrm{d}\Phi(\beta) \\ + \alpha \int_{K/Y}^{A} [w-Y-c-K(Z)+\beta Y-K(\Pi)]^{\lambda} \mathrm{d}\Phi(\beta) \Big\}. \end{aligned} \tag{A6.3}$$

With this formulation the problem may be viewed as one with variable rather than a fixed cost of transaction. For, by allowing the individual the choice of free disposal of assets in the second period, if the market value of assets becomes less than the cost of transaction K, we are implying that he incurs a cost in the second period which equals $\min[K, \beta Y]$. In other words, at the time of decision making in the current period, the individual knows only that any choice of the asset at a positive level would involve a cost of either K or βY in the second period over and above the current transaction cost, if any, and hence, from his point of view, it is not a fixed cost. In a two-period model the analysis and the form of optimal behavior is then dominated by the variable transaction cost in the second period. However, in a multi-period problem the importance of the last periods diminishes rapidly as the horizon increases and the fixed cost character of the problem becomes more important. It is this fact which lends further justification for the procedure in the text, and leads us to believe that the results obtained there will provide a closer representation of multi-period behavior than any outcomes obtained from the model in the appendix.

Returning to the model given in (A6.1) and (A6.2), properties of the optimal portfolio depend on

$$\begin{aligned} h(Y, y, w-c) = \int_a^{K/Y} [w-Y-c-K(Z)]^{\lambda} \mathrm{d}\Phi(\beta) \\ + \int_{K/Y}^{A} [w-Y-c-K(Z) + \beta Y-K(\Pi)]^{\lambda} \mathrm{d}\Phi(\beta). \end{aligned} \tag{A6.4}$$

The major problem here is that the limits of integration depend on the choice variable Y. An important consequence, which makes analysis intractable, is that in general the function $h(Y, y, w-c)$ is not concave in Y. In fact, we obtain after a number of simplifications,

$$h_{YY}(Y, y, w-c) = \lambda(\lambda-1) \int_{a}^{K/Y} D^{\lambda-2} \mathrm{d}\Phi(\beta)$$
$$+ \lambda(\lambda-1) \int_{K/Y}^{A} \{E^{\lambda-2}(\beta-1)^2\} \mathrm{d}\Phi(\beta) + \lambda D^{\lambda-1} \frac{K^2}{Y^3} f(K/Y), \tag{A6.5}$$

where $D=[w-Y-c-K(Z)]$ and $E=[w-Y-c-K(Z)+\beta Y-K(\Pi)]$. Here the first two terms of eq. (A6.5) are negative whereas the last term is positive. Hence it is in general not possible to say anything about the sign of h_{YY}. Consequently, while it is possible to obtain some properties of behavior rather easily, a full characterization of optimal behavior, which we have not attempted, promises to be exceedingly cumbersome.

COMMENTS

*On consumption and portfolio choices with transaction costs**

Hayne E. Leland

C6.1.1. Introduction

Using the class of utility functions exhibiting constant relative risk aversion, Mukherjee and Zabel (MZ) have shown that many results in portfolio/consumption theory depend upon the absence of transactions costs. While only one counterexample disproves a theory, counter-examples alone cannot provide a new theory. In this note we develop some general principles of optimal portfolio selection in the presence of

* This work was supported by National Science Foundation Grant GS-2874-A1 at the Institute for Mathematical Studies in the Social Sciences at Stanford University. A preliminary version was delivered at the NSF–NBER Conference on Decision Rules and Uncertainty, Iowa City, May 1972.

fixed transaction costs. These principles hold for all risk averse utility functions, not just for those exhibiting constant relative risk aversion. Our analysis implies that some, but not all, of MZ's results hold generally.

Given a single risky asset, we show the existence of an interval such that if the original risky asset holding belongs to this interval, no portfolio adjustment will be made. If the original holding lies outside the interval, it will be optimal to choose a risky holding which is independent of the initial holding.

We also examine the sensitivity of the 'no adjustment' interval to parametric changes. For example, the size of the interval is an increasing function of the fixed transaction costs. The interval moves to the right or left with increasing initial wealth as the index of absolute risk aversion decreases or increases with initial wealth. When consumption is also a decision variable, these results can be extended only if a crucial function is strictly concave.

C6.1.2. Portfolio selection with fixed transaction costs

Following MZ's notation, let

w = initial wealth,
c = consumption,
y = initial dollar holdings of the risky asset,
Y = dollar holdings of risky asset,
K = transactions cost if $Y \neq y$ (fixed transaction cost only),
β = random value per dollar invested in the risky asset,
$\phi(\beta)$ = (subjective) distribution function of β.

In this section, we assume c is fixed at zero. For simplicity, we further assume that there are no restrictions on the choice variable Y, that final wealth need not be converted to cash, and that the net return to the riskless asset is zero.

The investor is assumed to

$$\underset{Y}{\text{maximize}}\, EU[w - Km + (\beta - 1)Y],$$

where $m = 1$ if $Y \neq y$, $m = 0$ if $Y = y$. Following normal practice, we assume $U'(\cdot) > 0$, $U''(\cdot) < 0$, and a maximizing Y always exists[1].

[1] Sufficient conditions on U for the existence of an optimal portfolio are given in H. Leland, On the existence of optimal policies under uncertainty, *Journal of Economic Theory*, **4** (February 1972).

For the class of all risk averse utility functions, we show below that the optimal policies have a structure similar to that *MZ* have shown in their theorem 6.6 for utility functions exhibiting constant relative risk aversion.

THEOREM C6.1. *Given w, c, and K, the optimal portfolio strategy is characterized by a triple* (y^0, Y^*, y^{00}), *with* $y^{00} > Y^* > y^0$, *with the optimal strategy given by* $Y = y$ *if* $y \in [y^0, y^{00}]$, $Y = Y^*$ *otherwise.*

PROOF. Define

$$V(y) = EU[w + (\beta - 1)y],$$
$$M = EU[w - K + (\beta - 1)Y^*] = \max_Y EU[w - K + (\beta - 1)Y].$$

Note that $U''(\cdot) < 0$ implies $V(y)$ is strictly concave. Let y^* maximize $V(y)$. Because $U'(\cdot) > 0$, $V(y^*) > M$. It follows that the set $Z = \{y | V(y) \geq M\}$ is convex, compact and nonempty. We may represent Z as an interval $[y^0, y^{00}]$. Now $y \in [y^0, y^{00}]$ implies $V(y) \geq M$, so no portfolio change should be made: $Y = y$. If $y \notin [y^0, y^{00}]$, $V(y) < M$ and it is optimal to set $Y = Y^*$.

COMMENT. Theorem C6.1 shows that the interval structure of the 'no adjustment' strategy is a general property of optimal portfolio policies in the presence of fixed transactions costs, if there is a single risky asset. Multiple risky assets, however, complicate the problem. If a transaction cost is incurred for each asset whose level is changed, there will in general be a set of vectors Y^*; that which is optimal will depend on what region contains the initial portfolio.

C6.1.3. Sensitivity of $[y^o, y^{oo}]$ *to changes in w or K*

MZ show in eqs. (6.42) and (6.43) that the 'larger is wealth net of current consumption, the smaller is the interval in which asset holdings remain unchanged.... Analogous results apply in terms of y^0 and y^{00}.' In fact, while $\partial v^0/dw > 0$ implies $dy^0/\partial w > 0$, $\partial v^{00}/\partial w < 0$ does *not* imply $\partial y^{00}/\partial w < 0$. Below, we show that the behavior of $\partial y^{00}/\partial w$ depends upon the behavior of the index of absolute risk aversion $[-U''(Z)/U'(Z)]$ as wealth Z increases. Because the class of utility functions examined by MZ exhibits decreasing absolute risk aversion, theorem C6.2 below implies $\partial y^{00}/\partial w$ is positive. We therefore conclude that, although the size of the interval $[y^0, y^{00}]$ diminishes as a proportion of wealth for constant relative risk averse utility functions, we have not proved that the absolute magnitude of the interval $[y^0, y^{00}]$ is a decreasing function of wealth.

THEOREM C6.2. *$\partial y^0/\partial w$ and $\partial y^{00}/\partial w$ are positive, constant, or negative as the index of absolute risk aversion, $[-U'(Z)/U''(Z)]$ is a decreasing, constant, or increasing function of wealth Z, given $y^0 \geq 0$.*

Before we can prove theorem C6.2, we require the following:

LEMMA C6.1. *Let U be an increasing strictly concave utility function. Further, let*

$$\int_{\underline{x}}^{\bar{x}} U(a + bx)\mathrm{d}F(x) = \int_{\underline{x}}^{\bar{x}} U(c + dx)\mathrm{d}F(x), \tag{C6.1}$$

where $d > b \geq 0$ and $\underline{x}$, $\bar{x}$ are the inf and sup of the random variable x. Then, for any increasing strictly concave function ϕ,

$$\int_{\underline{x}}^{\bar{x}} \phi[U(a + bx)]\mathrm{d}F(x) > \int_{\underline{x}}^{\bar{x}} \phi[U(c + dx)]\mathrm{d}F(x). \tag{C6.2}$$

PROOF OF LEMMA C6.1. By the strict concavity of ϕ,

$$\begin{aligned} &\int_{\underline{x}}^{\bar{x}} \{\phi[U(a + bx)] - \phi[U(c + dx)]\}\mathrm{d}F(x) \\ &> \int_{\underline{x}}^{\bar{x}} \phi'[U(a + bx)]\,[U(a + bx) - U(c + dx)]\mathrm{d}F(x). \end{aligned} \tag{C6.3}$$

Integrating the right-hand side of eq. (C6.3) by parts and using eq. (C6.1) gives

$$\begin{aligned} &-\int_{\underline{x}}^{\bar{x}} \{\int_{\underline{x}}^{x} [U(a + by) \\ &- U(c + dy)]\mathrm{d}F(y)\}\, \phi''[U(a + bx)]U'[a + bx]b\mathrm{d}x \geq 0. \end{aligned} \tag{C6.4}$$

The inequality follows from the fact that the integral in brackets is always non-negative (shown below), whereas $\phi''(\cdot)U'(\cdot)b$ is always non-positive since $\phi''(\cdot) < 0$, $U'(\cdot) > 0$, and $b \geq 0$. Now eq. (C6.1) and the assumption that $d > b > 0$ imply there exists a y^0 such that for $y < y^0$, $a + by > c + dy$ and $U(a + by) > U(c + dy)$, while for $y > y^0$, the opposite holds. This single crossing property and eq. (C6.1) together imply the bracketed integral is non-negative for all $x \in [\underline{x}, \bar{x}]$. Combining eqs. (C6.3) and (C6.4) gives the desired result eq. (C6.2)[2].

The intuitive reasoning for lemma C6.1 is as follows: If a gamble $a + b\tilde{x}$ is indifferent to a gamble $c + d\tilde{x}$ for a given U, then the more risky gamble $c + d\tilde{x}$ is dispreferred by a person who has a more risk averse utility function $U^* = \phi(U), \phi' > 0, \phi'' < 0$.

PROOF OF THEOREM C6.2 (for decreasing absolute risk aversion). By the definitions of y^0, Y^* and y^{00}, we have

$$EU[w+(\beta-1)y^0]=EU[w-K+(\beta-1)Y^*]=EU[w+(\beta-1)y^{00}]$$

or (C6.5)

$$EU[w^0] = EU[w^*] = EU[w^{00}],$$

where w^0, w^* and w^{00} are defined as the arguments of U in the first line of eq. (C6.5). Recalling our earlier definition of y^*, we note for $K > 0$, $y^0 < (y^*, Y^*) < y^{00}$. Differentiating eq. (C6.5) with respect to w, and noting $EU'(w^*)(\beta - 1) = 0$, gives

$$\frac{\partial y^0}{\partial w} = \frac{\{E[U'(w^*)] - E[U'(w^0)]\}}{E[(\beta - 1)U'(w^0)]}, \tag{C6.6}$$

and

$$\frac{\partial y^{00}}{\partial w} = \frac{\{E[U'(w^*)] - E[U'(w^{00})]\}}{E[(\beta - 1)U'(w^{00})]}. \tag{C6.7}$$

Because y^* is interior to $[y^0, y^{00}]$ and $V(y)$ is strictly concave, $V'(y^0) > 0$ and $V'(y^{00}) < 0$. But $V'(y^0)$ is the denominator of eq. (C6.6), and $V'(y^{00})$ of eq. (C6.7). The signs of eqs. (C6.6) and (C6.7) will therefore depend on $E[U'(w^0)]$ vs. $E[U'(w^*)]$ vs. $E[U'(w^{00})]$.

Given $y^0 < Y^* < y^{00}$ and $y^0 \geq 0$, it follows that $w^0 = a + b\beta$, $w^* = c + d\beta$, and $w^{00} = e + f\beta$, with $f > d > b \geq 0$, and $a = w - y^0$, $b = y^0$, etc. Now the monotonic properties of $U(\cdot)$ and $U'(\cdot)$ imply there exists a transform ϕ such that $-U'(\cdot) = \phi[U(\cdot)]$. Furthermore,

$$\phi'(\cdot) = -U''(\cdot)/U'(\cdot) > 0,$$

[2] The author thanks Peter Diamond and Steve Ross for suggestions which led to this proof.

and

$$\phi''(\cdot) = \mathrm{d}\left[\frac{-U''(\cdot)}{U'(\cdot)}\right]\Big/\mathrm{d}(\cdot) < 0$$

if absolute risk aversion is decreasing.

Therefore $-U'(\cdot)$ is an increasing concave function of U, and using eq. (C6.5) and lemma C6.1 gives

$$EU'(w^0) > EU'(w^*) > EU'(w^{00}). \qquad \text{(C6.8)}$$

Using eq. (C6.8) in eqs. (C6.6) and (C6.7) implies $\partial y^0/\partial w, \partial y^{00}/\partial w > 0$. Obvious modifications of this approach yield the remainder of theorem C6.2.

Comment. We have shown that the behavior of the index of absolute risk aversion critically affects the position of the interval of initial portfolio holdings for which no change is optimal strategy. Not surprisingly, this interval moves to the right or left with increasing wealth – just as does the optimal holding Y^* of the risky asset if change is required – as absolute risk aversion decreases or increases with wealth. Our results do not tell us, however, about the behavior of the *length* of this interval with respect to increasing wealth.

Theorem C6.3.

$$\frac{\partial y^0}{\partial K} < 0; \qquad \frac{\partial y^{00}}{\partial K} > 0.$$

Proof. Differentiating eq. (C6.5) with respect to K and recalling $E(\beta - 1)U'(w^*) = 0$ gives

$$\frac{\partial y^0}{\partial K} = \frac{-EU'(w^*)}{E[(\beta - 1)U'(w^0)]} < 0,$$

and

$$\frac{\partial y^{00}}{\partial K} = \frac{-EU'(w^*)}{E[(\beta - 1)U'(w^{00})]} > 0.$$

Comment. Increasing transactions cost has the expected effect of increasing the amount of inertia, in that a wider range of initial portfolios will require no adjustment of asset holdings.

We now turn our attention to the case where consumption and portfolio decisions are made simultaneously to maximize expected utility in a two-period context. The utility function is assumed additive in consumptions in periods 1 and 2, and all wealth is consumed in the second period.

C6.1.4. Simultaneous portfolio and consumption decisions

Now assume that the investor wishes to choose Y and c to

$$\underset{Y,c}{\text{maximize}}\ U_1(c) + \alpha E U_2[w - c - Km + (\beta - 1)Y] \qquad \text{(C6.9)}$$

where $m = 1$ if $Y \neq y$; $m = 0$ if $Y = y$; $0 < \alpha < 1$; and U_1 and U_2 are increasing and strictly concave. As in the portfolio model without consumption, we can ask if there exist 'trigger values' y^0 and y^{00} which define an interval Z, with the property that if $y \in Z$, no portfolio change is required. It turns out that the introduction of consumption possibilities complicates the analysis. Define

$$\begin{aligned} H(y) &= \max_c \{U_1(c) + \alpha E U_2[w - c + (\beta - 1)y]\} \\ &= U_1[c^*(y)] + \alpha E U_2[w - c^*(y) + (\beta - 1)y]. \end{aligned} \qquad \text{(C6.10)}$$

That is, $c^*(y)$ maximizes expected utility conditional on no portfolio change.

Now define

$$M = \max_{Y,c} \{U_1(c) + \alpha E U_2[w - c - K + (\beta - 1)Y]\}. \qquad \text{(C6.11)}$$

Let c^{**}, Y^* provide the maximum to eq. (C6.11).

THEOREM C6.4. *Assume $H(y)$ is a strictly concave function of y, for any w. Then there exists an interval $[y^0, y^{00}]$, dependent on w, such that if $y \in [y^0, y^{00}]$, $Y = y$ and $c = c^*(y)$. If $y \in Z$, then $Y = Y^*$ and $c = c^{**}$.*

PROOF. Similar to the proof of theorem C6.1.

COMMENT. In contrast with theorem C6.1, $U'' < 0$ does *not* imply $H''(y) < 0$: the assumption that $H(y)$ is concave is nontrivial. Differentiating eq. (C6.10) twice with respect to y and simplifying gives

$$H''(y) = \alpha\{E(\beta - 1)^2 U_2''(\cdot) - [E(\beta - 1)U_2''(\cdot)]^2/D\}, \qquad \text{(C6.12)}$$

where $D = [U_1''(c^*) + \alpha E U_2''(\cdot)] < 0$. The first right-hand side term of eq. (C6.12) is negative, but the second term is positive. In the case of quadratic utility functions, $H''(y)$ is in fact negative. But quadratic functions exhibit certain well known behavioral aberrations, and this could be another one of them.

In sum, the introduction of consumption seriously complicates optimal portfolio policies in the presence of transactions costs, unless the key function eq. (C6.10) is concave in initial portfolio holdings *y*.

On consumer consumption and portfolio decisions with transactions costs

Stephen A. Ross

C6.2.1. Discussion

Professors Mukherjee and Zabel have given us a closely reasoned study of the individual consumption–savings choice in the presence of both uncertainty and transactions costs. The point I find to be of particular interest is the requirement in such problems, with or without transactions costs, that the feasible set of actions be assumed bounded. Alternatively, we can assume that preferences are such that actions are bounded by choice.

To be more specific, suppose we simplify the problem by ignoring the first period consumption withdrawal and focus on the maximization of the expected utility of end of period wealth. Fig. C6.1 illustrates a concave

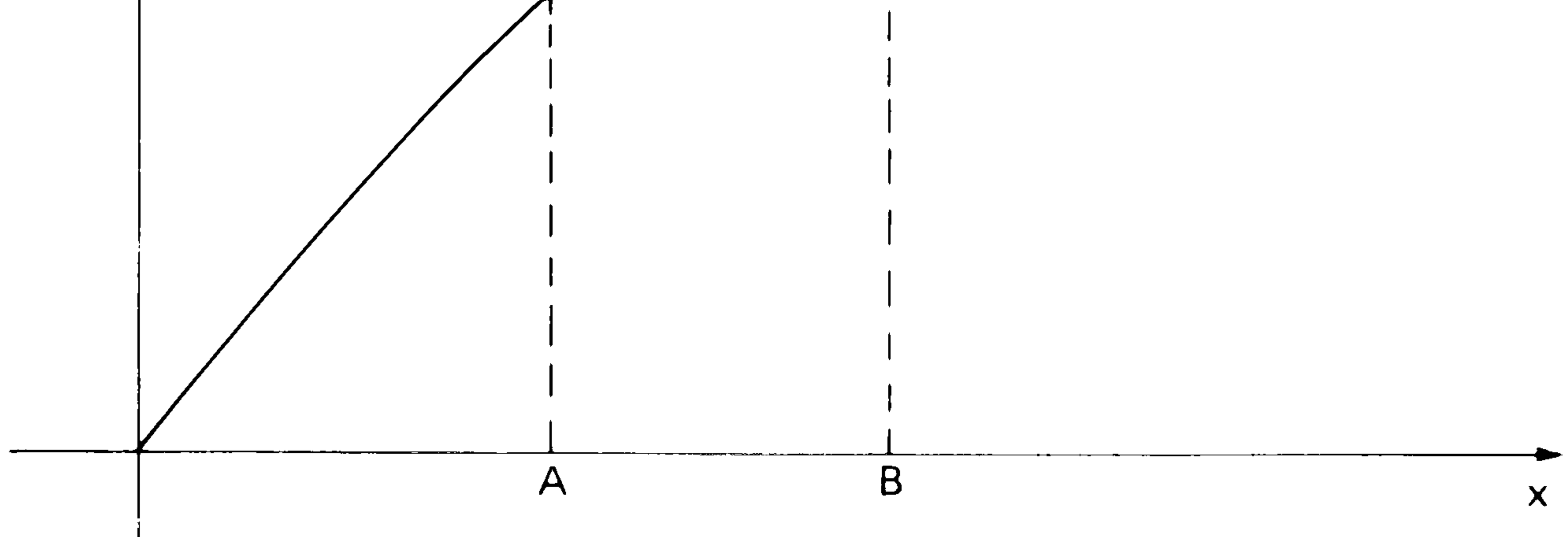

Fig. C6.1.

utility function that has been normalized so that $u(0) = 0$. Suppose too, for the sake of the argument, that among his other alternatives the individual can engage in even money lotteries of his choice. As Raiffa has pointed out [2], even if the utility function had a local non-concave region of the sort illustrated with the dotted curve in fig. C6.1, we could fill it in or concavify it with a straight line segment by purchasing fair gambles with the prizes A and B. This is a special case of the more general result that when the individual is faced with a collection of independent choices (not mutually exclusive) the relevant utility function to be applied on any subset is the envelope of the expected utility attainable by choices on the complement.

Now, if the individual seeks to maximize the expected utility of terminal wealth, then it is immediate from the concavity of $u(\cdot)$ that he will not take any independent even money gambles; a fair gamble has a zero expected return and also entails risk. But is it really so clear?

If the individual could obtain arbitrary fair lotteries there would be nothing to prevent him from choosing gambles which run the risk of bankruptcy. If the disutility of bankruptcy and its attendant social stigma is bounded from below, then as in fig. C6.2 the individual will choose to take fair lotteries with arbitrarily high losses. A sequence of the

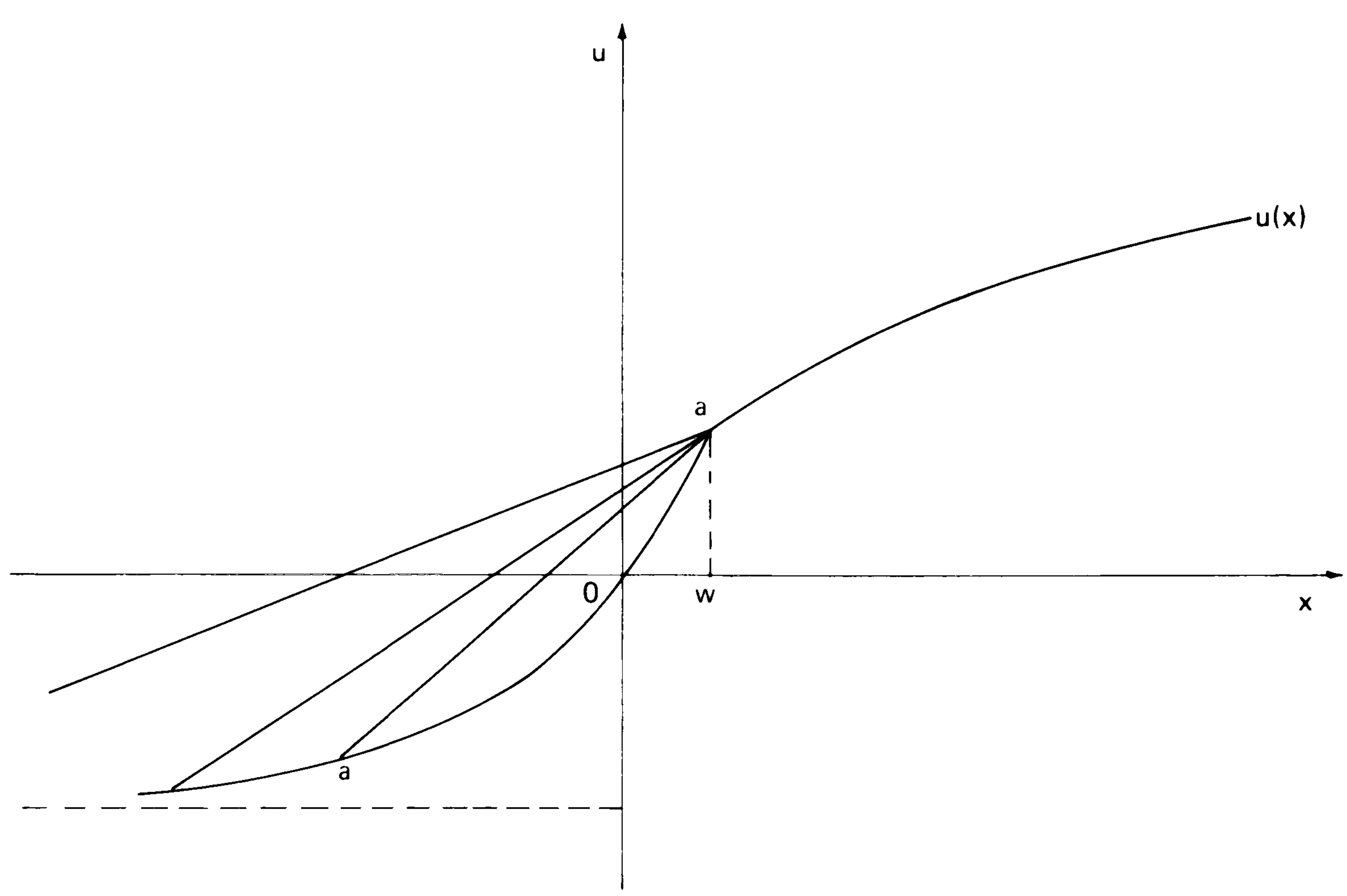

Fig. C6.2.

attainable utilities from a sequence of simple gambles with a fixed positive return of w and a decreasing probability of increasing loss is illustrated in fig. C6.2. The problem involves a fundamental non-concavity and the equilibrium for a 'lottery taker' (like a 'price taker') will be undefined as the agent takes arbitrarily large positions.

The assumption that utility is everywhere bounded from below, then, is anything but innocuous. For one thing, if bankruptcy is allowed such utility functions cannot be everywhere concave, and the difficulty noted above will arise. (The problem would be exacerbated in a welfare state where negative positions were wiped out and the individual was returned to a minimum living standard.) One way out of this is to assume that utility functions become improper at some wealth value A which need not be zero and may be negative. For $x \leq A$, then $u(x) = -\infty$ and for $x > A$, $u(x)$ is well-defined and everywhere concave. By choice such an agent will bound his positions, but a situation like this is too good to be true. It seems inevitable that the market will have to impose some constraints on individual action, perhaps in the form of collateral restrictions.

Consider an agent who starts with no wealth. The line *aa* in fig. C6.2 illustrates a lottery as seen by the agent but *aO* indicates the true lottery since the agent cannot pay for his loss. In such a situation, honesty on the part of the agent would preclude him taking such a position. In the absence of such honesty (and, like apples, one rotten one can ruin a perfect market) agents will have to come to the market announcing their wealth positions. Of course, as in the case of a sick man seeking health insurance, the individual has an incentive to misinform and, in this case, to overstate his true wealth. This could lead, as Akerlof pointed out [1], to a dimunition in the size of the market and to a loss of efficiency relative to the perfect information situation. More likely, though, compensating market institutions will arise. The market may attempt to infer what the individual's wealth is from observable market characteristics. Even this, however, is insufficient; in the first place it is inherently imperfect and in the second there is nothing to prevent the individual from making multiple contracts. Since the recording and dissemination to the market of such contracts would be costly and, perhaps, undesirable from the individual viewpoint, market response takes different forms. The legal penalty of fraud proceedings is one such response; we can argue that by increasing the penalty of negative positions laws against fraud concavify the utility function when viewed as a function of wealth alone.

The arguments presented above, the need to know individual wealth positions, and the need to prevent multiple contracts, provided the economic motivation for such laws. In fact much of the common law has its root in such simple economic arguments. More generally, though, it is in the exploration of the development of market and legal institutions that, I believe, the real value of models such as Mukherjee–Zabel's lies.

References

[1] Akerlof, G. The market for 'lemons': qualitative uncertainty and the market mechanism. *Quarterly Journal of Economics*, **84** (1970), 488–501.
[2] Raiffa, H. *Decision Analysis: Introductory Lectures on Choices Under Uncertainty.* Addison-Wesley: Reading, Massachusetts (1968).

CHAPTER 7

THE ECONOMICS OF QUEUES: A BRIEF SURVEY

David Levhari and Eytan Sheshinski*

7.1. Introduction

There is a well developed literature (for example, refs. [9] and [3]) on the statistical properties of queues with a variety of disciplines. However, there has been little discussion of the economic aspects of queues. Queues provide an obvious example of *externalities*, analogous to situations with *congestion*. It seems natural to try to use the quantitative statistical results of queueing theory for problems of optimization, pricing and queue regulation.

We intend to present here a brief and necessarily selective survey of the literature in order to bring out what we consider to be the main economic applications of queueing theory and to present some tentative results concerning Pareto-optimum pricing systems.

The only discussions in the queueing literature stemming from economic considerations, that got considerable attention, concern the creation of priority classes for waiting line phenomena. However, the assignment of priorities was discussed from an administrative point of view. No attempt was made to decentralize decisions by means of a price mechanism. The decisions about the number of the priorities and about the type of disciplines to be followed (preemptive or non-preemptive priorities), are typically decided by a criterion based on some statistical properties of the queue, but do not use tools of economic theory to find the best allocation of resources in queueing.

Another question that did not get any attention is the structure of prices of competitive firms supplying some service with a given distri-

* Supported in part by National Science Foundation Grant GS-3269. We would like to thank Michael Rothschild for helpful comments.

bution of costs per unit of time. It is intuitively clear that different wage distributions imply different price structures.

7.2. The Naor Model

Naor and others have written a number of papers intended to bring out the nature of the externalities in waiting line situations (in particular, refs. [7] and [10]).

Their model can be described as follows. Customers are arriving at a service station at a given rate. Every customer arriving at the service station observes the waiting line ahead of him and decides, in view of his expected waiting time on the one hand and the benefits derived on the other hand, whether to queue up or to leave the service station. Thus, the model is a queueing model that permits *balking*. The only cost considered is that of the waiting time, ignoring costs of arriving at the service station.

The following assumptions are standard ones in simple queueing situations ($M/M/1$), and are followed in the Naor model:

(i) a stationary Poisson stream of customers, with a parameter λ, arrives at a single service station;

(ii) the station renders service in such a way that the service times are independently, identically and exponentially distributed with a rate parameter μ;

(iii) all customers derive, upon completion of the service, the same benefit of R dollars;

(iv) the cost to a customer for staying in a queue is C dollars per unit of time; and

(v) each newly arrived customer is required to choose one of the following two alternatives: (*a*) he joins the queue and after waiting derives the benefits of the service or (*b*) he refuses to join the queue, an action which is âssured to be costless.

All customers are assumed to be risk neutral, that is, they take only the expected costs and benefits into account and disregard higher moments of their distributions.

A newly arrived customer will observe the queue size i at that instant. If the observed value falls short of some number n_p the new customer will join the queue. If the observed value is equal to n_p then the new customer leaves without joining the queue. The observed value i can never exceed n_p. If we define

$$\lambda/\mu = \rho \tag{7.1}$$

we obtain the following steady state equations

$$p_i\rho = p_{i+1} \qquad 0 \leq i < n_p, \tag{7.2}$$

and hence for $\lambda \neq \mu$ [7]

$$p_i = \frac{\rho^i}{1 + \rho + \ldots + \rho^{n_p}} = \frac{\rho^i(1-\rho)}{1-\rho^{(n_p+1)}} \qquad \lambda \neq \mu. \tag{7.3}$$

The expected value of customers in the system is

$$q = E(i) = \frac{\rho}{1-\rho} - \frac{(n_p+1)\rho^{(n_p+1)}}{1-\rho^{(n_p+1)}} \qquad \lambda \neq \mu. \tag{7.4}$$

The expected number of customers ζ diverted from the station in unit of time is

$$\zeta = \lambda\rho_{n_p} = \frac{\lambda\rho^{n_p}(1-\rho)}{1-\rho^{(n_p+1)}}. \tag{7.5}$$

The expected number of customers joining the queue in unit time equals

$$\lambda - \zeta = \lambda(1 - p_n) = \lambda\frac{1-\rho^{n_p}}{1-\rho^{(n_p+1)}}. \tag{7.6}$$

The expected number of customers leaving the service station in a unit of time equals

$$\mu(1 - p_0) = \mu\left[1 - \frac{1-\rho}{1-\rho^{(n_p+1)}}\right]. \tag{7.7}$$

These two quantities, the joining and the leaving rates, must be identical under steady state conditions, and this is easily verified.

n_p is selected by the customers in the following manner. When the newly arrived customer finds i customers in the system (one of them in service) he expects to incur an expected cost of $(i+1)C/\mu$ and this is weighed against the benefit R. If $R - (i+1)C/\mu \geq 0$ the customer remains in line. Thus n_p is found by

$$R - \frac{C}{\mu}n_p \geq 0, \tag{7.8}$$

and

$$R - (n_p + 1)\frac{C}{\mu} < 0. \tag{7.9}$$

Thus we find

$$n_p \leq \frac{R\mu}{C} < n_p + 1. \tag{7.10}$$

Alternatively we may say that

$$n_p = \left[\frac{R\mu}{C}\right], \tag{7.11}$$

where [ ] denotes the largest integer not exceeding the number in the brackets.

It is quite obvious that when a customer deliberates whether to join or not to join the queue he is just taking his own costs into account neglecting the effect of his joining the line on the expected costs of the following customers.

We now assume that there is a central planner with the aim of maximizing the expected sum of net gains accruing to customer per unit of time. The strategy that the planner follows is that of determining the maximum length of the waiting line, or the maximum number of customers in the system. Thus, the planner determines a natural number n_s, of the people allowed to remain in the system.

With a given n_s, the net gain per unit of time is given by

$$\begin{aligned} P &= (\lambda - \zeta)R - CE(i) = \lambda R(1 - p_{n_s}) - Cq \\ &= \lambda R\frac{1 - \rho^{n_s}}{1 - \rho^{(n_s+1)}} - C\left[\frac{\rho}{1 - \rho} - \frac{(n_s + 1)\rho^{(n_s+1)}}{1 - \rho^{(n_s+1)}}\right]. \end{aligned} \tag{7.12}$$

The planner chooses n_s to maximize eq. (7.12). Again, after some elementary but lengthy calculation it is shown that n_s should satisfy

$$\frac{n_s(1 - \rho) - (1 - \rho^{n_s})}{(1 - \rho)^2} \leq \frac{R\mu}{C} < \frac{(n_s + 1)(1 - \rho) - (1 - \rho^{(n_s+1)})}{(1 - \rho)^2}. \tag{7.13}$$

Again some simple manipulation proves that $n_s \leq n_p$ where the equality sign holds only when $n_p = 1$. The cause for the difference between the privately and the publicly determined n_p and n_s is obviously the neglected externalities in the single customer calculations.

As is quite usual under these conditions, it is possible with proper taxation to achieve the planner's optimum by a decentralized mechanism. In our case the optimal toll θ_s should fulfil the requirement

$$\frac{C}{\mu}(n_p - n_s - 1) = R - \frac{C(n_s + 1)}{\mu} < \theta_s \leq R - \frac{Cn_s}{\mu} = \frac{C}{\mu}(n_p - n_s). \quad (7.14)$$

If a toll is taken in this range the individual action will be socially optimal.

A third mechanism of providing and pricing the service can be discussed. Assume that the service station is controlled by a monopoly with the aim of maximizing revenue per unit of time. Thus the objective function is to maximize

$$M = (\lambda - \zeta)\theta = \lambda\frac{1 - \rho^n}{1 - \rho^{n+1}}\left(R - \frac{Cn}{\mu}\right), \quad (7.15)$$

where the privately determined n is a function of θ, or ζ is a function of θ. If we follow the calculations similar to the previous ones it is not surprising to find that n_m – the maximum prevailing number of customers in the system determined by the individuals in correspondence with θ_M, the value of θ that maximizes eq. (7.15) – is going to be smaller than the socially determined n_s. Thus we find

$$n_m \leq n_s \leq n_p \qquad (0 \leq \theta_s \leq \theta_M). \quad (7.16)$$

The results in qualitative form are independent of the specifics of the random process of the waiting line. Thus if service time were distributed other than exponentially we would still obtain similar results.

The Naor model brings out a familiar result in economic models, that a neglect of externalities creates overcongested situations, while monopolization of the service yields underutilization of the services provided.

It is also quite obvious that if an alternate facility with waiting time independent of customer flow were introduced into the model and balking were allowed, the optimal toll θ_s and the revenue-maximizing toll θ_M would be identical. Permission to balk and a competitive facility eliminate the 'monopoly-power' of the facility.

Following Edelson and Hildenbrand [2], let the toll at the alternate facility be τ and expected waiting time there be γ, constant and independent of customer flow. An irrevocable decision to join one of the queues must be made before observing the state of the system.

If the facility charges a toll θ, the equilibrium arrival rate $\lambda(\theta)$ must be such that customers are indifferent (*ex ante*) between patronizing the two facilities. Equating total costs

$$\theta + \frac{Cq}{\lambda(\theta)} = \tau + C\gamma, \tag{7.17}$$

where the expected queue size q is now also a function of θ.

$$q = \frac{\lambda(\theta)}{\mu - \lambda(\theta)} = \frac{\rho(\theta)}{1 - \rho(\theta)}. \tag{7.18}$$

If $\theta \gtreqless r$ then $q \lesseqgtr \lambda \cdot \gamma$, i.e. facilities charging a higher toll must offer shorter expected waiting plus service times. A higher toll reduces expected waiting time by decreasing the arrival rate, $\lambda'(\theta) < 0$.

Expected social welfare per unit time equals expected gross benefits R less expected service costs per unit time. In equilibrium

$$R = \tau + C\gamma. \tag{7.19}$$

Thus the optimal toll θ_s is such that $\lambda(\theta)$ maximizes

$$\lambda(\theta)(\tau + C\gamma) - C_q \tag{7.20}$$

given eq. (7.17). A revenue-maximizer, on the other hand, will seek to maximize $\lambda(\theta)\theta$. Let this toll be θ_r. By eq. (7.17)

$$\theta_r = \tau + C\gamma - \frac{Cq}{\lambda(\theta_r)}. \tag{7.21}$$

Therefore the entrepreneur's objective function is identical to eq. (7.12), which implies $\theta_s = \theta_r$.

In Naor's model, balking makes expected net benefits per customer greater than θ. If arrivals do not join when $i \geq n_p$, expected queue size must be less than n_p. Those customers arriving when $i < n_p$ obtain inframarginal benefits which are included in P, eq. (7.12), but not in M, eq. (7.15). By making $\theta_M > \theta_s$ the monopoly is able to expropriate part of his customer's 'consumer-surplus'. The setting of an alternate facility and the no-balking rule, eliminate this possibility.

7.2.1. A two-part tariff

Several writers (for example refs. [2] and [8]) have recently suggested a pricing scheme, termed 'the two-part tariff', according to which the firm sells rights to service valid for a given period, with a specific toll if service is rendered. It can be shown [2] that such a scheme also yields $\theta_s \equiv \theta_r$.

Given that a customer demands at most one service during the validation period, his expected gain (for assumptions, see ref. [8]) is

$$\hat{\lambda} \sum_{i=0}^{n-1} \pi_i \left[\tau + C\gamma - \theta - \frac{C(i+1)}{\mu} \right], \tag{7.22}$$

where $\hat{\lambda}$ is the probability that an individual arrives for service during the period, π_i is the probability of a queue of size i, θ is the toll paid if service is rendered, and n is the queue size at which a customer balks.

Expected total revenue is

$$\hat{\lambda} \sum_{i=0}^{n-1} \pi_i \left[\tau + C\gamma - \theta - \frac{(i+1)C}{\mu} \right] N + \hat{\lambda}\theta \sum_{i=0}^{n-1} \pi_i N$$
$$= \lambda \sum_{i=0}^{n-1} \pi_i \left[\tau + C\gamma - \frac{C(i+1)}{\mu} \right], \tag{7.23}$$

where $\lambda = N\hat{\lambda}$, N being the number of potential customers. The server's objective is to select queue size which maximizes eq. (7.23).

It is easily seen that eq. (7.23) is identical to Naor social welfare function, eq. (7.12), and hence $\theta_s = \theta_r$.

Another immediate extension of the Naor model is to consider a population of potential customers with *varying* opportunity costs C. It cannot be expected that a pricing schema with a limited number of parameters, such as the two-part tariff, will in general be Pareto optimal. Furthermore, there is also no reason to believe that the ranking $\theta_r > \theta_s$ will prevail for the model with heterogeneous population. Some tentative results for a model with *two* types of customers confirm this hypothesis [2].

7.3. Optimum Bribing Model

Kleinrock [4] has studied a queueing problem in which each entering customer is allowed to buy his relative priority by means of a bribe.

The size of the optimal bribe for the customer is determined by the economic opportunity costs, such as the wealth or wage of the customer.

The model consists of a single service facility that services customers who arrive randomly according to a Poisson process, with a random service time also distributed exponentially. The customer's bribe x is a random variable with cumulative distribution $B(x)$. When a new arrival to the system offers a bribe x, he is placed in the queue such that customers whose bribes $x' \geq x$ are in front of him, and all those with $x'' < x$ are behind him (customers with identical bribes are served on a first-come first-served basis).

The average waiting time for a customer with bribe x, $W(x)$, is shown to have the form (see ref. [4])

$$W(x) = W_0/[1 - \rho + \rho B(x)]^2, \tag{7.24}$$

where $W_0 = \lambda/2 \int_0^\infty \gamma^2 \mathrm{d}F(\gamma)$, $F(\gamma)$ the cumulative service time distribution.

Customers are distinguished according to an 'impatience' factor α which stands for the opportunity cost of waiting in line. The cost for the α-customer, $C(\alpha)$, is

$$C(\alpha) = x_\alpha + \alpha W(x_\alpha), \tag{7.25}$$

where x_α is the bribe offered by this customer. Kleinrock then solves the following optimization problem: find the function x_α which minimizes total expected cost C:

$$C = \int_0^\infty C(\alpha)\mathrm{d}P(\alpha) \tag{7.26}$$

subject to a given size of total bribes

$$B = \int_0^\infty x_\alpha \mathrm{d}P(\alpha), \tag{7.27}$$

where $P(\alpha)$ is the cumulative distribution of customers by α.

It is then shown that the optimal policy x_α is a strictly increasing function of α, and that various well-known priority disciplines may be viewed as bribing mechanisms.

It is obvious that Kleinrock avoids the main economic issue of finding a *Pareto-optimum* system of prices, or bribes. This is reflected in the weak result concerning the characterization of the bribing policy which is 'optimal' according to his definition. More fundamentally, since bribes are merely transfer payments between customers and owners, it would perhaps be more appropriate for social decision to minimize total waiting time

$$\int_0^\infty W(x_\alpha)\mathrm{d}P(\alpha) \tag{7.28}$$

subject to a given total cost, eq. (7.26). However, Kleinrock's model may be fruitfully used to explore the latter problem.

7.4. Optimization in Queues without Priorities

We wish to analyze here the problem of choosing optimal values for certain decision variables of the service mechanism (the mean service rate, mean arrival rate, etc.). Previous efforts in optimization of queueing systems have begged this basic question and instead solved a simpler problem, confining the feasible policies to a certain simple form. Thus, as we have seen, Naor [7] solved for the optimal value of the queue size at which to turn on a single server, assuming that the form of the policy is to turn on the server when the queue size reaches a certain figure. Here we are concerned with discovering the *form* of the optimal policy. Our approach follows the work of Marchand [6].

Consider a one service facility (one-server) problem, with a population of n types of customers. Arrival time of each customer is a random variable with a Poisson distribution, where the mean arrival rate of the ith customer is $1/\lambda_i$, $\lambda_i > 0$ a constant. The queue discipline is of the first-come first-serve type, with no priorities. Service time is also a random variable with Poisson distribution, and the mean service time of the ith customer is $1/\mu_i'$, $\mu_i' > 0$ a constant. Given these parameters, the formula for the expected waiting time in the queue w_q is well-known:

$$W_q = \left(\sum_{i=1}^{n} \frac{\lambda_i}{\mu_i'^2}\right) \bigg/ \left(1 - \sum_{i=1}^{n} \frac{\lambda_i}{\mu_i'}\right). \tag{7.29}$$

The expected delay time, i.e. the waiting time plus the service time, denoted D_i, is

$$D_i = W_q + (1/\mu_i'). \tag{7.30}$$

Delay time per arrival is thus $\lambda_i D_i$. The ith customer's utility function is assumed to depend on the benefit derived from the service, on the delay time, and on an aggregate commodity, whose quantity for the ith individual is denoted by x_i. For simplicity, utility is assumed to be linear in the delay time. The expected utility of the ith individual, U_i, can thus be written

$$U_i = u_i(x_i, \lambda_i) - \beta_i \lambda_i D_i \qquad (i = 1, 2, \ldots, n), \tag{7.31}$$

where u_i is a quasi-concave function, with $\partial u_i/\partial \lambda_i > 0$, and $\beta_i > 0$ is a constant. Making expected utility depend only on the expected arrival rate λ_i is of course, a rather strong simplifying assumption.

The optimization problem is to maximize eq. (7.31) subject to an aggregate constraint on resources. There are a number of possible decision variables. Here we concentrate on the aggregate commodities x_i, the arrival parameters λ_i and the service rates μ_i'. For the latter it is assumed that they depend upon a common parameter s, the speed at which the service facility operates:

$$\mu_i' = s \cdot \mu_i, \tag{7.32}$$

where μ_i are given constants. An increase in s means a decrease in service time. Such a change reflects an improvement in the service facility's capacity or equipment, which requires inputs of the aggregate commodity. The resource constraint can thus be written

$$f(x, s) = 0, \tag{7.33}$$

where $x = \sum_{i=1}^{n} x_i$.

Maximization of the sum utilities, eq. (7.31), with respect to x_i, λ_i and s subject to the constraints of eqs. (7.29), (7.30), (7.32) and (7.33) yields the following first-order conditions:

$$\frac{\partial u_j}{\partial x_j} - v\frac{\partial f}{\partial x} = 0 \qquad (j = 1, 2, \ldots, n), \tag{7.34}$$

$$\frac{\partial u_j}{\partial \lambda_j} - \beta_j D_j - \frac{\partial W_q}{\partial \lambda_j} \sum_{i=1}^{n} \beta_i \lambda_i = 0 \qquad (j = 1, 2, \ldots, n), \tag{7.35}$$

$$-\sum_{i=1}^{n} \beta_i \lambda_i \left(\frac{\partial W_q}{\partial s} - \frac{1}{s^2 \mu_i} \right) - v\frac{\partial f}{\partial s} = 0 \qquad (j = 1, 2, \ldots, n), \tag{7.36}$$

where $v \geq 0$ is a Lagrange multiplier. By eq. (7.29),

$$\frac{\partial W_q}{\partial \lambda_j} = \frac{1}{1 - \sum_{i=1}^{n} \frac{\lambda_i}{s\mu_i}} \left(W_q \frac{1}{s\mu_j} + \frac{1}{(s\mu_j)^2} \right) \qquad (j = 1, 2, \ldots, n) \qquad (7.37)$$

and

$$\frac{\partial W_q}{\partial s} = \frac{W_q}{s} \left[-2 - \frac{\sum_{i=1}^{n} \frac{\lambda_i}{\mu_i}}{1 - \sum_{i=1}^{n} \frac{\lambda_i}{s\mu_i}} \right]. \qquad (7.38)$$

Substituting eq. (7.37) into eq. (7.35) and combining with eq. (7.34) yields

$$\frac{\partial u_j}{\partial \lambda_j} \Big/ \frac{\partial u_j}{\partial x_j} = \frac{\sum_{i=1}^{n} \beta_i \lambda_i + \frac{\beta_j}{s\mu_j}}{\left[1 - \sum_{i=1}^{n} \frac{\lambda_i}{s\mu_i} \right] v \frac{\partial f}{\partial x_j}} \left[W_q \frac{1}{s\mu_j} + \frac{1}{(s\mu_j)^2} \right] = A \frac{1}{s\mu_j} + B \frac{1}{(s\mu_j)^2}, \qquad (7.39)$$

where

$$A = W_q \frac{\Sigma \beta_i \lambda_i}{1 - \sum_{i=1}^{n} \frac{\lambda_i}{s\mu_i}} \Big/ v \frac{\partial f}{\partial s} \qquad \text{and} \qquad B = \frac{A}{W_q}$$

The right-hand side of eq. (7.39) is the marginal rate of substitution between x_j and λ_j (for a given queue size), which can be given an interpretation of a price. Hence, the *optimal price formula for λ is quadratic* (see fig. 7.1). Since $(s\mu_j)^2$ is inversely proportional to the expected

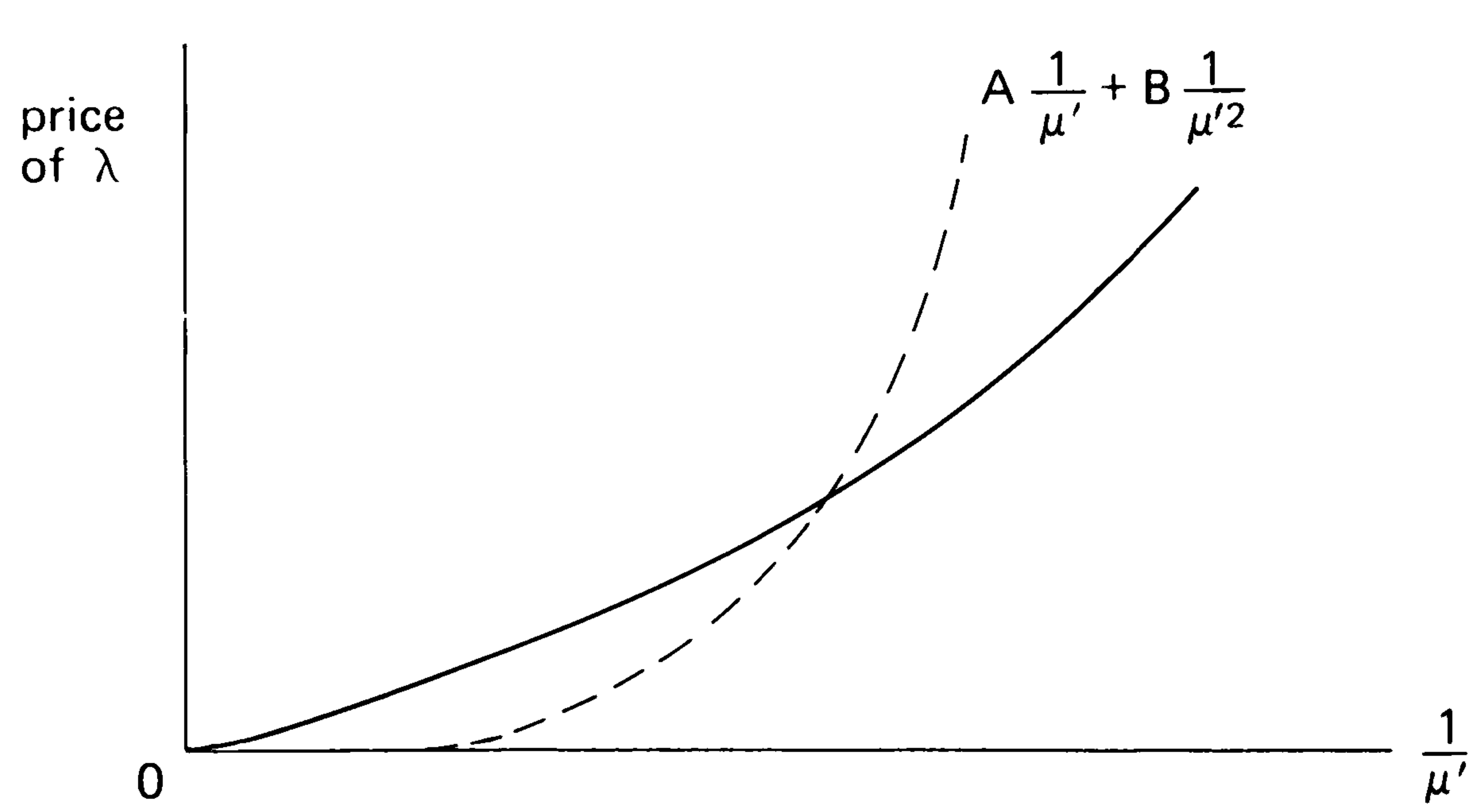

Fig. 7.1.

service variance, this condition implies a price formula that depends linearly on the mean and variance of the service rate. A similar formula has been suggested by Drèze [1] in the context of peak-load pricing under uncertainty.

The quadratic formula implies that customers with longer service rates are penalized more than proportionately. The rate of change in the price depends on the ratio $A/B = W_q$. The larger the waiting line W_q, the steeper will the price function rise.

The marginal rate of transformation between x and s is obtained from eqs. (7.38), (7.36) and (7.34):

$$\frac{\partial f}{\partial s}\bigg/\frac{\partial f}{\partial x} = -\sum_{i=1}^{n} \beta_i \lambda_i \left(\frac{\partial W_q}{\partial s} - \frac{1}{s^2 \mu_i}\right)\bigg/\frac{\partial u_i}{\partial x_i}. \tag{7.40}$$

The queue discipline of first-come first-served is of course not optimal. Since individuals have different costs for delay time, β_i, priority should be given to those with high β and low μ.

The next step should thus be to divide customers into priority groups and to determine the price structure for each group. We shall postpone discussion of this problem to another occasion.

7.5. Queues with Priorities

There is vast literature describing queues disciplines such that some types of customers receive priority (for example, ref. [3]). It is surprising how little attention the regulation of priorities through a price mechanism got.

The cost per unit of time of keeping certain customers queueing may be particularly high and it may then be reasonable to give them a high priority. If the cost per unit queueing time is constant, it will be desirable to reduce the overall mean queueing time and this, as expected, can be achieved by giving high priority to customers expected to have a small service time.

It is simplest to deal with a priority system in which once a customer is at the service point he remains there until his service is completed. Then the next customer is the one of highest priority among those queueing. This is called non-preemptive priority. A preemptive priority system is one in which a customer of high priority takes, on arrival, immediate precedence over customers of lower priority, the customer whose service is interrupted returning to the service point only when there are no higher priority customers remaining in the system. In the non-preemptive case each customer has a priority class $1, \ldots, k$ where 1 is the highest priority and k the lowest. Customers of a given class are served in order of arrival.

Assume that customers of different classes arrive independently at random rates $\lambda_1, \ldots, \lambda_k$ and that the unit of time is chosen so that the total arrival rates $\lambda_1 + \ldots + \lambda_k = 1$. Let the service time of different customers be independently distributed with $F_j(t)$, the distribution function for the j customers. The 'overall' service time distribution is

$$F(t) = \sum_{j=1}^{k} \lambda_j F_j(t).$$

Let us denote

$$b^j = \int_0^\infty t \mathrm{d}F_j(t) \qquad \text{and} \qquad c_j = \int_0^\infty t^2 \mathrm{d}F_j(t).$$

The moments for $F(t)$ will be respectively

$$b = \Sigma \lambda_j b_j \qquad \text{and} \qquad c = \Sigma \lambda_j c_j.$$

We assume that $b < 1$, to assure the existence of a steady state.

The case of exponential service time is particularly simple. If μ_j is the rate of service of the j customers with $1/\mu_j$ the mean service time, then

$$b = \Sigma \frac{\lambda_j}{\mu_j} \qquad \text{and} \qquad c = 2\Sigma \frac{\lambda_j}{\mu_j^2}.$$

The mean queueing time of j-customer W_j is, on allowing for the chance $(1 - b)$ of not having to queue,

$$W_j + \frac{\frac{1}{2}c}{\left(1 - \sum_{i=1}^{j-1} \lambda_i b_i\right)\left(1 - \sum_{i=1}^{j} \lambda_i b_i\right)}. \tag{7.41}$$

The mean queueing time W of all customers is

$$W = \sum_{j=1}^{k} \lambda_j W_j = \frac{c}{2} \sum_{j=1}^{k} \lambda_j \frac{1}{\left(1 - \sum_{i=1}^{j-1} \lambda_i b_i\right)\left(1 - \sum_{i=1}^{j} \lambda_i b_i\right)}. \tag{7.42}$$

These equations enable one to discuss the effects on mean queueing time of any proposed system of priorities. Suppose that there are k types of customers and that the cost of keeping a customer of the jth type queueing for unit time is constant and equal to w_j – this can represent the wage rate of the jth customer. Assuming risk neutrality, the cost of queueing depends only on the mean queueing time and on the mean cost:

$$C = \sum_{j=1}^{k} \lambda_j w_j W_j = \frac{c}{2} \sum_{j=1}^{k} \frac{\lambda_j w_j}{\left(1 - \sum_{i=1}^{j-1} \lambda_i b_i\right)\left(1 - \sum_{i=1}^{j} \lambda_i b_i\right)}. \tag{7.43}$$

We are interested in choosing priorities $(1, 2, \ldots, k)$ so as to minimize C. Thus if we permute 2 and 3 (assuming $k = 3$) we find a new mean cost C'. After simple calculation we find

$$C - C' = \frac{c\Delta}{2}\left(\frac{w_2}{b_2} - \frac{w_3}{b_3}\right), \tag{7.44}$$

where

$$\Delta = (1 - \lambda_1 b_1 - \lambda_2 b_2)^{-1} + (1 - \lambda_1 b_1 - \lambda_3 b_3)^{-1} - (1 - \lambda_1 b_1)^{-1} - (1 - \lambda_1 b_1 - \lambda_2 b_2 - \lambda_3 b_3)^{-1}.$$

Now if x, y, z are positive and different it may be verified that

$$\frac{1}{x - y} + \frac{1}{x - z} - \frac{1}{x} - \frac{1}{x - y - z} < 0. \tag{7.45}$$

Hence $\Delta < 0$ so that $C < C'$ if and only if $b_2/w_2 < b_3/w_3$; thus if $b_2/w_2 > b_3/w_3$ we can reduce the mean cost by changing priority classifications. The same holds for any $j - 1, j$. If $b_{j-1}/w_{j-1} > b_j/w_j$ we can reduce the mean cost by changing the priority classification for the $j - 1$ customer and the j customer. Thus for the priority system to be cost minimizing we need to order the priority according to

$$\frac{\text{mean service time}}{\text{cost of queueing per unit of time}}$$

– the lower the value, the higher the priority. This should be expected intuitively. It is not surprising that the priorities are independent of the arrival rate.

Now if we would like to implement this optimal (minimizing) cost arrangement through a price mechanism, we should charge prices P_j for each priority class, and we choose $P_{j-1} - P_j$ in the following range:

$$w_j(W_j - W_{j-1}) = P_{j-1} - P_j = w_{j-1}(W_j' - W_{j-1}'), \qquad (7.46)$$

where W_j', W_{j-1}' are the expected waiting times of priorities $j-1$ and j when we permute priority classes $j-1$ and j. The fact that we have minimized C implies

$$w_j(W_j - W_{j-1}) \leq w_{j-1}(W_j' - W_{j-1}'). \qquad (7.47)$$

Otherwise we could reduce the costs by permutation of the j and the $j-1$ priority classes. It is quite obvious that by this method we can get a consistent pricing procedure, and if we let each customer choose his priority class we shall get the same pattern that would be achieved by the planner.

The price of the lowest priority class can be set arbitrarily. Then we set the rest of the prices to agree with the above rate. A special case is when the cost w_j is the same for all groups, when the optimum priority number depends only on the mean service time of a class. A limiting case of this arises when it is possible to predict the service time on arrival and the service time has a distribution $F(t)$. We can then have a continuum of priority classes, the customer selected for service being the one with the lowest service time. That is, we assume that new repair jobs are generated by a Poisson law with an average of λ jobs per unit of time. Without loss of generality we may choose the time unit so that $\lambda = 1$. If we denote by λ_t the arrival rate of jobs of duration t, then in steady state $\lambda_t \mathrm{d}t = \mathrm{d}F(t)$. The expected waiting time W_t of a job of duration t by analogy with the discrete priority case is

$$W_t = \frac{c}{2} \frac{1}{\left[1 - \int_0^t s\mathrm{d}F(s)\right]^2}, \qquad (7.48)$$

where as before

$$c = \int_0^\infty t^2 \mathrm{d}F(t).$$

(We neglect the possibility of saturation – of an infinite length queue.)

The resulting mean queueing time is then easily calculated to be

$$W_{\min} = \frac{c}{2}\int_0^\infty \frac{dF(t)}{\left[1 - \int_0^t s\,dF(s)\right]^2}. \tag{7.49}$$

If, for example, $F(t) = 1 - e^{-\mu t}$ corresponding to exponential repair time (assuming $\mu > 1$), then $c/2 = 1/\mu$ and the expected waiting time for service of duration t is

$$W_t = \frac{\frac{1}{\mu^2}}{\left\{1 - \frac{1}{\mu}\left[1 - e^{-\mu t}\left(1 + \mu t\right)\right]\right\}^2}\,dt. \tag{7.50}$$

The mean queueing time is then

$$W_{\min} = \int_0^\infty \frac{1}{\mu}\cdot\frac{e^{-\mu t}}{\left\{1 - \frac{1}{\mu}[1 - e^{-\mu t}(1 + \mu t)]\right\}^2}\,dt. \tag{7.51}$$

If the queue discipline 'first-come first-serviced' is used, then the mean queueing time is $W = 1/\mu(\mu - 1)$. $W_{\min}$ can be appreciably less than this.

The strategy based on service time is optimal if the service time of each customer can be predicted accurately on arrival and it is known that $\lambda = 1$. A simpler priority system in the continuous case can be as follows. Suppose again $f(t) = \mu e^{-\lambda t}$ so that the mean service time is $1/\mu$. Suppose that we count all customers whose service times are less than or equal to α/μ as 1-priority customers, all other customers being 2-customers. The mean waiting time in this case is

$$W = \frac{c\left(1 - \frac{1}{\mu} + \frac{1}{\mu}e^{-\phi}\right)}{2\left(1 - \frac{1}{\mu}\right)\left(1 - \frac{1}{\mu} + \frac{1}{\mu}e^{-\phi} + \frac{\phi}{\mu}e^{-\phi}\right)}. \tag{7.52}$$

By simple calculus we find that the optimal ϕ has to fulfill $\mu = 1 + \{e^{-\phi}/(\phi - 1)\}$.

This can be supported by a decentralized simple price function $p(t)$, for priority 1 service of duration t. Denote by W_1, W_2 the expected waiting time in priority classes 1 and 2. Let w denote the common wage or alternative time cost. We set $p(\phi) = (W_2 - W_1)w$. Now any increasing monotonic price schedule $p(t)$ – the price for 1 priority of a customer asking a service of duration t – fulfilling this requirement is a price schedule that will make people choose their priority class according to the planner's wish.

As an example, if $1/\mu = \rho = 0.75$ there will be a 37% reduction in mean queueing time by introduction of this priority system rather than 'first-come, first-serviced'. Further gains will be possible if we set more than two priority classes. The more variable is the distribution of service time, the greater will be the advantages of a priority system based on service time.

We mention just briefly the case of preemptive priority. A customer of low priority is displaced by a customer with higher priority immediately on arrival. Assume two types of customers with arrival rates λ_1, λ_2 and with rate of exponential service μ_1, μ_2. Again the time scale is chosen so that $\lambda_1 + \lambda_2 = 1$. The mean queueing time of priorities 1, 2 are

$$W_1 = \frac{\lambda_1}{\mu_1}\Big/\mu_1\left(1 - \frac{\lambda_1}{\mu_1}\right), \tag{7.53}$$

$$W_2 = \left(\frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2}\right)\Big/\mu_2\left(1 - \frac{\lambda_1}{\mu_1} - \frac{\lambda_2}{\mu_2}\right) + \mu_2\frac{\lambda_1}{\mu_1}\Big/\mu_1\left(1 - \frac{\lambda_1}{\mu_1}\right). \tag{7.54}$$

The corresponding waiting time for the non-preemptive case is

$$W_1 = \left(\frac{\lambda_1}{\mu_1^2} + \frac{\lambda_2}{\mu_2^2}\right)\Big/\left(1 - \frac{\lambda_1}{\mu_1}\right), \tag{7.55}$$

$$W_2 = \left(\frac{\lambda_1}{\mu_1^2} + \frac{\lambda_2}{\mu_2^2}\right)\Big/\left(1 - \frac{\lambda_1}{\mu_1}\right)\left(1 - \frac{\lambda_1}{\mu_1} - \frac{\lambda_2}{\mu_2}\right). \tag{7.56}$$

The mean queueing time for both cases is $\lambda_1 W_1 + \lambda_2 W_2$. As expected, 1-priority customers queue a shorter time under a preemptive discipline. It is quite clear that which of the two systems is more advantageous depends on the per-time costs of the two cases.

The question of optimal priority in waiting line situations may assume a different form in the case where the service is provided by numerous competitive firms. The extreme case is one in which the service is provided by a perfectly competitive industry. In that case we may expect that different firms will specialize in different priority classes. This will occur through differential prices charged for the same kind of service. Thus individuals with high alternative time cost will use the services with short waiting lines but higher prices for the service. It seems that it is worth while to investigate the implications of a model in which a given wage distribution generates a certain price structure. Hence, suppliers of the service will charge different prices, each specializing in a given group of wage earners.

References

[1] Drèze, J. H. Postwar contributions of French economists. *American Economic Review* Supplement, **54** (1964), 1–64.
[2] Edelson, N. and Hildebrand, W. *Congestion tolls for Poisson queueing processes.* Discussion Paper No. 252, Department of Economics, University of Pennsylvania, Philadelphia (1973).
[3] Jaiswal, N. K. *Priority Queues.* Academic Press: New York (1968).
[4] Kleinrock, L. Optimal bribing for queue position. *Operations Research,* **15** (1967), 304–318.
[5] Lusky, I. *Queues and social welfare.* Unpublished PhD Dissertation, The Hebrew University, Jerusalem (1971).
[6] Marchand, M. Priority pricing with application to time-shared computers. *AFIPS – Conference Proceedings*, Vol. 33 (1968), 511–519.
[7] Naor, P. On the regulation of queues size by levying tolls. *Econometrica,* **37** (1969), 15–24.
[8] Oi, W. Y. A Disneyland dilemma: two-part tariffs for a Mickey-Mouse monopoly. *Quarterly Journal of Economics,* **85** (1971), 77–96.
[9] Saaty, T. L. *Mathematical Methods of Operations Research.* McGraw-Hill: New York (1959).
[10] Yechiell, V. On optimal balking rules and toll charges in the GI/M/1 queueing process. *Operations Research* (forthcoming).

COMMENTS

On 'The economics of queues: a brief survey'

Warren J. Boe

C7.1.1. Discussion

In any decision problem under uncertainty, a first step to solving the problem is to associate with each pair consisting of an alternative and a state of nature a cost. Levhari and Sheshinski have proposed looking at queueing theory in a way which is essentially that of a decision problem under uncertainty. Their states of nature are the arrival rates of customers, and the alternatives are the various priority classes which may be specified. The problem then is to specify the cost to be associated with a given set of arrival rates and a set of priority classes. The authors' objective in this paper is to decentralize the decisions of creating priority classes and assigning individuals to them by means of a price mechanism.

As a basis for their work, the authors suggest the models of Naor [2] and Kleinrock [1]. This model is a standard Poisson input exponential service time model in which the customer derives R dollars of benefit from the service while his cost of waiting for the service is C dollars per unit time. A customer arriving at the queue may either decide to join the queue and wait for service, or he may decide to leave without service. If his expected cost of waiting exceeds the benefit he derives from the service, he will leave; otherwise he will stay. He determines his expected costs on the basis of the number in the queue already, say n_p.

A central planner who wants to maximize total benefits then also obtains a maximum queue length n_s. Finally, a monopolist who wants to maximize revenue per unit of time determines a maximum queue length n_m. The relationship between these queue lengths is $n_m \leq n_s \leq n_p$.

The authors' discussion of Kleinrock's model treats a special case of that model. Their special case corresponds to the assumptions which describe Naor's model. Kleinrock's model, which is described somewhat carelessly by the authors, allows customers to determine their own priority in the system by offering bribes to the queue administrators.

Levhari and Sheshinski then proceed to develop a queueing model in which they apply Naor's techniques to Kleinrock's model. That is, they determine prices to be associated with different priorities in the queue. While Kleinrock lets individuals decide how large a price they are willing to pay to obtain a certain priority class, Levhari and Sheshinski determine the price to be charged for each priority class by minimizing the total cost. These prices are determined by a central planner after the manner of Naor. It would be possible to determine other priority prices as well, such as the price for a monopolist's operation or a purely competitive operation.

Of course, it is necessary to consider the manner in which priority classes are determined. It is obvious that first-come first-served is not an optimal way to determine priority classes since it does not consider the individual costs of queueing time. The authors state, without proof but with an intuitive appeal, that ordering the priority classes on the basis of the measure

$$\frac{\text{mean service time}}{\text{cost of queueing per unit of time}}$$

minimizes the total cost for the system.

A few 'obvious' results were obfuscated by the authors. The specification of the price mechanism for controlling queue priorities was not clearly stated. The authors only presented intuitive speculations about the nature of price determination by several firms offering the same services. They have, however, provided a basis for this type of price determination by the non-priority and priority models developed in this paper.

References

[1] Kleinrock, L. Optimal bribing for queue position. *Operations Research*, **15** (1967), 304–318.

[2] Naor, P. On the regulation of queues size by levying tolls. *Econometrica*, **37** (1969), 15–24.

CHAPTER 8

ON THE ECONOMIC THEORY OF AGENCY AND THE PRINCIPLE OF SIMILARITY*

Stephen A. Ross

8.1. Introduction

The term 'agency' has its origins in Roman law and has come to be the generic title for a variety of instrumentalities. In the law, a relationship of agency exists between two (or more) parties when one of these, designated the agent, acts on behalf of or as representative for the other, the principal. In a decision-theoretic context it is also necessary to precisely specify the environment of agency, i.e. the domain or particular class of situations within which the relationship exists, the feasible set of actions the agent may take and the consequences of these actions. It is particularly interesting to examine the agency relationship when the agent must make choices and then, in some fashion, share their consequences with the principal in an environment involving uncertainty.

It is easy to think of a number of relationships in economics involving agency. The employee–employer relationship is one, and in general any situation where labor services are hired gives rise to a meaningful agency relationship to the extent to which the employee possesses some decision-making authority[1]. Another example can be found in the class of problems raised by moral hazard. It is illuminating to consider these problems

* This research was supported by grants from the National Science Foundation and Rodney L. White Center for Financial Research. The author wishes to thank the participants of the NSF–NBER Conference for their helpful comments.

[1] In fact, until machines are made which possess self-will, it might be argued that the degree of agency is the primary distinction between labor and capital services (surely not the problems of time, as Marx's embodiment and Arrow's learning by doing and Becker's concept of human capital have taught us).

as belonging to the theory of agency, since the agency relationship, between the insurer and the insured or between the government and the governed (although here the roles of principal and agent are less easy to assign), is the common thread running between them. As a final example, the problem of the fiduciary or the financial intermediary is fundamentally a problem in the theory of agency. The investment counselor, to the extent to which he makes decisions for his client, is his client's agent.

Despite the plethora of examples, though, the relationship of agency has received very little study from economists or decision theorists[2]. The intent of this paper, then, is to formalize the relationship of agency as a problem in decision theory and to propose and examine the implications of one particular type of solution.

8.2. Basic Problem

We will begin by assuming that there is a single agent and a single principal. The agent and the principal will agree to some sort of arrangement in which the agent makes a decision or takes an action and he and the principal share in the consequences. The complete structure of the problem facing the agent and the principal can be described by a payoff function $w = w(\alpha, \theta)$ describing the (one-dimensional) consequence or payoff, most generally a wealth variable, resulting when act α is chosen and state of the world θ occurs or, in game theoretic parlance, is 'chosen by nature'. The action is to be chosen before the random variable θ is known to the parties and the problem is one of choice under uncertainty. To make such decisions, we will let $U(\cdot)$ and $G(\cdot)$ denote the von Neumann–Morgenstern utility functions of the principal and the agent respectively and assume that each evaluates his own position by its expected utility. Both $U(\cdot)$ and $G(\cdot)$ are assumed to be independent of the state of nature θ, ruling out situations, for example, where the principal would evaluate wealth differently if he were sick

[2] The closest work is by Karl Borch [3] and Robert Wilson [8] on the theory of syndicates. After I had written this article, it was pointed out to me that Marvin Berhold [2] had formulated a similar problem and had considered the linear rule L, in some detail. Berhold was not, however, concerned with a theoretical justification of L or its relationship to other rules.

(θ_1) than if he were well (θ_2), and both $U(\cdot)$ and $G(\cdot)$ are assumed to be monotone increasing and, unless otherwise specified, risk averse (i.e. concave).

Having chosen the decision structure, the agency arrangement can be described by the fee schedule f defining the payoff or fee that is paid to the agent for his services. In general, $f = f(\alpha, \theta)$, but we will assume that the only impact of the action taken upon the fee is through the resulting payoff and write $f = f(w(\alpha, \theta), \theta)$. (This would be the case if, for example, it were not possible to monitor or observe the action itself.) The additional dependence upon the state of the world θ arises from the possibility that the agent and the principal may differ in the subjective probability distributions they hold for θ.

Given a fee schedule the agent will now seek

$$\max_{\alpha \in A} \underset{\theta}{E} \{G[f(w(\alpha, \theta), \theta)]\}, \tag{8.1}$$

where A is the set of feasible actions and the expectation is taken over the agent's subjective distribution for θ. (In what follows, the subjective distribution will be understood to be that associated with the relevant party.) Clearly the optimal action chosen by the agent is a function, among other things, of the fee schedule that has been specified, and the agent obviously has preferences among fee schedules.

Suppose first that the agent was asked to pick the optimal fee schedule from his point of view subject to the constraint

$$\underset{\theta}{E} \{f(w(\alpha, \theta), \theta)\} \leq c.$$

The solution to this problem is simply to choose a flat schedule $f(w(\alpha, \theta), \theta) = c$. To see this, observe that (since more is better) the constraint will be binding and since $G(\cdot)$ is risk averse any other fee but the constant one will have the same expected return and, by Jensen's inequality, a lower expected utility. Of course, from the principal's point of view a constant fee might be quite a poor choice since it leaves the agent indifferent as to which act to choose. Notice too that the solution will not in general be *ex ante* Pareto efficient.

The Paretian problem would be to maximize the agent's expected utility subject to the constraint that

$$\underset{\theta}{E} \{U[w(\alpha, \theta) - f(w(\alpha, \theta), \theta)]\} \geq c,$$

i.e. the principal, who receives the payoff less the fee, $w - f$, must attain a minimal level of expected utility. If we assume that the agent and the principal both share the same subjective probability distribution, then there is no reason for the fee to be directly functional on the state, and we can write $f = f(w)$.

By augmenting the act space A, by randomized strategies if necessary, it is not difficult to show that the Paretian frontier of optimal pairs

$$(\underset{\theta}{E}\{(U[w-f]\}, \underset{\theta}{E}\{G[f]\})$$

must be convex (see ref. [6] for a good exposition). Hence we can also pick out efficient fee schedules by seeking

$$\max_{f} \max_{\alpha \in A} \underset{\theta}{E}\{U[w-f] + \lambda G[f]\} = \max_{\alpha \in A} \max_{f} \underset{\theta}{E}\{U[w-f] + \lambda G[f]\},$$

where $\lambda \geq 0$. But the inner maxima simply requires that

$$\text{(PE)} \qquad U'[w-f] = \lambda G'[f],$$

and PE may be used[3] to define $f(w)$. Notice that the PE condition defines a function $f(w)$ and not simply a point value. By changing λ, the complete family of Pareto efficient (PE) fee schedules can now be traced out. Furthermore, $f(w)$ is *not* dependent upon the probabilistic structure of the problem, i.e. given λ, then $f(w)$ is independent of the functional form of $w(\alpha, \theta)$ or of the state of the world θ. This result depends critically on the assumption that the principal and the agent share the same subjective probability distribution over states.

The problem of choosing the optimal fee schedule is considerably more complex when viewed by the principal rather than the agent. Since the act will be chosen by the agent, the principal must choose a fee schedule that appropriately motivates the agent to act in a fashion that is best for the principal. Letting $\alpha(\langle f \rangle)$ denote the action chosen by

[3] The reader can verify the PE condition more formally by considering a variational perturbation, δf, of the function satisfying PE. Now

$$\begin{aligned}\underset{\theta}{E}\{U(w - f - \delta f) + \lambda G(f + \delta f)\} &= \underset{\theta}{E}\{U(w-f) + \lambda G(f)\} \\ &\quad + \underset{\theta}{E}\{(U''(w-f) + \lambda G(f))\delta^2 f\} \\ &< \underset{\theta}{E}\{U(w-f) + \lambda G(f)\}\end{aligned}$$

for non-trivial variations.

the agent given the fee schedule $\langle f \rangle$ (obtained by solving eq. (8.1)), the principal seeks

$$\max_{\langle f \rangle} \mathop{E}_{\theta} \{ U[w(\alpha(\langle f \rangle), \theta) - f(w(\alpha(\langle f \rangle), \theta))] \},$$

subject to a (perhaps market-imposed) constraint on the expected fee schedule or on the expected utility of the agent[4]. Since the principal is also concerned with motivating the agent, i.e. since he must work through $\alpha(\langle f \rangle)$, unlike the agent's problem there is no assurance that the resulting fee schedule will be Pareto efficient. The problem is further complicated if the principal and the agent differ in their subjective distributions or, equivalently, if we fix the distribution of θ for both and recognize that the principal may have a subjective distribution over payoff structures $w(\alpha, \theta)$. In this case the principal will seek

$$\max_{\langle f \rangle} \mathop{E}_{\langle w(\alpha, \theta) \rangle \theta} \{ U[w(\alpha(\langle f \rangle), \theta) - f(w(\alpha(\langle f \rangle), \theta)] \},$$

subject to the above mentioned constraints and where, as before, f may be functionally dependent on θ.[5]

Suppose, however, that the principal is simply unable to say anything about the payoff structure $w(\alpha, \theta)$. In fact, aside from ordinary comparative advantage, one reason for forming an agency relationship in risky situations might be the assumption that the agent is more knowledgeable than the principal. (Or, alternatively, the principal might be certain of $w(\alpha, \theta)$ but ignorant of the agent's assessment.) In addition, it might be too costly to communicate information on their relative probability assessments and, for that matter, the principal might simply find it too costly to even assess a distribution over payoff structures. Finally, even if the principal were to solve the problem exactly, the resulting fee schedule would be dependent on the particular payoff structure $w(\alpha, \theta)$. If the agency relationship were ongoing, then every new problem would

---

[4] Without such a constraint the problem will in general be noncompact and lack a solution. There will be a sequence of fees $\langle f^{\nu} \rangle$ converging to zero while $\alpha(\langle f^{\nu} \rangle)$ converges to the optimal act the principal would choose if he sought

$$\max \mathop{E}_{\theta} \{ U[w(\alpha, \theta)] \}$$

and did not avail himself of the services of the agent.

[5] The exact solution to the principal's problem as it has been posed above is quite difficult to characterize in general. This characterization and a number of related market phenomena are discussed in a second paper [7].

require a new fee schedule and this instability in the fee schedule might severely strain the assumption that computation costs are negligible[6].

Given these considerations the principal might choose an optimal fee schedule by an alternative criterion. In this paper we will explore the implications of using a conservative maximin criterion that insures the principal against suffering an opportunity loss with the 'worst' possible payoff structure. The formalization of this is to choose the optimal fee schedule $\langle f \rangle$ to minimize

$$\Phi \equiv \max_{\langle w(\alpha,\theta)\rangle} \{\max_{\alpha} \underset{\theta}{E}\{U[w(\alpha,\theta) - f(w(\alpha,\theta))]\} - \underset{\theta}{E}\{U[w(\alpha(\langle f \rangle),\theta) - f(w(\alpha(\langle f \rangle),\theta))]\}\} \geq 0.$$

In words, the term inside the first brackets represents an opportunity cost; given the fee schedule, it is the difference between the expected utility of the principal when the optimal act from his viewpoint is chosen, and his expected utility from the agent's choice of an act. The criterion Φ is the maximum of such opportunity costs over all payoff structures and it is clearly non-negative. The principal seeks a fee schedule to minimize this maximum (perhaps subject to some outside constraints of the sort discussed above).

The solution to this problem is actually much simpler than might at first appear. Since the agent chooses an act by maximizing the expected value of his utility function $G(\cdot)$, we need only choose a fee schedule that makes the agent's and the principal's evaluations of payoffs equivalent to ensure that, given the fee schedule, the agent takes the action the principal would wish him to. If we set the fee schedule $\langle f(w) \rangle$ so that for some constants, $a > 0$, b

$$\text{(S)} \qquad U[w - f(w)] = aG[f(w)] + b,$$

then the agent will always choose the act that maximizes the principal's expected utility and $\Phi = 0$. The constants a and b can be chosen to satisfy outside constraints. Intuitively, the fee schedule has been chosen so that the agent and the principal have equivalent utility assessments of wealth and will make the same decisions in all risky situations. We refer to this equivalence as the similarity rule (S).

In the next sections we will examine the implications of using the

---

[6] This point has also been made by Wilson [8].

above similarity rule (S) and/or Pareto efficiency (PE) to define the fee schedule of the agency relationship.

8.3. Similarity, Efficiency, and Linearity

The fee schedules that are actually observed tend to be of a very simple form. Wage contracts are typically fixed payments and bonus compensation is roughly proportional to performance. Mutual fund fees, for example, usually consist of a fixed component, the load, and a variable component taken as a stated percentage ($\frac{1}{4}$–1%) of asset value. Formally, we will say that a fee schedule is linear (L) if

$$\text{(L)} \qquad f(w) = \alpha w + \beta$$

for some constants α, β.

Our first task is to study the relationships between S, PE, and L. Aside from its simplicity, it is not even clear on *a priori* grounds that L is of any theoretical interest, but as the following theorem reveals its role is pivotal.

THEOREM 8.1. *Any two of the three conditions* S, PE *and* L *imply the third.*

PROOF. (i) S and PE ⇒ L.

This result (first obtained by Wilson [9]) highlights the importance of L schedules. The proof is straightforward. Differentiating S we obtain

$$[1 - f']U' = aG'f',$$

and from the necessary condition for Pareto efficiency

$$\text{(PE)} \qquad U' = \lambda G'.$$

It follows that

$$\lambda[1 - f']G' = aG'f'$$

or

$$f' = \left(1 + \frac{a}{\lambda}\right)^{-1} \in (0, 1),$$

and since λ is a constant, integration over w yields L with $\alpha = \{1 + a/\lambda\}^{-1}$ and β an arbitrary integration constant.

(ii) S and L $\Rightarrow$ PE.

Differentiating S with respect to wealth and using L we obtain

$$[1-\alpha]U' = a\,\alpha\,G',$$

and defining $\lambda = a\alpha/[1-\alpha]$ yields PE. (The sufficiency of this for PE follows from the concavity of $U(\cdot)$ and $G(\cdot)$.)

(iii) PE and L $\Rightarrow$ S.

Substituting L into PE we have

$$U'((1-\alpha)w-\beta) = \lambda G'(\alpha w+\beta),$$

and integrating (over w) we obtain

$$(1-\alpha)^{-1}U((1-\alpha)w-\beta) = \lambda\alpha^{-1}G(\alpha w+\beta)+c,$$

where c is a constant of integration. Defining $a = \lambda(1-\alpha)/\alpha > 0$ and $b = c(1-\alpha)$ yields S. *Q.E.D.*

In general, though, we should expect S, PE and L to limit rather severely the class of feasible utility functions. Similarity (S) for example, uniquely defines the fee schedule as parametrized by the two constants a and b. Pareto efficiency, however, also uniquely defines the fee schedule parametrized by λ. It would be surprising, then, if all pairs $\langle U, G\rangle$ allowed the simultaneous satisfaction of S, PE and L. It is clear, though, that given any agent's utility function G (or the principal's utility function U) we can generate a family of 'conformable' principal's (agent's) utility functions by simply setting

$$U((1-\alpha)w-\beta) \equiv a\,G(\alpha w+\beta)+b \tag{8.2}$$

for arbitrary $a > 0$, b, $\alpha \in (0,1)$ and β. Conversely, it is easy to see that a $\langle U, G\rangle$ pair can satisfy S and PE only if there exists $\alpha \in (0,1)$, β and $a > 0$, b such that eq. (8.2) is satisfied. (I am indebted to L. Hurwicz for this observation.) This leads to the following somewhat trivial representation result based on the risk tolerance, $\ell_v \equiv -v'/v''$, of the relevant utility functions. (The risk tolerance is simply the reciprocal of the coefficient of absolute risk aversion.)

THEOREM 8.2 [9]. *If the pair $\langle U, G\rangle$ allow a fee schedule satisfying* S *and* PE (*or, by theorem* 8.1, *any two of* S, PE *and* L), *then*

$$\ell'_u = \ell'_G \tag{8.3}$$

where ℓ_u and ℓ_G are evaluated at $[(1-\alpha)w-\beta]$ and $[\alpha w+\beta]$ respectively.

PROOF. From theorem 8.1 the fee schedule must be linear, hence for some $\alpha \in (0, 1)$ and β:

$$U((1-\alpha)w-\beta) = aG(\alpha w+\beta)+b,$$

and by performing the requisite differentiations we obtain eq. (8.3) in a straightforward manner. *Q.E.D.*

This characterization, however, is not particularly powerful since there is no assurance either that the particular PE fee schedule obtainable from the $\langle U, G\rangle$ pair is reasonable given market conditions or the relative bargaining strengths of the parties, or that $\langle U, G\rangle$ admit any other PE schedules. In general they will not, and we can make a particularly strong statement when $\langle U, G\rangle$ allow a range of alternative fee schedules satisfying S and PE.

THEOREM 8.3. *The pair $\langle U, G\rangle$ allows a range of* PE *fee schedules satisfying* S *if and only if* U *and* G *are members of the linear risk tolerance class with $\ell'_u = \ell'_G$, i.e.*

$$-U'(x)/U''(x) = cx + d \tag{8.4a}$$

and

$$-G'(x)/G''(x) = cx + e, \tag{8.4b}$$

where c, d and e are constants.

PROOF. By a range of PE fee schedules we mean that there is some non-trivial interval of λ weights in PE which can be satisfied with S. Let λ be a value in the interior of the interval. Differentiating PE with respect to λ and using theorem 8.1 we get

$$[-\alpha'w-\beta']U'' = G' + \lambda[\alpha'w+\beta']G''. \tag{8.5}$$

Differentiating PE with respect to w yields

$$[1-\alpha]U'' = \lambda G''. \tag{8.6}$$

The primes on α and β indicate that they are derivatives with respect to λ, and from theorem 8.1 we know that the fee schedules must remain linear as λ is altered. Combining eqs. (8.5) and (8.6) and eliminating U'' we have

$$\ell_G = -G'/G'' = cx + e,$$

where

$$x \equiv \alpha w + \beta, \qquad c \equiv \lambda\alpha'\left[\frac{1}{1-\alpha} + \frac{1}{\alpha}\right],$$

and

$$e \equiv \lambda[\alpha\beta' - \alpha'\beta]\left[\frac{1}{1-\alpha} + \frac{1}{\alpha}\right].$$

From theorem 8.2, $\ell'_U = \ell'_G = c$, and $\ell_U(x) = cx + d$. The converse is straightforward. *Q.E.D.*

Notice that even though we only required λ to be variable within some interval, the result is a global one. The class of utility functions satisfying eq. (8.4) is well known and is composed only of the functions

$$U(X) = \frac{1}{\gamma}(X + \delta)^{\gamma} \quad \text{and} \quad U(X) = \log(X + \delta) \quad \text{and} \quad U(X) = -e^{-\xi X}.$$

If $U(X)$ is required to be concave, then we must have $\gamma \leq 1$ and $\xi \geq 0$.

The formation of a team, then, in the sense that both the agent and the principal have preferences which allow S and PE, places severe restrictions upon their permissible utility functions. In general, though, given any two utility functions, there will exist some fee schedule satisfying S and another satisfying PE (on some interval). In the next section we will try to characterize some of the qualitative properties of these fee schedules.

8.4. Property of Fee Schedule Under Similarity

Assume first that the fee schedule satisfies S but not necessarily PE. Without loss of generality we will assume that $U(\cdot)$ and $G(\cdot)$ have been scaled to eliminate the arbitrary constants a and b. The following lemma characterizes Ω, the domain of wealth values that permits S to be satisfied given $U(\cdot)$ and $G(\cdot)$.

LEMMA. *Let* $\Omega \equiv \{w \mid (\exists f)\ U(w - f) = G(f)\}$. *Then* Ω *is either an unbounded interval closed on the left or* Ω *is null.*

PROOF. Closure follows immediately from continuity. Suppose that $X_0 \in \Omega$:

$$U(X_0 - f_0) = G(f_0).$$

Let $X > X_0$; we wish to show $(\exists f)$

$$U(X - f) = G(f). \tag{8.7}$$

Now

$$U(X - f_0) \geq U(X_0 - f_0) = G(f_0).$$

and for $f \geq f_0 + (X - X_0)$ we have

$$U(X - f) \leq U(X_0 - f_0) = G(f_0) \leq G(f).$$

Hence, by continuity, Bolzano's theorem implies that $\exists f$ satisfying eq. (8.7). *Q.E.D.*

In what follows we will assume that w lies in the interior of Ω, and we will concern ourselves only with concave $U(\cdot)$ and $G(\cdot)$. Differentiating S implicitly we have

$$f'(w) = \frac{U'}{U' + G'} \in [0, 1].$$

The marginal fee is positive, unless the principal's marginal utility is zero, but it is less than unity to insure that the principal's marginal share is non-negative. This is precisely what we would expect S to imply.

Differentiating again yields

$$\begin{aligned} f'' &\sim \frac{U''}{(U')^2} - \frac{G''}{(G')^2} \\ &= \frac{R_G}{fG'} - \frac{R_U}{(w - f)U'} \\ &\sim \frac{(w - f)U'}{fG'} - \frac{R_U}{R_G} \\ &= \frac{f'}{f}\bigg/\frac{(1 - f')}{(w - f)} - \frac{R_U}{R_G}, \end{aligned} \tag{8.8}$$

where $\sim$ stands for 'has the same sign as' and R_v denotes the coefficient of relative risk aversion $R_v \equiv -XV''(X)/V'(X)$.

Thus the fee schedule is concave or convex as the ratio of the wealth elasticities of the agent's to the principal's share falls short of or exceeds the reciprocal of the ratio of their coefficients of relative risk aversion. Unfortunately this is not a very illuminating criterion, and what we would really like is a more direct means of deducing the shape and, specifically, the concavity or convexity of the fee schedule from the characteristics of the agent's and the principal's utility functions. To do so, however, we will need some preliminary lemmas on the asymptotic properties of concave utility functions. These results are somewhat tangential to the main argument and have been put in the appendix. In addition, the proofs of theorems 8.4–8.6 are somewhat technical and have also been deferred to the appendix. We begin with a theorem applicable when $U(\cdot)$ and $G(\cdot)$ are unbounded from above. It verifies the intuition that if U is more risk averse than G, then the fee schedule must (eventually) be concave to induce the agent to view risky choices as the principal does. Conversely, if the agent is more risk averse than the principal, the fee schedule is (eventually) convex to motivate the agent to accept additional risk.

THEOREM 8.4. *Assume that condition* S *holds. If U and G are unbounded above, and if*

$$\text{(i)}\ \mathbf{R}_U \equiv \lim_{X\in\Omega} \inf R_U > \lim_{X\in\Omega} \sup R_G \equiv \bar{R}_G,$$

then $(\exists \hat{w})(\forall w > \hat{w})$ *the fee schedule is concave and if*

$$\text{(ii)}\ \mathbf{R}_G \equiv \lim_{X\in\Omega} \inf R_G > \lim_{X\in\Omega} \sup R_U \equiv \bar{R}_U,$$

then $(\exists \hat{w})(\forall w > \hat{w})$ *the fee schedule is convex.*

PROOF. See the appendix.

The next theorem deals with the case illustrated in fig. 8.1 (see appendix). If theorem 8.4 was intuitive, theorem 8.5 may appear counterintuitive.

THEOREM 8.5. *Assume that condition* S *holds. If U is bounded above and* $\lim\inf R_U > 0$, *and if* $(\exists w_0)\ G(w_0) > \sup_{X\in\Omega} U(X)$, *then* $(\exists \hat{w})(\forall w > \hat{w})$ *the* *fee schedule is concave. Conversely if G is bounded above and* lim inf $R_G > 0$, *and if* $(\exists w_0)\, U(w_0) > \sup_{X \in \Omega} G(X)$, *then* $(\exists \hat{w})(\forall w > \hat{w})$ *the fee schedule is convex.*

PROOF. See the appendix.

The need to require lim inf $R_u > 0$ in the proof of theorem 8.5 can be illustrated by counterexample. From lemma 8.2 (see appendix) the boundedness of U assures that lim sup $R_u \geq 1$, but there is nothing to prevent R_u from cyclically approaching zero as $w \to \infty$. In fact, take any bounded, concave, monotone increasing function $U(\cdot)$ and define $V(\cdot)$ to be the function formed by the chords joining $U(n)$ and $U(n+1)$ where n runs over the integers. $V(\cdot)$ can clearly be smoothed to make any desired finite degree of differentiability (i.e. high contact) on, say, $[n - \varepsilon, n + \varepsilon]$ where $\varepsilon < \frac{1}{2}$, but while $V(\cdot)$ is bounded $R_n = 0$ on $(n + \varepsilon, n + 1 - \varepsilon)$ and, hence, lim inf $R_n = 0$ and $f''(w) > 0$ on all intervals of the form $(n - \varepsilon, n + 1 - \varepsilon)$.

The final theorem concerns the case where both U and G are bounded by the same bound. This will clearly exhaust the cases, and reverse our intuition.

THEOREM 8.6. *Assume that condition* S *holds. If U and G are both bounded above with*

$$\sup_{X \in \Omega} U(X) = \sup_{X \in \Omega} G(X)$$

and if $R_u(X) \to R_u^*$ and $R_G(X) \to R_G^*$, then $(\exists w_0)(\forall w > w_0)$ *the fee schedule is convex or concave as $R_u^* > R_G^*$ or $R_u^* < R_G^*$.*

PROOF. See the appendix.

This result is clearly much weaker than our previous ones but it does not appear possible to strengthen it much. By the mean value theorem it can be shown that there exists a non-negative function θ_w such that

$$\frac{(w - f)U'}{fG'} = \frac{1 - R_u(w + \theta_w - f(w + \theta_w))}{1 - R_G(f(w + \theta_w))}$$

but if R_u and R_G do not converge to limit values, theorems such as we obtained for the unbounded utility cases will not be available. Suppose that we were to assume that $\mathbf{R}_G > \bar{R}_u$. By boundedness we have $\bar{R}_u \geq 1$, but we will also assume that $\mathbf{R}_u > 1$. It follows that we can bound f'' above in sign by

$$f'' = \frac{(w-f)U'}{fG'} - \frac{R_u}{R_G} = \frac{1-R_u}{R_G} - \frac{R_u}{R_G} < \frac{\bar{R}_u - 1}{R_G - 1} - \frac{\mathbf{R}_u}{\bar{R}_G}$$

which will be positive for $\bar{R}_G$ or $\bar{R}_u$ sufficiently large even when the criterion $\mathbf{R}_G > \bar{R}_u \geq \mathbf{R}_u > 1$ is satisfied. Unless the bound can be made tighter or unless we assume R_u and R_G are sufficiently smooth to assure that $(1 - R_u)/(1 - R_G)$ actually 'tracks' $(w-f)U'/fG'$, it seems difficult to get a stronger result.

It remains to explain our somewhat counterintuitive results for the case where one of the utility functions is bounded above. When U and G are unbounded, $\mathbf{R}_u > \bar{R}_G$ implies that the principal is more risk averse than the agent and, as intuition might suggest, we set up a concave fee schedule to 'riskify' the agent. Conversely, a convex schedule is required by S when $\mathbf{R}_G > \bar{R}_u$. On the other hand, if both U and G have the same upper bound and the limit values of R_u and R_G exist, we choose a convex schedule if $R_u > R_G$ and a concave one if $R_u < R_G$. From a strictly mathematical (or geometric) viewpoint the issue is clear. If both U and G are bounded with the same bound, the one with the greater risk aversion approaches the bound more rapidly. As a consequence its share must be a concave function of wealth and the other is convex ($f'' \lesseqgtr 0$ as $d^2(w-f)/dw^2 = -f'' \gtreqless 0$). The argument is similar when one utility is bounded below the other. From an economic standpoint, though, the issue is more murky. In particular, if $R_u > R_G$ but the fee schedule is convex, how can the agent, who is less risk averse than the principal, possibly mirror the principal's preferences toward risk? The answer lies in the mathematics of concave transforms. While a concave f is concavified further by a concave G to $G(f)$, it is not necessarily true that a concave $G(\cdot)$ is less risk averse than $G(f(\cdot))$ where f is concave. This point has nothing to do with the fact that the fee is subtracted from terminal wealth to give the principal's return. In fact, if f is convex, $w-f$ is concave and the paradox is exacerbated even further.

The case where some third party pays the agent is simpler and interesting in its own right and since it can produce the same paradoxes we will study it. Now S is modified to

$$\text{(S}'\text{)} \quad U(w) = G(f(w)], \quad f' = U'/G' \quad \text{and} \quad f'' \sim (wU'/fG') - (R_u/R_G),$$

as we obtained from S. Some algebra yields

$$R_f \equiv -wf''/f' = R_u - (wf'/f)R_G.$$

Since G is concave, $R_G > 0$ and $R_u > R_f$ for all w, i.e. U is uniformly more risk averse than f. Notice, though, that even when f is concave, $U(\cdot)$ is not necessarily more risk averse than $G(\cdot)$, with G evaluated not at $f(w)$ but at w. For example, if we choose $R_G > 1$ to be constant and f concave with $R_f < 1$, then $R_u < R_G$. To verify this, let

$$f(w) = w^{1-R_f} \qquad \text{and} \qquad G(w) = -w^{1-R_G}.$$

Now

$$U(w) = G[f(w)] = -(w^{1-R_f})^{1-R_G} = -w^{(1-R_f)(1-R_G)}$$

and

$$0 > (1 - R_f)(1 - R_G) > 1 - R_G$$

implies that

$$R_u = 1 - (1 - R_f)(1 - R_G) < R_G.$$

Similarly, an example can be constructed where f is convex but $R_u > R_G$.

To sum up, the concavity of f is neither necessary nor sufficient to insure that U is more concave than G. The case where f is convex is equally equivocal. Our intuition was misguided; R_u uniformly greater than R_G requires only that

$$R_f > R_G(w) - (wf'/f)R_G(f) = [1 - (wf'/f)]R_G$$

when R_G is constant, and this can be satisfied even if f is convex and is not necessarily satisfied if f is concave; we require f to be sufficiently concave.

An alternative approach that affords an interesting look at the problem is the following. U is said to be more concave than G if and only if there exists a *concave* transform $H(\cdot)$ such that $U = G[f(w)] = H[G(w)]$.[7] This condition is equivalent to requiring that $R_u(w) > R_G(w)$. Differentiating once yields $G'(f)f' = H'G'(w)$, and again,

$$H'' \sim G''(f)(f')^2 + G'(f)f'' - H'G''(w), \tag{8.9}$$

which can be positive if $G''(w)$ is sufficiently negative relative to $G''(f)$.

[7] This terminology and the equivalence of this definition of greater concavity and the definition of greater risk aversion are due to Pratt [5].

In particular, then, the concavity of the fee schedule depends not only on *whether* U and G are concave, but also on the magnitude of their concavity through eq. (8.9).

As a final example suppose $f(w) = \alpha w + \beta$. Now,

$$R_f = 0 = R_u - \left(\frac{\alpha w}{\alpha w + \beta}\right) R_G$$

or

$$R_u(w) = \left(\frac{\alpha w}{\alpha w + \beta}\right) R_G \gtreqless R_G(f)$$

as $\beta \lesseqgtr 0$. If R_G is constant, U can be more or less concave than G even though f is linear. If $\beta = 0$, however, intuition returns and U and G are identically concave when R_G is constant. Even when $\beta = 0$, though, if R_G is not constant, then we are still unsure of the relative concavity of U and G; we only know that

$$R_u(w) = R_G(f(w)) = R_G(\alpha w) \gtreqless R_G(w).$$

8.5. Property of Fee Schedule Under Efficiency

It is also of interest to consider the properties of those fee schedules defined by Pareto efficient (PE) bargains between agent and principal. Fortunately, by appropriate interpretation of our previous results most of the work has already been done. Let us assume then that PE defines the fee schedule and put aside S. If we substitute U' for U and G' for G, the PE rule $U'(w - f) = \lambda G'(f)$ is formally identical to S. Now

$$f'(w) = \frac{U''}{U'' + \lambda G''} \in (0, 1)$$

if U and G are strictly concave. As with S, we will suppress the constant λ into the $G(\cdot)$ function since we will not have occasion to vary it in what follows.

If U''', $G''' > 0$, then $-U'$ and $-G'$ are monotone strictly concave functions and the identification is complete. In particular, defining

$$P_u \equiv -U'''X/U'' \qquad \text{and} \qquad P_G \equiv -G'''X/G'',$$

theorems 8.4, 8.5 and 8.6 hold with P_u substituted for R_u and P_G for R_G. Is U''', $G''' > 0$ likely or even reasonable? Many have argued that the assumption of a decreasing coefficient of absolute risk aversion, $A \equiv -U''/U'$, is quite sensible. In particular, it is a sufficient condition for the risky asset to be superior in a two asset world with a riskless and a risky asset. Furthermore, the risk premium for a sufficiently small gamble will decline with increasing wealth if and only if A declines (see ref. [5]). As is easily shown, if absolute risk aversion is to decline with wealth, then the utility function must have positive third derivatives.

8.6. Conclusion

This paper represents a first and tentative attempt at developing the economic theory of agency. A number of extensions suggest themselves. The similarity condition (S) is easily generalized to the many-agent–many-principal problem, as is the criterion of Pareto efficiency. The results of this paper pose no additional difficulties in this context, but with many principals and agents, problems of coordination of information arise of the sort studied in the theory of teams, most notably by Marschak and Radner [4]. In its many-player form, a theory similar to that of agency was developed as the theory of syndicates by Borch [3] and Wilson [8]. Borch [3] first proved the PE condition, although in a somewhat different problem, and Wilson [9] suggested condition S as a way for a syndicate to motivate its managers and showed that S and PE imply L.

In a subsequent paper we develop the theory of agency from the principal's exact constrained maximum problem (as posed initially) and make a start towards embedding agency in a general market context. The problems raised by the existence of intermediaries, or agents, in general equilibrium models are quite difficult, but it is not unlikely that research in this area will aid our understanding of much of the institutional structure of modern economies.

References

[1] Arrow, K. J. *Theory of Risk-Bearing*. Markham: Chicago (1971).
[2] Berhold, M. A theory of linear profit sharing incentives. *Quarterly Journal of Economics*, **85** (1971), 460–482.

[3] Borch, K. Equilibrium in a reinsurance market. *Econometrica*, **30** (1962), 424–444.
[4] Marschak, J. and Radner, R. *The Economic Theory of Teams*. Yale University Press: New Haven and London (1972).
[5] Pratt, J. W. Risk aversion in the small and in the large. *Econometrica*, **32** (1964), 122–137.
[6] Raiffa, H. *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*. Addison-Wesley: Reading, Massachusetts (1968).
[7] Ross, S. The economic theory of agency: the principal's problem. *American Economic Review Proceedings* (1973).
[8] Wilson, R. B. On the theory of syndicates. *Econometrica*, **36** (1968), 119–132.
[9] Wilson, R. B. The structure of incentives for decentralization under uncertainty. In *La Décision*. Éditions du Centre national de la Recherche scientifique: Paris (1969).

APPENDIX

LEMMA 8.1. *Let $U(\cdot)$ be a concave monotone increasing function; then $U(\cdot)$ is bounded only if*

$$\lim_{w \to \infty} wU'(w) = 0$$

and if $(\exists\ \varepsilon > 0)$ *such that*

$$\lim_{w \to \infty} w^{1+\varepsilon} U'(w) = 0$$

then $U(\cdot)$ is bounded.

PROOF. Since $U'(\cdot)$ is non-negative, lim sup $wU'(w) \geq 0$. If lim sup $wU'(w) > 0$, then there exists a divergent sequence $\langle w^{\nu} \rangle$ and a $\delta > 0$ such that

$$w^{\nu}U'(w^{\nu}) > \delta > 0 \qquad \text{or} \qquad U'(w^{\nu}) > \delta / w^{\nu}.$$

Since $U'(\cdot)$ declines monotonically, we can thus bound $U'(w)$ from below by a simple function $h(\cdot)$ with steps of height δ / w^{ν}. Integrating back we obtain

$$U(X) - U(X_0) = \int_{X_0}^{X} h(w) dw \geq \sum_{\nu=1}^{\nu_x} \frac{\delta}{w^{\nu}} (w^{\nu} - w^{\nu-1}),$$

where $w^0 \equiv 0$ and v_x is chosen so that

$$w^{v_x} < X \leq w^{v_x + 1}.$$

Since $\langle w^v \rangle$ is a divergent sequence it is possible to pass to a subsequence $\langle w^i \rangle$ such that

$$(w^i - w^{i-1})/w^i > 1/i,$$

implying that

$$U(\infty) - U(X_0) \geq \sum_{v=1}^{\infty} \frac{\delta}{w^v}(w^v - w^{v-1}) \geq \sum_{i=1}^{\infty} \frac{\delta}{w^i}(w^i - w^{i-1}) \geq \sum_{i=1}^{\infty} \frac{\delta}{i} = \infty$$

and $U(\cdot)$ is unbounded. This establishes the first proposition by contradiction.

Now suppose that $(\exists\, \varepsilon > 0)$ such that

$$\lim_{w \to \infty} w^{1+\varepsilon} U'(w) = 0.$$

Defining

$$c(w) \equiv w^{1+\varepsilon} U'(w),$$

we have

$$U'(w) = \frac{c(w)}{w^{1+\varepsilon}}$$

or

$$U(w) - U(w_0) = \int_{w_0}^{w} \frac{c(X)}{X^{1+\varepsilon}} \mathrm{d}X \leq \{\sup_{X > w_0} c(X)\} \int_{w_0}^{w} \frac{\mathrm{d}X}{X^{1+\varepsilon}}$$

$$= \{\sup_{X > w_0} c(X)\} \left[\frac{w_0^{-\varepsilon}}{\varepsilon} - \frac{w^{-\varepsilon}}{\varepsilon}\right] < \{\sup_{X > w_0} c(X)\} \frac{w_0^{-\varepsilon}}{\varepsilon}.$$

However, since $c(w) \to 0$ as $w \to \infty$, $\sup_{X > w_0} c(X)$ must be finite for some w_0.
Q.E.D.

LEMMA 8.2. *Let $U(\cdot)$ be a concave monotone increasing function; then $U(\cdot)$ is bounded if* lim inf $R(\cdot) > 1$ *and, if $U(\cdot)$ is bounded then* lim sup $R(\cdot) \geq 1$.

PROOF. If lim inf $R(\cdot) > 1$, then $(\exists w_0, \varepsilon > 0)(\forall w > w_0)$, $R(w) > 1 + 2\varepsilon$.

Now

$$R(w) \equiv -wU''(w)/U'(w)$$

implies that

$$U'(w) = c\exp\left(-\int_{w_0}^{w}\frac{R}{z}\,dz\right) < c\exp\left(-\int_{w_0}^{w}\frac{1+2\varepsilon}{z}\,dz\right) = c\left(\frac{w_0}{w}\right)^{1+2\varepsilon}.$$

Hence $w^{1+\varepsilon}U'(w) < cw_0^{1+2\varepsilon}/w^{\varepsilon} \to 0$ and, by lemma 8.1, $U(\cdot)$ is bounded.

Assume that $(\exists X_0)(\forall X > X_0)R(X) \leq 1$. It follows that

$$U'(w) \geq c\exp\left(-\int_{w_0}^{w}\frac{dz}{z}\right) = \frac{cw_0}{w}$$

and by lemma 8.1, $U(\cdot)$ is unbounded. This result is due to Arrow [1], and is actually somewhat stronger than we require. Of course, it now follows that lim sup $R(\cdot) \geq 1$.

THEOREM 8.4. *Assume that condition* S *holds. If U and G are unbounded above, and if*

$$\text{(i)}\ \mathbf{R}_U \equiv \liminf_{X\in\Omega} R_U > \limsup_{X\in\Omega} R_G \equiv \bar{R}_G,$$

then $(\exists\hat{w})(\forall w > \hat{w})$ *the fee schedule is concave and if*

$$\text{(ii)}\ \mathbf{R}_G \equiv \liminf_{X\in\Omega} R_G > \limsup_{X\in\Omega} R_U \equiv \bar{R}_U,$$

then $(\exists\hat{w})(\forall w > \hat{w})$ *the fee schedule is convex.*

PROOF. If U and G are both unbounded it follows immediately that Ω is an unbounded interval and that both f and $w - f$ are unbounded.

From

$$(d/dX)[X\,U'(X)] = X\,U'' + U' = [1 - R_U]U'$$

we have

$$X\,U'(X) = \int_{X_0}^{X}[1 - R_U]\,U'\,dz + X_0U_0'$$

$$\equiv [U - U_0]\,\theta_U + X_0U_0',$$

and similarly

$$X\,G'(X) \equiv [G - G_0]\theta_G + X_0G_0'.$$

It follows from eq. (8.8) that

$$f''(w) \sim \frac{(w-f)U'}{fG'} - \frac{R_U}{R_G}$$

$$= \frac{\theta_U[U-U_0] + (w_0 - f_0)U'_0}{\theta_G[G-G_0] + f_0 G'_0} - \frac{R_U}{R_G}$$

$$= \left[\left\{\theta_U + \frac{(w_0 - f_0)U'_0}{(G-G_0)}\right\} \Big/ \left\{\theta_G + \frac{f_0 G'_0}{(U-U_0)}\right\}\right] - \frac{R_U}{R_G}.$$

Now, if (i) holds, then, since $U(\cdot)$ is unbounded from above, lemma 8.2 implies that

$$\limsup_{X\in\Omega} R_G < \liminf_{X\in\Omega} R_U \leq 1.$$

Hence $(\exists\delta > 0$ and $w_0)(\forall w > w_0)$, $R_G(w) < 1 - \delta$, and therefore $(\forall w > w_0)$, $\theta_G > \delta$.

Hence, if $\bar{R}_G > 0$, then

$$\limsup f''(w) = \limsup \left\{\frac{\theta_U}{\theta_G} - \frac{R_U}{R_G}\right\} \leq \frac{1 - \mathbf{R}_U}{1 - \bar{R}_G} - \frac{\mathbf{R}_U}{\bar{R}_G}$$

$$= (\bar{R}_G - \mathbf{R}_U)([1 - \bar{R}_G]\bar{R}_G)^{-1} < 0,$$

where we have made use of $\bar{R}_G < 1$. If $\bar{R}_G = 0$ then R_U/R_G diverges and the result is immediate.

Proposition (ii) may be proved interchanging the role of U and G in the statement of the proof. *Q.E.D.*

THEOREM 8.5. *Assume that condition* S *holds. If U is bounded above and* lim inf $R_U > 0$, *and if*

$$(\exists w_0)\, G(w_0) > \sup_{X\in\Omega} U(X),$$

then $(\exists\hat{w})(\forall w > \hat{w})$ *the fee schedule is concave. Conversely, if G is bounded above and* lim inf $R_G > 0$, *and if*

$$(\exists w_0)\, U(w_0) > \sup_{X\in\Omega} G(X),$$

then $(\exists\hat{w})(\forall w > \hat{w})$ *the fee schedule is convex.*

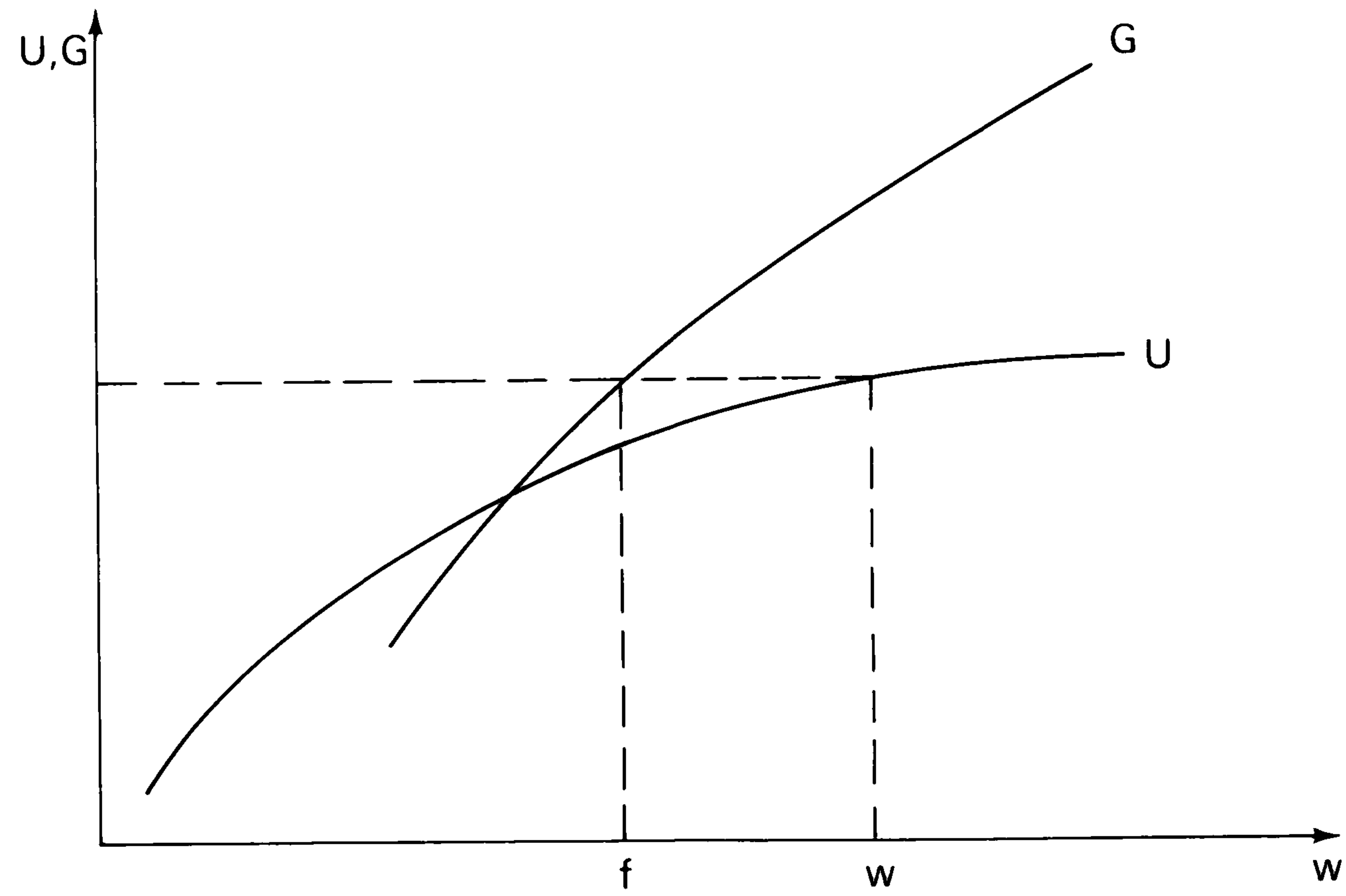

Fig. 8.1.

Proof. Figure 8.1 illustrates the situation where U is bounded and G rises above U. Since

$$G(f) = U(w - f) \leq \sup_{X \in \Omega} U(X) < \infty,$$

it must be the case that $f(w) \to \bar{f} < \infty$. From monotonicity $fG' \to \bar{f}G'$, $(\bar{f}) > 0$, and from lemma 8.1 the boundedness of $U(\cdot)$ implies that

$$(w - f)\, U'(w - f) \to (w - \bar{f})\, U'(w - \bar{f}) \to 0.$$

It follows that

$$(w - f)U'/fG' \to 0$$

and, since lim inf $R_U > 0$, $(\exists \varepsilon > 0$ and $w_0)$ such that $(\forall w > w_0)$

$$\frac{R_U}{R_G} > \frac{\varepsilon}{R_G(\bar{f})} > 0.$$

Hence $(\exists \hat{w}, \delta > 0)(\forall w > \hat{w})$

$$f''(w) \sim \frac{(w - f)U'}{fG'} - \frac{R_U}{R_G} < -\delta < 0.$$

The converse proposition is proved in an identical fashion. *Q.E.D.*

THEOREM 8.6. *Assume that condition* S *holds. If* U *and* G *are both bounded above with*

$$\sup_{X\in\Omega} U(X) = \sup_{X\in\Omega} G(X)$$

and if

$$R_u(X) \to R_u^* \qquad \text{and} \qquad R_G(X) \to R_G^*,$$

then $(\exists w_0)(\forall w > w_0)$ *the fee schedule is convex or concave as* $R_u^* > R_G^*$ *or* $R_u^* < R_G^*$.

PROOF. By strict monotonicity both U and G attain their suprema only at ∞; hence both f and $w - f$ must be unbounded. (If, for example, $f \to \bar{f} < \infty$, then by monotonicity

$$G(\bar{f}) < \sup_{X\in\Omega} G(X).$$

However,

$$U(w - f) \to U(w - \bar{f}) \to \sup_{X\in\Omega} U(X) = \sup_{X\in\Omega} G(X),$$

violating $U(w - f) = G(f)$ for sufficiently large w.) From lemma 8.1 we have

$$(w - f)U' \to 0 \qquad \text{and} \qquad fG' \to 0.$$

Suppose first that $R_G^* > R_u^*$. From lemma 8.2 we must have $R_u^* \geq 1$. Applying l'Hospital's rule we have

$$\lim_{w\to\infty} \frac{(w - f)U'}{fG'} = \lim_{w\to\infty} \frac{[1 - R_u]U'}{[1 - R_G]G'} \frac{(1 - f')}{f'} = \lim_{w\to\infty} \frac{1 - R_u}{1 - R_G} = \frac{1 - R_u^*}{1 - R_G^*}.$$

From eq. (8.8) it now follows that

$$f''(w) \sim \frac{(w - f)U'}{fG'} - \frac{R_u}{R_G} \to \frac{1 - R_u^*}{1 - R_G^*} - \frac{R_u^*}{R_G^*} < 0.$$

The converse result is obtained by applying the above analysis to

$$f''(w) \sim \frac{R_G}{R_u} - \frac{fG'}{(w - f)U'}.$$

Q.E.D.

COMMENTS

On the theory of economic agency

Leonard Mirman

C8.1.1. Discussion

Professor Ross has presented an important discussion of the problem of agency. Although the name of the economic relationship described and analyzed is novel, the content extends the work of Wilson [4]. It is also similar to the problems studied by Berhold on linear profit sharing incentives [1]. The importance of the work is in laying a sound theoretical structure for the study of observed relationships in the real world. Among the more important areas in which economic relationships can be described by Ross's theoretical structure are incentive contracting, share cropping, profit sharing, insurance, etc. The connecting theme of these diverse areas of economic institutions is the relationship between two competing – although, in some sense, cooperating – agents who must decide how to share the return and 'risk' of a random variable which requires inputs from at least one of the agents in a way that is consistent with the risk preferences of both agents.

More precisely the object of the principal is to provide an incentive for the agent to choose an input, from the set of feasible inputs, so as to maximize the agent's expected utility while simultaneously providing the highest possible expected utility for the principal. His only problem is to decide on the payoff function which is consistent with his own utility maximization while giving the agent enough incentive for maximizing his own expected utility and to take the risk the principal would have him take. The difficulty is that there might be a divergence of attitudes toward risk between the principal and agent, in which case the agent might choose an input which maximizes his own expected utility which is not optimal for the principal. For example, let $w(\alpha, \theta)$ be the payoff function for input α and state of the world θ, the payoff to the agent being $f(w(\alpha, \theta))$. Suppose there are two possible inputs α_1

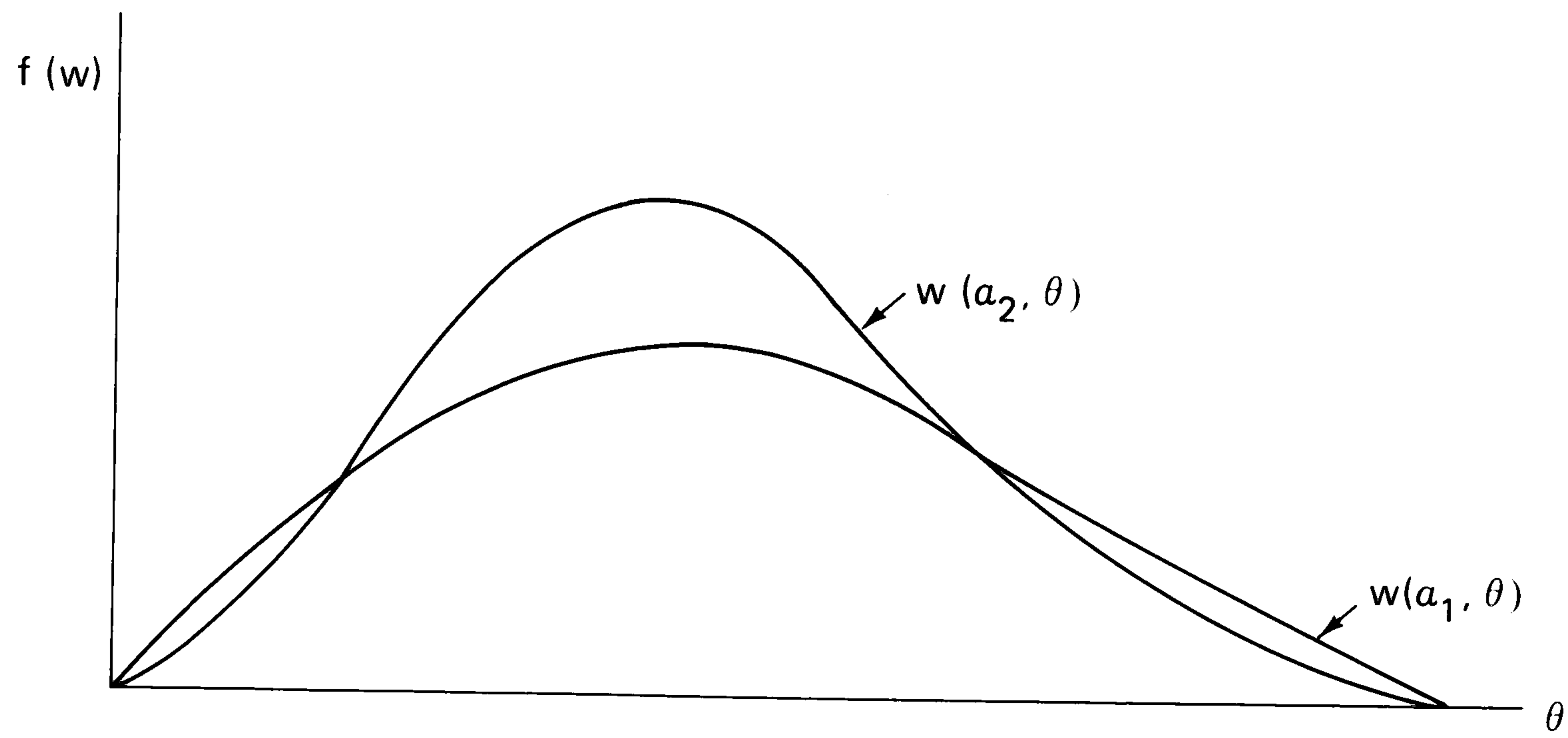

Fig. C8.1.

and α_2 as in fig. C8.1. Suppose that the principal prefers α_1 over α_2 but the agent prefers α_2 over α_1. It is then important for the principal to choose the payoff function $f(w)$ which will provide an incentive for the agent to choose α_1 over α_2. The wrong incentive will lead to α_2 rather than α_1 being chosen.

Choice criteria used by Ross to derive decision rules which correspond to those observed in the world are the principle of similarity and Pareto optimality. The principle of similarity turns out to be the criterion – used previously by Wilson – on which the principal chooses his fee schedule. The similarity criterion is motivated by its relationship to a minimax strategy followed by the principal when minimizing the loss due to the '"worst" possible payoff structure'. Pareto optimality is based on an efficiency criterion and yields an efficient allocation for each value of the random payoff.

The similarity criterion plays the crucial role in selecting a payoff function which equates the attitudes towards risk of the principal and the agent. In view of the fact that linear payoff functions are usually observed in real world situations and the expectation that Pareto optimal relations will arise from utility maximizing behavior, it seems reasonable to believe that there is a strong relationship between linear payoff functions, Pareto optimality and similarity. Indeed, this is the content of theorem 8.1, which shows that any two of the three imply the other.

Unfortunately there are limitations in simultaneously assuming Pareto optimality and similarity which are discussed in theorem 8.3. It is shown that restrictions of the type implied by Pareto optimality and similarity restrict the admissible class of utility functions. It is interesting, however, that the class of utility functions allowed are the Cobb–Douglas and exponential type utility functions. An important aspect of the Ross model is that the principal does not – as the agent does – decide on an action. The type of models which readily come to mind in this situation are those used in the theory of teams [3] and in the theory of games [2]. The agency relationship is different from each of these, since like the team and unlike a game situation there is an incentive for both the principal and agent to cooperate, i.e. after a sharing rule is determined both agent and principal have a stake in having the best decision made. Although unlike a team, the decision makers do not have the same interests. In the spirit of the theory of games, interests are competitive although, in the choice of a sharing rule, the principal makes the decision keeping the agent's reactions in mind.

Finally, some extensions of the models should consider differences in information between principal and agent as well as differences in subjective probabilities. The question of how new information affects the choice of a sharing rule and the final choice of an input would be interesting. In addition, simultaneous decisions on the part of the principal and agent for consumption or other projects have been assumed away and should eventually enter the model.

References

[1] Berhold, M. A theory of linear profit sharing incentives. *Quartely Journal of Economics,* **85** (1971), 460–482.
[2] Luce, R. D. and Raiffa, H. *Games and Decisions.* Wiley: New York (1958).
[3] Marshak, J. and Radner, R. *The Economic Theory of Teams.* Yale University Press: New Haven (1972).
[4] Wilson, R. On the theory of syndicates. *Econometrica,* **36** (1968), 119–132.

PART 3

CHAPTER 9

NOTES ON WELFARE ECONOMICS, INFORMATION AND UNCERTAINTY

J. A. Mirrlees

9.1. *Introduction*

These notes, although not entirely tentative, are less systematic than I would wish. Their purpose is to show how, in a world with imperfect and unreliable information, the Arrow–Debreu framework[1] for welfare economics is unsatisfactory, and that models akin to those now used in the theory of public finance may be more appropriate. I shall not discuss all the reasons that urge one to extend or avoid the standard models. In particular, as Arrow[2] and others have pointed out, the production of information – as in medical care, invention or, presumably, the educational system – requires special treatment; but I shall not say much about it.

What I am going to consider is the information a government might have about consumers, or consumers about government. I shall be thinking of welfare economics as a mode of discussing alternative government policies, and also as part of the discussion about alternative systems of government. The policies a government can adopt, and the policies it should adopt, depend upon information about consumers, what they *do* and what they *are*. Thus the fundamental theorem of welfare economics invites governments to distribute to households quantities of resources that are a function of what the households are (not what they choose to do), in the hope that the right competitive equilibrium will establish itself. In the Arrow–Debreu theory, this distribution of resources is carried out in advance of knowing the state of

---

[1] See ref. [1], chapter 4.
[2] [1], chapters 6 and 8.

the world, but with complete information about the characteristics of the households. A great many insurance and futures markets are required if the optimum is to be a competitive equilibrium, but I shall not concern myself directly with that well-known difficulty. I shall ask what should be done – for reasons to be explained – (1) if the distribution depends on the state of nature, and (2) if information about the characteristics of households is imperfect. I shall also include some remarks on preferences regarding uncertainty.

9.2. Redistribution and Risk-taking

The Arrow–Debreu formulation of welfare economics accepts each household's beliefs – possibly expressible by means of subjective probabilities – in the same way that it accepts the household's tastes. If a man believes strongly but wrongly that the end of the world is at hand, he will be given his wealth now and allowed to spend it all at once. He will then starve, in circumstances he believed would not occur, but an Arrow–Debreu welfare function does not care. We should like to be able to discuss policies for a government that does care about such outcomes and is, in some respects, better informed about possible states of nature than some of the households for which it claims responsibility. Among these policies would be income distributions that are a function of the state of nature[3].

Another reason for studying such distributions is the impossibility of identifying all states of nature 'objectively' – this is the phenomenon of moral hazard, well recognized in the literature. A farmer cannot perfectly insure his crop against adverse circumstances, for the degree of adversity can, in practice, be assessed perfectly only by looking at the size of the crop, and that is affected by actions under the farmer's control. If perfect insurance were possible, distribution by government could (apart from the difficulties just mentioned) without disadvantage be independent of the state of nature. Since perfect insurance is not possible, the government presumably ought to relate distribution to the actual outcome of the harvest, as a proxy for the state of nature.

In this case, we clearly have what has come to be called a 'second-best'

[3] This distinction between *ex ante* and *ex post* optimality is fairly well known. See ref. [2].

problem – that is, a problem that will not have for solution the straightforward competitive equilibrium familiar in simpler problems. The term 'second-best' may be a bit misleading since, in such a case as that of the risk-taking farmer – or the assistant professor – the first-best is an even more unattainable theoretical construction than the 'second-best'.

The first case, of 'Allais optimality', seems to have an easy formal solution, at least in the extreme case where household's probability beliefs are irrelevant, and policies are discussed in terms of a welfare function which has as arguments each household's utility in each state of nature (for example, expected welfare, with the government's probabilities, or yours or mine). As usual one wants to have shadow prices for each commodity in each state of nature. Producers should maximize profits (interpreting the shadow prices conditional upon the state of nature); and households should work out their plans separately for each state of nature, in the light of these prices and the government's plans for distribution of wealth, which are conditional upon the state of nature. Formally, if x_s^h is the vector of h's consumption in state of nature s, $u_s^h = u^h(x_s^h)$ is h's utility[4], we want to maximize

$$W(\dot{u}) = W(u_1^1, u_2^1, \ldots, u_1^2, u_2^2, \ldots) \qquad \text{(monotonically increasing)}$$

with the competitive equilibrium conditions satisfied for each assigned budget b_s^h. Then, aggregate excess demands $\Sigma_h x_s^h$ in the various states of nature being feasible together, the production plan $(y_1, y_2 \ldots)$ (equal to aggregate excess demands in the various states of nature) maximizes $\Sigma p_s \cdot y_s$, and each x_s^h maximizes u_s^h subject to $p_s \cdot x_s^h$ being no greater than b_s^h where p_s is the competitive price vector for state s. This is a standard argument.

The problem associated with 'Allais optimality' is that consumers must not trade insurance: it must be impossible to trade a quantity of a commodity contingent upon a different state. If there were perfect contingent markets between consumers it would not in general be desirable to allow producers to trade in these same markets: one would want to see commodity taxes imposed, for example. The Arrow–Debreu equilibrium would not be optimal. We shall come upon this need to prohibit markets again.

---

[4] The utility function itself could vary with the state of nature s.

When the moral-hazard aspect of the economy is brought in, even this rather unstable competitive result is not optimal. I have no interesting general results as yet, but the following special example seems to capture the essence of the matter[5].

Consider an economy of independent peasant farmers, producing corn on their own farms with their own labor-enterprise-and-attention. The probability density of corn output y when there has been labor input z is $f(y, z)$. The government relates the farmer's consumption x to his output through a function

$$x = c(y) \tag{9.1}$$

which represents redistribution of output between farmers. The farmers, identical to one another, and each interested only in himself, choose z to maximize expected utility

$$\int u(x, y, z) f(y, z) \mathrm{d}y \tag{9.2}$$

subject to eq. (9.1). Let us take u to be a concave increasing function of x, and a decreasing function of y and z. I shall also assume that u tends to $-\infty$ when x tends to zero, so that farmers would give first priority to avoiding zero consumption. The government accepts eq. (9.2) as its own welfare function, so that all the farmers should find its policies appealing. Everything has to take place under the aggregate production constraint which, on the assumption of a very large number of farmers, insensitive to small variations in consumption and with stochastically independent production possibilities, can be taken to be

$$\int y f(y, z) \mathrm{d}y - \int c(y) f(y, z) \mathrm{d}y = 0. \tag{9.3}$$

There is no harm in assuming everything is suitably differentiable and that the requisite Lagrange multipliers exist. I use the multiplier r for the constraint (9.3), and a multiplier s for the constraint

$$\int u_z f \mathrm{d}y + \int u f_z \mathrm{d}y = 0 \tag{9.4}$$

which arises from utility maximization by the farmers. Notice that we have a second-order condition from that maximization,

$$A \equiv \int u_{zz} f \mathrm{d}y + 2 \int u_z f_z \mathrm{d}y + \int u f_{zz} \mathrm{d}y \leq 0. \tag{9.5}$$

[5] I first came across problems of this kind in connection with population policy. That analysis is given in ref. [5].

The first-order conditions for the government's maximization are that, for each y,

$$u_x f - rf + su_x f_z + su_{xz} f = 0, \tag{9.6}$$

from variations of c, and

$$r \int (y - c(y)) f_z \mathrm{d}y + sA = 0, \tag{9.7}$$

from variation of z. Notice that this last condition is simplified by use of eq. (9.4). It is more illuminating to write eq. (9.6) in the form

$$(r - su_{xz})/u_x = 1 + s(f_z/f). \tag{9.8}$$

A 'first-best' optimization would of course have made u_x the same for all y. It is to be expected that if, as in the application I have in mind here, larger z is to be encouraged, c will be so chosen that u_x diminishes as y increases. To show this, it is convenient first to impose the convention that

$$\int y f(y, z) \mathrm{d}y = z \tag{9.9}$$

for all z, and natural to suppose that

f_z/f is an increasing function of y, negative for small y and positive for large y. (9.10)

I must also assume that $u_{xz} = 0$.

It follows from (9.10) that r must be positive, since u_x always is, and the right hand side of eq. (9.8) is sometimes positive. We have to show that s is positive. Write $h(y) = f_z/f$, with the optimum z. Then eq. (9.8), which now reads $r/u_x = 1 + sh(y)$, tells us that x is a function of $sh(y)$ and y, $x = g(sh(y), y)$, where the derivatives of g are

$$g_1 = -\frac{1}{r}\frac{u_x^2}{u_{xx}} \quad \text{and} \quad g_2 = -\frac{u_{xy}}{u_{xx}}. \tag{9.11}$$

Assuming that u is concave, $g_1 > 0$. With this notation, we can write eq. (9.7) in the form

$$r = -sA + r \int g(sh(y), y) h(y) f \mathrm{d}y. \tag{9.12}$$

Here we have used the fact that $\int y f_z \mathrm{d}y = 1$, which follows from eq. (9.9) by differentiation.

Suppose that

$$u_{xy} \leq 0. \tag{9.13}$$

I shall show that this assumption implies that s is positive. If we had $s \leq 0$, $g(sh(y), y)$ would be a nonincreasing function of y, by eqs. (9.10), (9.11) and (9.13), and, y_1 being such that $h(y_1) = 0$,

$$\int g(sh(y), y)h(y)f\mathrm{d}y = \int [g(sh(y), y) - g(0, y_1)]h(y)f\mathrm{d}y \leq 0.$$

The first step is implied by $\int hf\mathrm{d}y = \int f_z\mathrm{d}y = 0$, since $\int f\mathrm{d}y = 1$; and the second is implied by eq. (9.10). Also, we know from eq. (9.6) that $A \leq 0$. Therefore, under assumption (9.13), the right-hand side of eq. (9.12) is non-positive if s is non-positive: but that is impossible, since $r > 0$. I have proved, then, that when x and y are weakly complementary, in the sense that (9.13) holds, and x and z are independent, $s > 0$; and, by eq. (9.8), u_x is a decreasing function of y. In the special case $u_{xy} = 0$ (which might hold, for example, because the farmer does not care about output itself, and needs no labor to gather it in) we can further conclude that x is an increasing function of y.[6]

The assumptions used to obtain these results are rather strongly sufficient, but it will be clear from the analysis that the 'perverse' case, where x decreases with y, is not entirely impossible.

It can be seen from formula (9.8) that the government's optimal policy c is not generally linear, or even particularly simple. To extend the argument, one would expect, administrative and political reasons apart, to recommend quite complicated allocation rules for medical care, police protection, car insurance and educational expenditures. One curious feature of rule (9.8) is highlighted by its response to the apparently sensible assumption that agricultural output is distributed lognormally:

$$f(y, z) = \frac{C}{y}\exp\left[-\left(\log\frac{y}{z} + 1/2\sigma^2\right)^2\Big/(2\sigma^2)\right] \tag{9.14}$$

(C is a constant). Eq. (9.14) implies that

$$f_z/f = \left(\log\frac{y}{z} + 1/2\sigma^2\right)\Big/(\sigma^2 z). \tag{9.15}$$

According to eq. (9.15), f_z/f tends to $-\infty$ when y tends to 0. But that is, by eq. (9.8), inconsistent with any value of s other than 0. Yet $s = 0$ does

[6] One might also think that in the optimum one would have $c'(y) < 1$; or at least that $y - c(y)$ would change sign only once, from negative to positive, as y increased. I have not found any nice assumptions that I can prove imply these results; and I suspect that they cannot easily be guaranteed.

not give an optimal policy. This is obvious because (depending on the utility function) a rule for distributing the available output that leaves everyone with the same u_x may leave no incentive for the farmers to produce anything! From a technical point of view, for this apparently well-set problem, no optimum exists.

The fact is, in this case, that one does better the more one penalizes those farmers who turn out to have very low output. The point can be made rigorously by considering what happens if all farmers whose output is less than a small number η receive consumption ε (another small number), while the others receive what they would receive in the *first-best* optimum – call it $c^*(y)$. This will be possible if it is possible at the same time to induce farmers to adopt the first-best optimum level of z – call it z^*. For that we require

$$\int_0^{\eta} u(\varepsilon, y, z^*) f_z(y, z^*) \mathrm{d}y + \int_{\eta}^{\infty} u(c^*(y), y, z^*) f_z(y, z^*) \mathrm{d}y$$
$$+ \int_0^{\eta} u_z(\varepsilon, y, z^*) f(y, z^*) \mathrm{d}y + \int_{\eta}^{\infty} u_z(c^*(y), y, z^*) f(y, z^*) \mathrm{d}y = 0.$$

Recollect that u_z is independent of x, so that the third and fourth terms can be written, together, as $\int_0^{\infty} u_z(y, z^*) f(y, z^*) \mathrm{d}y$. Thus we have to choose ε and η so that

$$\int_0^{\eta} \{u(c^*(y), y, z^*) - u(\varepsilon, y, z^*)\} f_z(y, z^*) \mathrm{d}y$$
$$= \int_0^{\infty} \{u(c^*(y), y, z^*) f_z(y, z^*) + u_z(y, z^*) f(y, z^*)\} \mathrm{d}y.$$

The right-hand side of this equation is equal to

$$r^* \int_0^{\infty} \{c^*(y) - y\} f_z(y, z^*) \mathrm{d}y,$$

where $r^* = u_x(c^*(y), y, z^*)$, by first-best optimality; and this expression will normally be negative, since

$$\int_0^{\infty} (c^* - y) f_z \mathrm{d}y = \int_0^{\infty} (c^* - y) h(y) f \mathrm{d}y < 0 \qquad \text{if } c^{*\prime}(y) < 1$$

by assumption (9.10), and this is ensured by the weak assumption that $u_{xy} \leq -u_{xx}$.

Thus, for any $\eta > 0$, we can choose ε so as to get z^* chosen. Yet, given any number M, we can choose η so small that $f_z < -Mf(y < \eta)$, so that

$$\int_0^\eta u(c^*(y), y, z^*)f(y, z^*)\mathrm{d}y - \int_0^\eta u(\varepsilon, y, z^*)f(y, z^*)\mathrm{d}y <$$

$$-\frac{1}{M}\int_0^\eta \{u(c^*(y), y, z^*) - u(\varepsilon, y, z^*)\}f_z(y, z^*)\mathrm{d}y =$$

$$-\frac{1}{M}\int_0^\infty \{u(c^*(y), y, z^*)f_z(y, z^*) + u_z(y, z^*)f(y, z^*)\}\mathrm{d}y \to 0 \text{ as } M \to \infty.$$

In this way, we can approximate as closely as we wish to the first-best optimum, by imposing penalties (presumably of great severity) on a small proportion of the population.

Although these farmers suffer severely, there are so few of them that their sufferings are outweighed by the encouragement their fate, or rather the prospect of it, gives to farmers taking production decisions. It seems that models of this kind can in certain cases provide some justification for extreme punishment of negligibly small groups.

The problem has been presented as one of government policy, but with the coincidence between private and government ends postulated, the solution may instead be interpreted as a prediction of the kind of insurance system that would arise in the society considered. It is interesting and important to consider further solutions in which the government adopts criteria different from those of the farmers *ex ante*, on utilitarian or egalitarian grounds; but I do not consider this further in the present notes.

9.3. The Characterization of Households

We may think of the standard problem of welfare economics in the following form. A household of type h has utility function $u(x, h)$ in terms of its trades x with the rest of the economy (i.e. excess demands). Production constrains $y = \Sigma_h x(h)$ to lie in the production set Y. $W(u)$ is to be maximized. The 'fundamental theorem' asserts that the optimum is a competitive equilibrium if there is a suitable distribution of budgets defined by a function $b(h)$. The point I want to emphasize, obvious though it is, is that the consumers are then supposed to choose what trade they will do rationally in terms of their own self-interest, but are supposed to reveal the necessary information about themselves, symbolized by the variable h, without regard to their own self-interest. The usual notation obscures this. The following example, which has some correspondence with reality, may highlight the difficulty[7].

Consider an economy with one consumer good, produced with labor. A household of type n provides labor of quantity ny when it works for time y. Every household has utility function $u(x, y)$ in terms of consumption x and labor time y. The welfare function is completely separable in terms of individual consumption, so that the utility function can be chosen to ensure that the welfare function is $W = \int u(x(n), y(n)) f(n) dn$, f being the density function giving the distribution of ability n. Welfare is quasi-concave in individuals' consumption, since u is concave. The optimum policy allocates consumption and time as functions of n in such a way that $\int x(n) f(n) dn$ is producible with labour input $\int ny(n) f(n) dn$. Denote the optimum by $x^*(n), y^*(n)$.

PROPOSITION. $u^*(n) = u(x^*(n), y^*(n))$ *is a* decreasing *function of* n *if (and only if) time is a strictly normal commodity.*

Recollect that to say time is a strictly normal commodity for the consumer means, by definition, that an increase in non-labor income would, in a market economy, lead the consumer to reduce his labor supply. It is a very plausible assumption.

The proof of the proposition, which is routine, is given in the appendix. The point of the proposition is that one would naturally assume that individuals have some control over the information they convey to government about their abilities; it is presumably easier to pretend to less ability than one has than it is to pretend to more. In any case there is no incentive to provide the information that the government must have if it is to bring the optimum about; on the contrary, there is an incentive for any individual not to provide the information. The model has some unrealistic features which serve to overstate the difficulties; but it is plausible that a government which attempted to realize the optimum of basic welfare economics would fail because of these difficulties.

---

7 It generalizes a special case mentioned in ref. [4].

In order to capture this feature of welfare economics, I propose to reformulate the basic problem in the form of two-level maximization, with households maximizing under a government-imposed constraint, and the government choosing the constraint in order to maximize welfare. Let us then interpret the vector x not only as trades but, more generally, as behavior. Thus, if some kinds of work can be observed directly, as amount of time or energy spent as well as through productive effects, both aspects could appear in the list x. Denote by k those aspects of individuals that are publicly known, independently of behavior (age, sex, place of birth, ...), and by h those aspects which, although they affect behavior through the consumer's choices, are not 'visible', at least to government. The individual will choose x so as to maximize $U(x, h, k)$, but will be constrained, first by his consumption set $X(h, k)$, and secondly by the constraints imposed by government and the markets of the economy, $A(k)$.

PROOF. $x(h, k)$ maximizes $U(x, h, k)$ subject to $x \in X(h, k) \cap A(k)$. (9.16)

The production constraint is that

$$\Sigma x(h, k) \in Y. \tag{9.17}$$

(Some of the components of Σx, corresponding to non-trade behavior, are redundant.) Write $u(\cdot)$ for the function of h and k defined by $u(h, k) = U(x(h, k), h, k)$. Then the government seeks to maximize $W(u(\cdot))$, subject to the constraints (9.16) and (9.17), by choice of $A(\cdot)$.

The government is supposed to know the nature of the population, the number of people for each h and k. Specifically, it is most interesting to suppose h and k continuously distributed with density function $f(h, k)$; so that the production constraint should be written

$$\int x(h, k) f(h, k) \mathrm{d}h \, \mathrm{d}k \in Y. \tag{9.17'}$$

The government's maximand would depend both upon the utility outcome $u(\cdot)$ and upon the distribution $f(\cdot)$. For example, it might take the completely separable form – which is the easiest to handle –

$$W = \int u(h, k) f(h, k) \mathrm{d}h \, \mathrm{d}k.$$

Two objections to this formulation of welfare economics will so readily occur to the reader that I must answer them now. The first is that the government's knowledge of $f(\cdot)$ is hard to reconcile with its ignorance of any particular man's h. To this I answer that $f(\cdot)$ embodies *statistical* information, as opposed to personal information, which could be collected under the secrecy normal for census and sample. There is no incentive to hide information that will be used in this non-personal way. Of course the information obtained by census or sample may be quite poor – but $f(\cdot)$ represents the government's beliefs, even if they are inaccurate. The only correspondence between $f(\cdot)$ and reality that is required for application of the theory is that, in the outcome, markets clear: if they did not, the government would have to change its beliefs. Finally, I note that it will, in general, be possible to deduce the distribution of h and k from observation of the distribution excess demands x, as they actually appear (given the model of consumers expressed by $U(x, h, k)$).

The second objection, to which the problem as posed is open, is that it is too hard to answer usefully. This objection can be answered only by a complicated analysis, for which this is not the place. A special case has been analyzed in ref. [4], and in unpublished work; and I can derive necessary conditions for optimality in reasonably simple form. The power of the theory is illustrated by a result quoted in section 9.3.2.3. In some cases, this problem may have a simple solution.

9.3.1. Example

Consider the special case to which the above pro sition applies. Denote labor supply by z. Then behavior is (x, y, z) and is constrained by $z \leq ny$. This assumes unproductive work is possible. The government does not know n. In terms of the model, maximization can be achieved (subject to the usual convexity assumptions, etc.) by distributing income according to observed ability z/y in such a way that everyone receives the same utility; specifically by assigning the same income to everyone, whatever his z/y. But this assumes that, by some means or other, individuals are induced to make z as large as possible, for given y, when it matters to them not at all. Otherwise, when this inducement is absent, there is, strictly speaking, no optimum; the government can always do better by making the utility distribution more equal. With that proviso, we have an example in which the optimum, though not the first-best optimum, is a competitive equilibrium.

Generally speaking, with u depending on h, the optimum for the above problem is not a competitive equilibrium (which would have A described by $p \cdot x \leq b(h', k)$, where h' is an estimate of h deduced from x by means of the consumption set $X(\cdot, k)$, and p are the correct producer prices), because A can be nonlinear. (I have discussed the above example, with y invisible, in ref. [4].) One really general result is the desirability of production efficiency [3]. Under very general assumptions, optimum production for the above problem lies on the frontier of the production set; thus decentralization of production decisions may be possible. The argument in support of this conclusion is simple: if production were in the interior of the production set, a sufficiently small subsidy (an equal lump-sum payment to everyone) would make everyone better off and, presumably, change aggregate demands by an amount small enough to leave them in the production set.

9.3.2. Some points

9.3.2.1. Uncertainty. The model as it stands allows for misleading information, but not for imperfect information. In general, we expect the government to have imperfect information about consumer behavior and about the consumer characteristics labelled k. Suppose first that all consumer characteristics are visible, so that the variable h does not appear in the formulation. The government observes k'. k and k' have a joint probability density function $f(k, k')$. (This expresses the imperfect information about k that observation of k' provides.) The government imposes on consumers a budget constraint of the form $x \in A(k')$, but the individual knows k and therefore maximizes $U(x, k)$ subject to $x \in X(k) \cap A(k')$.

The point I want to make is that this problem has essentially the same features as the problem set in eqs. (9.16) and (9.17) above. The optimum is not (in general) a competitive equilibrium, except on the production side. The optimum budget constraint will be nonlinear. To see this, consider what the optimum must be. Optimum production y^* is equal to $\int y^*_{k'} f(k, k') \mathrm{d}k$, where $y^*_{k'}$ is the aggregate production made available to all consumers who appear to have characteristic k'. Consider one particular value of k'. Then, given the optimal allocations to everyone else, $A(k')$ must have been chosen so as to solve a problem of the form of eqs. (9.16) and (9.17) with $y^*_{k'}$ given.

A special case may make the point clearer. Consider the example discussed before, with the additional assumption that the government, unable to observe y, makes inaccurate observations n of n'. The joint density function for n and n' is $f(n, n')$. The government wants to maximize

$$\int u(x(n, n'), y(n, n'))f(n, n')\mathrm{d}n\, \mathrm{d}n', \tag{9.18}$$

where the budget constraint may be taken to have the form, for an (n, n')-man, $x \leq c(ny, n')$. since ny is his supply of labor, and n' is information the government possesses about him. Then for each (or, I suppose, for almost all) n', the government should choose the optimum income tax schedule (leading to the constraining consumption function) given the part of production allocated to those who are labelled n', to maximize $\int uf\mathrm{d}n$. Naturally, in income tax theory, the optimum schedule is related to the shadow prices of the commodities – in this example to the marginal productivity of labor – and these must be the same for all n'. But that does not ensure that the various budget constraints have a similar form, for the form of the optimum depends upon the distribution of skills in the population, i.e. on $f(\cdot, n')$ for each n', and also on the production made available to the class in question.

Thus imperfect information leads to essentially the same economic considerations as misinformation. By the same argument, a satisfactory theory for the problem of eqs. (9.16) and (9.17) could be very easily extended to the case of imperfect information.

9.3.2.2. Rationing versus the price system. The particular form of the general welfare economics problem stated above in which the budget constraint is linear in x (i.e. A is a convex cone) has been studied in ref. [3] (where the variable k is not considered explicitly). There it is stated that when optimum commodity taxes prevail – i.e. when A is optimal subject to being linear – it may actually be desirable to introduce rationing for some commodities, thus replacing the market. Rationing should be understood in this context as a constraint imposing maximum levels of 'trade', which are functions only of k, while leaving the rest of the budget constraint unaffected by the amounts of the rationed commodities actually traded. (There may be a problem of ensuring that the ration quantity allows a consumption plan in the consumption set. For instance, in the special example we have been discussing, it is not possible to impose a ration on z if all abilities down to zero are represented.) Many social services have the characteristics of rationing.

It is an interesting question how far rationing schemes are desirable.

The problem is not susceptible to general theorems, but is a question of what is plausible, given what we know of the real world. Some of the main reasons for rationing are omitted from the welfare economics problem we are considering. Within the context of the model, it is easy to see that universal rationing is generally far from optimal. Universal rationing means that every consumer supplies the same and receives the same from the market. In the case of a convex aggregate production set almost everyone will be better off if he faces a budget constraint set $(1/H)Y$, where H is the number of consumers, and the resulting demands are feasible. In general, this will not be the optimal A, but it is better, and usually far better, than the rationing proposal. It seems likely that one can similarly construct a fairly liberal constraint set that is better than rationing of a single commodity, but I do not yet see how to do it.

In this context, it should be noted that some kinds of 'rationing' may be obtained by social arrangements not accessible to government. Many aspects of behavior, above all care and efficiency at work, are not directly visible to government but, to some extent, visible to other individuals. Social norms may make behavior in these respects more nearly uniform, and perhaps socially more useful, than it would otherwise be. Uniform behavior by work groups might be regarded as regrettable in a world with perfect distribution, but may actually be desirable, though an infringement of individual preferences, in a world of imperfect information.

9.3.2.3. Uniform prices versus progression and discrimination. At the opposite extreme to rationing we may consider tax systems with uniform nondiscriminatory consumer prices. If the optimum had this character, no undesirable trading among households would take place. In general, there is no reason why uniform prices should be optimal: it is then desirable that trading among households be prevented if possible, since it would change the constraint set actually operating. (Of course it might change the constraint set differently for different groups, and in that case it *might* be desirable.) I suggest two interesting questions: first, when is a policy of uniform subsidies supplemented by commodity taxes approximately optimal? Second, when will uniform commodity taxes be part of the optimal system?

In connection with the second question, the following particular case may be of interest. (I believe Stiglitz has obtained results of the same kind.) Generalize the particular example above to many consumer goods, assuming that the typical household has utility function

$$u(v(x), z/h), \tag{9.19}$$

where x is now a vector of consumer goods. We may take it that the optimal budget constraint, defining A, will have the form $b(x) \leq z$.

I can prove that in fact the optimum can be obtained by a budget constraint of the form

$$\beta(p \cdot x) \leq z, \tag{9.20}$$

where p is the vector of producer prices (in terms of which the optimum aggregate $\int x(h) f(h) \mathrm{d}h$ maximizes profits within the production set). In effect this enables the economy to provide the desired levels of v at minimum cost. The optimum is achieved by having a tax, generally non-linear, on labor income, and no taxes on other commodities. Considerable generalization of this result is possible.

9.4. The Consumer's Information

It may be conjectured that it would usually be desirable in terms of the problem presented in the previous section, that consumers be as well informed as possible about their own characteristics, measured by the variables h. To take the extreme case, if no-one knew anything to differentiate himself from anyone else, we should have the 'rationing' solution, and we have seen that that can be improved upon. But this is surely misleading. It seems that many people get satisfaction from beliefs about their relative intelligence, strength, beauty, charm or quality of judgment. The requirements of the economy apart, it might be a pity if their information on these subjects were very accurate. Perhaps some of us are better for a precise knowledge of our failings and abilities, but a meritocracy in which everyone knew his ability beyond doubt is not, I think, an attractive prospect.

Therefore, even if it were possible accurately to ascertain each person's abilities and other characteristics, it would not, I suggest, be desirable to do so – although it may well be desirable to obtain statistical information of this type from small samples. If a man believes he receives a low labor income because he chooses not to work very hard, rather than because he lacks the required abilities, there is a case for leaving things that way, rather than attempting to assess his abilities, and subsidize him accordingly. This is a reason for welcoming uncertainty and for going beyond the Arrow–Debreu version of welfare economics under uncertainty.

References

[1] Arrow, K. J. *Essays in the Theory of Risk-Bearing*. North-Holland: Amsterdam (1971).

[2] Diamond, P. A. Cardinal welfare, individualistic ethics and interpersonal comparisons of utility: comment. *Journal of Public Economics* (October 1967).

[3] Diamond, P. A. and Mirrlees, J. A. Optimum taxation and public production. *American Economic Review* (March and June 1971).

[4] Mirrlees, J. A. An exploration in the theory of income taxation. *Review of Economic Studies* (April 1971).

[5] Mirrlees, J. A. Population policy and the taxation of family size. *Journal of Public Economics*, **1** (1972), 169–198.

APPENDIX

The proposition on p. 251 is to be proved. In the optimum, we must have, for shadow prices p, q independent of n,

$$u_x = p \qquad u_y = -qn.$$

Differentiating with respect to n, we obtain

$$u_{xx}x' + u_{xy}y' = 0, \qquad u_{xy}x' + u_{yy}y' = -q,$$

where primes denote derivatives with respect to n. Therefore

$$\frac{\mathrm{d}u}{\mathrm{d}n} = u_x x' + u_y y' = \frac{(u_x u_{xy} - u_y u_{xx})q}{u_{xx}u_{yy} - u_{xy}^2}$$

which has the same sign as $u_{xx} - (u_x/u_y)u_{xy}$. This last expression is negative if and only if y is a normal commodity.

COMMENTS

Notes on welfare economics, information and uncertainty

Peter Diamond

Economists have not yet learned to incorporate an interesting theory of information into general models of resource allocation. I suspect that the difficulty arises from the basically dynamic nature of the former and the essentially static nature of most of the latter. Given this situation, Professor Mirrlees has followed the standard approach of treating information (or policy tools) as either costlessly available or totally unusable. This approach permits progress (to which he has made major contributions) in the understanding of resource allocation where the fundamental welfare theorem of first-best economics is not applicable. Since the problems at hand are so difficult – mathematically complicated and not amenable to the intuitions we have developed so far – it seems good research strategy to proceed on this basis without attempting to incorporate information modelling into these problems too (and conversely to examine information in models where resource allocation questions are very simple).

Unfortunately, the models used in these notes are mathematically very complicated. But the difficulty lies with the problems not the author since the model he has used to start this paper is the simplest possible of general models – one can't consider resource allocation without a choice variable; one can't have uncertainty without a random variable; and one can't have public policy without an observable variable (which can't coincide with either of the other two variables if the problem is not to vanish). Add the assumption that everyone is identical to the three good model and we have the simplest possible situation, short of going to specific functional forms. In this setup, the model focuses on the desirability in general of nonlinear policy tools to deal with the moral hazard problem – the problem arising from individual maximization given the alternatives made available by public policy which result in individuals facing shadow prices which differ from social shadow prices. (This is the same problem as the deadweight burden of taxation – if individuals determined demand by prices which ignored taxes, the excise tax world would coincide with the lump sum tax world.)

Moral hazard comes in three forms. Individuals may falsify reports (report nonexistent accidents or incorrect damage estimates). For my division of types this represents an attempt to alter distributions but not a resource using decision. Resource allocation decisions can be distinguished as occurring before or after the outcome of the random event is known. The most familiar example of the latter arises in medical care where people presumably purchase excessive services, once sick, since they bear only a fraction or none of the costs. An example of the former, which is the focus of Mirrlees' analysis, would be fire insurance where insufficient protection is taken *ex ante* to limit either the probability or extent of damage of fire. Presumably the natures of the information difficulties in the two cases are different (although possibly not significantly so) in that everyone is making (socially) poor resource allocation decisions in the second case, but only those who have accidents (are sick) are doing so in the first. If one were to try to improve decisions within the framework of the information structure, one might try subsidizing fire extinguishers (or fertilizers) in the first case. Taxing medical care, however, would defeat the purpose of the insurance with the second case.

The second conventional insurance problem is adverse selection. Because of an inability to distinguish among individuals for whom the cost of service varies, the insurer charges a uniform price – an average cost price rather than a marginal cost price. When individuals know more about themselves than the company does, good risks don't buy insurance (or buy little) while bad risks buy (or buy much). (Mirrlees also mentions the reverse case where individuals know less about their true accident probabilities than the government or the insurance company.) In terms of trading commodities of uncertain quality the problem has been nicely characterized by George Akerlof as a 'market for lemons' (*Quarterly Journal of Economics*, **84** (1970), 488–500). Mirrlees approaches this problem by means of a double index of observed and unobserved traits. Since he is considering social policy, individuals are not allowed to opt out, although the inefficiency from average cost pricing remains present, but less starkly so than in the Akerlof case. (Again the problem has an analogue in tax theory in consideration of taxes to correct externalities where the same tax is levied on everyone although different people give rise to different degrees of externality (for example, fast *versus* slow drivers). As with the Akerlof example, this can take the extreme form of making it impossible to have any corrective value in taxation.)

Since Mirrlees is focusing on a single activity or single decisions he does not mention the problem which has been called 'externalization by transfer' by Guido Calabresi (*The Costs of Accidents*, Yale (1970)). Here many activities get lumped together in a single insurance policy for administrative reasons. Thus the price to the individual does not vary with his choice of activities. (A similar problem arises with prices that depend on the choice of activity (like car owning) but not the level of activity (miles driven). Thus summarizing these three problems the shadow prices may be wrong for an individual choosing his level of activity because of the needs of insurance, the grouping of the individual with others, or of the activity with others.)

The information problem is further complicated by the passage of time and the presence of many individuals who may have relevant information. Since the past is given, taxing it will not alter past behavior. However, present behavior will be altered when it is anticipated or feared that future taxation will depend on (then) past behavior. This relationship does not seriously allow the past to be a base for lump sum taxation, but the complexities of time and individual rigidities also may not be well caught in a static equilibrium model which has instantaneous response to tax policy. Mirrlees mentions social arrangements which may take advantage of different sources of knowledge. An additional example he does not cite is private law suits serving as a deterrent to externality generating behavior, which is frequently dubbed a 'private attorney general'. This indicates a role for social policy to encourage good social use of private information.

The problems of welfare economics in the presence of complexities of information and uncertainty are great. These arise both in the difficulty in analyzing individual models and in the absence, so far, of an overall framework in which to locate different models. Mirrlees' notes are very valuable as approaches to dealing with a number of these problems. The next instalment is eagerly awaited.

CHAPTER 10

PREEXISTING CONTRACTS AND TEMPORARY GENERAL EQUILIBRIUM*

Jerry Green

10.1. Introduction

The object of this study is to consider a model of general economic equilibrium over time in which the markets for trading commodities are open at every date. This type of system is to be contrasted with that of the Arrow–Debreu model in which an equilibrium is reached at an initial point in time and the remainder of economic activity consists of carrying out the equilibrium plans formulated at that date. Two distinct approaches have been taken to the problem of modelling systems of this type. The first, introduced by Radner [13] and used subsequently by Hahn [6] and Kurz [10] in studies of transactions costs, is to ask whether there exists a set of prices on current markets and a sequence of point forecasts regarding prices on markets in the future such that, if these beliefs were held by everyone, the resulting course of the economy would see them fulfilled.

The second approach is that of viewing the economic system as a sequence of temporary equilibria. This idea goes back at least to Hicks [7]. It was introduced in a formal general equilibrium model of a monetary economy without futures markets by Grandmont [5]. Subsequently, Sondermann [4] treated a monetary economy with production in this framework and Green [3] studied a non-monetary economy with futures markets. In the temporary equilibrium approach, expectations are taken as data of the system, although they may depend on

* In the early stages of this work I benefited greatly from discussions with Jean-Michel Grandmont and Frank Hahn. A substantial debt is owed to the writings of Grandmont which provided many useful insights. Financial support under National Science Foundation Grant GS-31688 to Harvard University is gratefully acknowledged.

variables currently determined in an equilibrium. It is not required that individuals agree, though some degree of compatibility of expectations appears to be necessary to ensure the existence of an equilibrium (see ref. [3]).

The present paper is a continuation of ref. [3] in the following sense: if a temporary equilibrium in a previous period resulted in the exchange of futures contracts, then it is necessary to take into account the fact that the current period endowments will reflect these trades. This is not a problem if endowments, including preexisting contracts, remain in the consumption set. However, a sale of futures contracts in a previous equilibrium in excess of the endowment of that commodity may lead to the nonexistence of an equilibrium in the present period. In particular, the possibility of bankruptcy exists. This must be faced squarely if there is to be any hope that the economy can be modeled as a sequence of temporary equilibria when futures contracts are permitted.

10.2. Previous Work on Preexisting Contracts

There have been several previous attempts to incorporate preexisting contracts in a model of general equilibrium. In Arrow and Hahn [1] a single-period model is considered in which the endowments might contain negative amounts of some commodities due to debts incurred by the household in earlier periods[1]. They prove that there exists a price system and a set of transfers (in units of account) among the agents such that all markets clear.

A second treatment of this topic is in the thesis of Grandmont [4]. Grandmont's primary interest is in an economy with money and in which money, or debt measured in units of account, is the only store of value. His model has two periods. In the first, endowments are non-negative and trade takes place only on spot markets. Debt, in units of account, can also be incurred, or claims to future units of account can be acquired. The price of obtaining one unit of account in the second period is determined in the equilibrium of the first period in the market

[1] Arrow and Hahn actually are more general on this point since they deal with a general consumption set instead of the non-negative orthant of the commodity space. Simplifying their model in this way will, however, facilitate comparison with what we shall do later.

for 'debts'. If an individual debtor is not solvent in period 2, because the value of his endowment is less than the amount of his debt, he is assigned the consumption of zero and his endowment is used to pay the debt to whatever extent this is possible. Naturally, if there is bankruptcy in the economy, the amount of debt repayments will be insufficient to cover the quantity owed. The repayments are made proportionately for all creditors. That is, every creditor is refunded the same proportion of what he is owed. The individuals in the economy know this rule in period 1, when they consider becoming creditors. They view loaning units of account as a risky prospect in the sense that they have subjective probability assessments over the fraction of debt that will be refunded.

Grandmont shows that there will be an equilibrium in each period: period 1, when endowments are non-negative and future bankruptcies are viewed with uncertainty; and period 2, when bankruptcies occur, debts are settled and economic activity ends.

The first introduction of the potential for bankruptcy in a general equilibrium model is due to Stigum [16], though his conditions for existence are not immediately comparable with those of the other authors.

Our model differs from each of these in several respects. As mentioned in the introduction, the primary objective is to be able to view the economic system as progressing from one temporary equilibrium to the next. Thus we will be considering a period in which there has been a past – preexisting contracts may be present – and there will be a future – plans are made viewing the future with subjective uncertainty. This differs from the Arrow–Hahn case in which there has been a past, but there will be no markets convening in the future. In the Grandmont model, period 1 has a future but no past, period 2 a past but no future. We shall attempt to synthesize these, so that the period in question can be viewed as a typical 'snapshot' of the economy.

One consequence of there being a future period is that individuals view their own potential bankruptcies, as well as those of others, with subjective uncertainty. If there is a past, the value of holding debts of others must be determined. Thus the two sides of the bankruptcy problem – anticipation and settlement – are simultaneously present in our system.

Our model has further differences with the Grandmont system because contracts sold include all future commodities as well as those currently deliverable. Debt in our model will be represented by the selling of a futures contract for real goods instead of a claim to a future unit of account. Thus we will not be able to determine a price level as Grandmont does, but rather relative prices only will emerge. The system we propose, however, has many stores of value – every futures contract could serve as such. Typically, an individual may sell some contracts and buy others so that he can be classified as neither debtor nor creditor. The net value of his preexisting contracts is determined in the equilibrium. This gives rise to the possibility that the bankruptcy of one agent may cause others to be bankrupt if they are his creditors but in debt to others. This phenomenon cannot occur in thc Grandmont system since all creditors will be solvent because their endowments are non-negative, even if no debtors repay any debts.

10.3. Codes of Conduct and Non-Economic Penalties in General Equilibrium Models

The Arrow–Debreu model presumes an institutional structure in which every agent has access to a system of markets that he can use to trade commodities. Since some of the commodities are not deliverable until a date beyond that at which trades are made, one typically speaks of these markets as involving futures contracts. An individual who buys such a contract believes that it will, in fact, be fulfilled: that is, that when the appointed date arrives, the exact amount of the agreed upon physical commodity will come into his possession. On the other side of the coin, therefore, the Arrow–Debreu model presupposes honesty: no one is allowed to sell a contract without having the required amount of the good available in his endowment (or production plan).

Thus the Arrow–Debreu model embodies an ethical code of conduct and a set of beliefs about the performance of the system, as well as an institutional framework. It should be noted that these are consistent in the sense that if everyone follows this code of conduct, the belief that all contracts will be fulfilled is justified. Further, the belief that the markets will only meet once is validated by the fact that even if it were possible to reopen them, no one would want to engage in trade at the equilibrium prices that would emerge.

However, certain generalizations of the Arrow–Debreu model may destroy this self-realization property. Knowledge that markets will reconvene will cause people to behave initially in such a way that they will desire to trade at future dates as well. In ref. [3] the existence problem was studied for such a system.

The result of an equilibrium in this model is a set of realized consumptions of current goods and an altered endowment distribution for the following period. In the following period endowments will be modified to reflect futures contracts traded. Thus this model tacitly assumes the Arrow–Debreu code of conduct and resulting beliefs about futures contracts. But, unless the endowments remain non-negative, these beliefs will prove unjustified. There may be no equilibrium because there are prices at which some individuals have endowments with negative value and therefore cannot choose any point in their consumption set (assumed to be the non-negative orthant). That is, the possibility of bankruptcy exists in this system even though every individual makes only those futures contracts that he can fulfill with subjective probability one. Some people hold subjective beliefs that are 'wrong' in the sense that the prices next period would have been deemed 'impossible' *ex ante*[2].

The problem is that the institutional structure of sequential trading is incompatible (except under overly strong conditions on expectations) with the presumption that contracts are fulfilled with certainty. Thus a broader concept of the nature of economic contracts is required if we are to be able to view the system as generating a sequence of temporary equilibria in which the environment in each period is a result of past actions.

Since the possibility of bankruptcy cannot be avoided, a viable economic system must have rules governing this circumstance. The rule must be feasible to implement. That is, it must prescribe a feasible redistribution of commodities corresponding to any state of the economic system – including those states in which some bankruptcy is occurring.

Further, if one interprets consumption sets as describing consumption bundles that are the minimum required for sustenance of the individual, then some non-economic penalties for bankruptcy are required, for

---

2 Strictly speaking, the actual prices are such that a neighborhood of them lies outside the convex hull of the support of the subjective distribution previously held.

otherwise an individual would not care about the extent of his bankruptcy – all bankruptcies would leave him with a consumption on the lower boundary of his feasible set. Problems concerning the determinateness of individual behavior would arise.

Thus, it seems that a natural consequence of the sequential markets setting is a need for explicit consideration of social rules of conduct and a penalty structure which are unnecessary in the Arrow–Debreu system and are usually thought to be outside the scope of economic analysis.

At this point a crucial choice in model-building arises: in the real world, the vast majority of contracts are bilateral in that the buyer and seller know each other's identity. For the purposes of the Arrow–Debreu model this was unimportant because, since all contracts are always fulfilled and transaction costs are absent, the identity of one's trading partner is economically irrelevant. If, however, we contemplate constructing a system in which these properties may fail, the identity of one's trading partner will matter and contracts with different people must be viewed differently; that is, they may have different prices in an equilibrium and must be viewed as distinct commodities.

One possibility would be to take explicit account of this phenomenon, making assumptions about the information each agent has regarding the portfolios of others. In an equilibrium of such a system, the prices of currently deliverable commodities will not differ among sellers, but the prices of futures contracts will.

Proving the existence of an equilibrium – and even stating its definition – in such a system raises several interesting and atypical questions. We hope to treat this in a separate paper.

For the present model, we shall assume that all trades are made with an abstract market. The contract obliges its seller to deliver the good, if he is solvent. A buyer, however, knows that some sellers may be bankrupt. The contract entitles the buyer to a proportion of its face value equal to the ratio of actual repayments to the total quantity of outstanding contracts for this commodity. This actual payout ratio will vary from commodity to commodity since the sellers of different futures contracts are, in general, different.

One way of interpreting this system is that every buyer is assigned, by the market, contracts of various sellers in proportion to his share of the total contracts held. If buyers were risk averse and had no information about the asset positions of any sellers, then this is the allocation that they would choose in order to spread the risk evenly among themselves. Viewing our allocation rule in this way, it is apparent that this system (the use of an abstract market) is the polar opposite of the full information situation in which every buyer knows the complete asset position of every seller. However, since buyers' information is rarely perfect, and often very imperfect, this approach seems justified as a modeling technique by its symmetry and simplicity. It also seems to be a reasonable institutional rule because of its risk spreading properties under complete ignorance. We now proceed to a formal statement of the model and assumptions.

10.4. The Model

We will be studying the existence of a temporary equilibrium at a date that will be denoted by 1. There will be a future, denoted by period 2. Past economic activity enters the model through the date of period 1, as will be described shortly. There are ℓ_1 commodities at date 1 and ℓ_2 at date 2; we write $\ell \doteq \ell_1 + \ell_2$. Commodities are non-durable. A typical individual at date 1 is described as follows.

Naturally occurring endowment that accrues to him regardless of the actions of others and is known with no uncertainty is written

$$\boldsymbol{\omega} = (\omega^1, \omega^2) \in R^{\ell}_{+}.$$

It is further assumed that $\omega^1 > 0$, $\omega^2 > 0$ and the aggregate is strictly positive[3].

Preexisting contracts that reflect the economic activities of previous periods and are subject to default as we shall describe below are written

$$\boldsymbol{e} = (e^1, e^2) \in R^{\ell},$$

that is, in previous equilibria the individual may have traded contracts that are now due (e^1), or he may have made futures transactions involving periods still to come (e^2). Notice that no restriction is placed on the

[3] We adopt the following conventions for vector inequalities:

$\boldsymbol{x} \geq \boldsymbol{y}$ implies $x_i \geq y_i$ for all i;
$\boldsymbol{x} > \boldsymbol{y}$ implies $x_i \geq y_i$ for all i, and $x_i > y_i$ for some i;
$\boldsymbol{x} \gg \boldsymbol{y}$ implies $x_i > y_i$ for all i.

sign of $\boldsymbol{e}$. However, we will later use the fact that the sum of preexisting contracts over individuals must be zero.

Let us denote that $\boldsymbol{e} = \boldsymbol{e}_+ + \boldsymbol{e}_-$, where $\boldsymbol{e}_{+k} = \max(\boldsymbol{e}_k, 0)$ and $\boldsymbol{e}_{-k} = \min(\boldsymbol{e}_k, 0)$. Contracts held are $\boldsymbol{e}_+$ and contracts owed are $\boldsymbol{e}_-$. Some of the contracts held may be in default. The extent of default will be determined shortly. At present let us write $\boldsymbol{r} = (r^1, r^2) \in [0, 1]^\ell$ as the vector of *returns ratios on preexisting contracts*. This vector will be determined, along with prices, in the temporary equilibrium. If the individual is holding preexisting contracts $\boldsymbol{e}_+$, and the returns ratios are $\boldsymbol{r}$, he is actually paid

$$\boldsymbol{re}_+ \equiv (r_1 e_{+1}, \ldots, r_\ell e_{+\ell}).$$

Prices at date 1, to be determined in equilibrium, are denoted

$$\boldsymbol{p} = (p^1, p^2) \in \Delta^\ell, \text{ where } \Delta^\ell = \{\boldsymbol{p} \in R^\ell_+ \mid \sum_k p_k = 1\}.$$

If the individual faces prices–returns $(\boldsymbol{p}, \boldsymbol{r})$, his wealth is given by $\boldsymbol{p} \cdot (\omega + \boldsymbol{re}_+ + \boldsymbol{e}_-)$. We shall say that the individual is *solvent* or *bankrupt at* $(\boldsymbol{p}, \boldsymbol{r})$ according to whether or not this expression is non-negative.

If the individual is bankrupt, the *default rule* modifies his endowment in the following way: he is allowed to default an equal proportion of all the contracts on which he is a debtor. This *default proportion at prices* $\boldsymbol{p}$ *and returns* $\boldsymbol{r}$ is given by

$$d(\boldsymbol{p}, \boldsymbol{r}) = \min\{\delta \geq 0 \mid \boldsymbol{p} \cdot (\omega + \boldsymbol{re}_+ + (1 - \delta)\boldsymbol{e}_-) \geq 0\}.$$

This is the bankruptcy law to which we have referred in the last section. At each $(\boldsymbol{p}, \boldsymbol{r})$ the endowment of the individual is given by

$$\omega + \boldsymbol{\eta}(\boldsymbol{p}, \boldsymbol{r}) = \omega + \boldsymbol{re}_+ + (1 - d(\boldsymbol{p}, \boldsymbol{r}))\boldsymbol{e}_-,$$
$$(\boldsymbol{\eta}(\boldsymbol{p}, \boldsymbol{r}) = (\boldsymbol{\eta}^1(\boldsymbol{p}, \boldsymbol{r}), \boldsymbol{\eta}^2(\boldsymbol{p}, \boldsymbol{r})).$$

Thus the interpretation of $\boldsymbol{\eta}(\boldsymbol{p}, \boldsymbol{r})$ is the vector of preexisting contracts, after the bankruptcy law has been applied. By definition, the value of the altered endowment is non-negative.

We shall let the number of individuals in the economy be I and denote the set of all individuals by $S = \{1, \ldots, I\}$. We denote quantities referring to a particular individual by a pre-superscript i. In order for the default rule to be feasible when applied simultaneously to all individuals, we must have that the amount of contracts actually repaid equals the amount of outstanding obligations after the preexisting contracts have been adjusted. This can be expressed as

$$-\sum_{i\in S}(1 - {}^i d(\boldsymbol{p}, \boldsymbol{r}))^i \boldsymbol{e}_{-k} = r_k \sum_{i\in S} {}^i \boldsymbol{e}_{+k} \tag{10.1}$$

for all k.

If eq. (10.1) is satisfied, we will say that $\boldsymbol{r}$ is a *consistent* returns vector at prices $\boldsymbol{p}$.

It can easily be observed that condition (10.1) may fail at an arbitrary $(\boldsymbol{p}, \boldsymbol{r})$. The principle result of this section will be to show that for each strictly positive $p \in \Delta^{\ell}$, there is a *unique* consistent returns vector.

This result will be useful when we come to proving the existence of an equilibrium, for in searching for an equilibrium it will suffice to vary only $\boldsymbol{p}$, letting $\boldsymbol{r}$ be determined endogenously as the unique returns vector whose existence is asserted above. A second, technical, reason for being interested in this result is that we will eventually apply a fixed point theorem to the domain of the aggregate excess demand function. For this, we will use the fact that the domain can be approximated by a sequence of compact convex sets. If, in $\Delta^{\ell} \times [0, 1]^{\ell}$, the set of consistent $\boldsymbol{r}$ at each $\boldsymbol{p}$ had a pathological shape, the application of fixed point methods would be inappropriate. The uniqueness of the consistent returns ratio vector avoids this potential difficulty.

In the model we have been discussing, debts are centralized in the sense that everyone trades with an abstract market. However, it is clear that this system is equivalent to a bilateral trading system with debts appropriated in a particular way. Suppose that, on each contract, every creditor holds the same proportion of the contracts of each debtor. Thus, the proportion of his debt on contract k that debtor i owes to creditor i' is the share of total outstanding contracts on good k that i''s contracts represent.

Let $\boldsymbol{p} \gg 0$ be fixed. Let the debts owed by the various individuals, when appropriated in this way, be written as a matrix $T = [t_{ij}]$, where t_{ij} is the value at prices $\boldsymbol{p}$ of the claims of individual i on individual j. Let $t_{ii} \equiv 0$.

Let $\boldsymbol{W}$ be the vector of values of naturally occurring endowments of the individuals at these prices. Let $\tau_j = \sum_{i\in S} t_{ij}$; τ_j is the total value of the claims against individual j.

Let

$$E = T + \begin{bmatrix} -\tau_1 & & 0 \\ & \ddots & \\ 0 & & -\tau_I \end{bmatrix}.$$

Thus, if in the bilateral appropriation of the debt we have chosen, individuals are defaulting with proportions $\boldsymbol{d} = ({}^1d, \ldots, {}^Id)$, the vector of wealth post-default is given by $\boldsymbol{W} + E(\mathbf{1} - \boldsymbol{d})$, where $\mathbf{1}$ is the vector each element of which is one. For $\boldsymbol{d}$ to be an equilibrium default proportions vector for this way of appropriating debt, we must have

$$\left.\begin{array}{l} \boldsymbol{W} + E(\mathbf{1} - \boldsymbol{d}) \geq 0 \\ 0 \leq \boldsymbol{d} \leq \mathbf{1} \\ (\boldsymbol{W} + E(\mathbf{1} - \boldsymbol{d})) \cdot \boldsymbol{d} = 0 \end{array}\right\}. \tag{10.2}$$

The interpretation of the last equation is that any individual who is defaulting on some contracts (${}^id > 0$) must have zero wealth.

It is clear that this appropriation of debt gives rise to a system that is equivalent to the original one in the following sense: if $\boldsymbol{d}$ is an equilibrium of the bilateral trade system (i.e. satisfies eq. (10.2)) and prices are $\boldsymbol{p}$, then the returns ratios $\boldsymbol{r} = (r_1, \ldots, r_\ell)$ calculated from the equations

$$r_k = \frac{-\sum_{i \in S} (1 - {}^id){}^ie_{-k}}{\sum_{i \in S} {}^ie_{+k}} \qquad k = 1, \ldots \ell$$

are consistent at $\boldsymbol{p}$. Conversely, if $\boldsymbol{r}$ is consistent at $\boldsymbol{p}$, then the default proportions associated with $(\boldsymbol{p}, \boldsymbol{r})$, ${}^id = {}^id(\boldsymbol{p}, \boldsymbol{r})$, satisfy (10.2).

We will use this equivalence in proving the existence and uniqueness of a consistent $\boldsymbol{r}$ at each $\boldsymbol{p} \gg 0$.

LEMMA 10.1. *There exists a consistent* $\boldsymbol{r}$ *for any* $\boldsymbol{p} \gg 0$.

PROOF. Assume that, for each k, $\sum_{i \in S} {}^ie_{-k} \neq 0$. If, for some k, this sum is zero, we will be able to set $r_k = 1$ without loss of generality, since contract k will be held by no one. From the definition of ${}^id(\boldsymbol{p}, \boldsymbol{r})$, it is clear that

$$1 - {}^id(\boldsymbol{p}, \boldsymbol{r}) = \min\left[1, \frac{\boldsymbol{p} \cdot ({}^i\omega + \boldsymbol{r}{}^ie_+)}{-\boldsymbol{p} \cdot {}^ie_-}\right].$$

Thus define, for $\boldsymbol{r} \in [0, 1]^{\ell}$ and $k = 1, \ldots, \ell$.

$$F_k(\boldsymbol{r}) = \frac{\sum_i \min\left[1, \frac{\boldsymbol{p}\cdot({}^i\omega + r^i\boldsymbol{e}_+)}{-\boldsymbol{p}\cdot{}^i\boldsymbol{e}_-}\right] {}^i\boldsymbol{e}_{-k}}{\sum_i {}^ie_{-k}}.$$

Because $\sum_i {}^ie_{+k} = -\sum_i {}^ie_{-k}$, it is clear that $F_k(\boldsymbol{r}) \in [0, 1]$. Let $\boldsymbol{F}(\boldsymbol{r}) = (F_1(r), \ldots, F_\ell(r))$. The continuity of $\boldsymbol{F}$ is obvious. Hence $\boldsymbol{F}$ has a fixed point, applying the Brower fixed point theorem. A fixed point of $\boldsymbol{F}$ is a consistent value of $\boldsymbol{r}$ at prices $\boldsymbol{p}$, by definition. (Note that the condition $\boldsymbol{p} \gg 0$ is used to ensure that

$$\boldsymbol{p}\cdot({}^i\omega + r^i\boldsymbol{e}_+)/-\boldsymbol{p}\cdot{}^i\boldsymbol{e}_- \neq 0/0,$$

and hence that $\boldsymbol{F}$ is well-defined.) *Q.E.D.*

Let $\boldsymbol{r}^*$ be a consistent returns vector and let $\boldsymbol{d}^* = \boldsymbol{d}(\boldsymbol{p}, \boldsymbol{r}^*)$ be a vector of default proportions satisfying eq. (10.2). We will show that there exists no other solution to eq. (10.2). Hence $\boldsymbol{r}^*$ will be unique, except if there is some commodity on which no contracts are outstanding in which case the r_k^* corresponding to it can be set equal to 1, with no effect on any real variables.

Our proof of the uniqueness of $\boldsymbol{d}^*$ relies on two basic facts. First, aggregate wealth remains constant since the default rule only redistributes preexisting contracts. Second, any individual i with ${}^id > 0$ has zero wealth, by the third relation in eq. (10.2).

Lemma 10.2. *The system* 10.2 *has a unique solution.*

Proof. Let $\boldsymbol{d}^*$ and $\boldsymbol{d}'$ be solutions such that $\boldsymbol{d}^* \neq \boldsymbol{d}'$. Let $Q = \{i \mid {}^id' < {}^id^*\} \neq 0$. Let $Q' = \{i \mid {}^id' \geq {}^id^*\}$. We must have that $Q' \neq \phi$, since otherwise ${}^id^* > 0$ for all i, every individual would be bankrupt, and aggregate wealth would be zero, a contradiction. If any member of Q is in debt to any member of Q', the aggregate wealth of members of Q' is higher in the $\boldsymbol{d}'$ situation than in $\boldsymbol{d}^*$. Every member of Q is bankrupt in the $\boldsymbol{d}^*$ situation, since ${}^id^* > 0$ for every $i \in Q$. Thus no member of Q is in debt to any member of Q', for if so, the aggregate wealth of members of Q would be less than zero in the $\boldsymbol{d}'$ situation, violating the consistency of $\boldsymbol{d}'$. Thus all preexisting contracts between members of Q and Q' involve the former as creditors and the latter as debtors. Since contracts between two members of Q cannot affect the aggregate wealth of Q, this aggregate must be at least as great as the value of the naturally occurring endowment of Q. But this contradicts the fact that every member of Q is bankrupt in the d^* situation. *Q.E.D.*

We denote by $\boldsymbol{r}^*(\boldsymbol{p})$, the unique returns ratio vector associated with $\boldsymbol{p} \gg 0$. It is obvious from definitions that $\boldsymbol{r}^*(\cdot)$ is continuous on int Δ^{ℓ}. It should be kept in mind that $\boldsymbol{r}^*(\cdot)$ also depends on $\{({}^i\omega, {}^ie)\}$ which are data of the system at date 1.

10.5. Individual Behavior

Consider an individual who is faced with prices $\boldsymbol{p} \in \Delta^{\ell}$ and returns ratios $\boldsymbol{r} \in [0, 1]^{\ell}$. His endowment is modified by the default rule to $\boldsymbol{\omega} + \boldsymbol{\eta}(\boldsymbol{p}, \boldsymbol{r})$. The individual is to select an *action* which consists of a vector for current consumption, $\boldsymbol{x}$, and a vector of purchases and sales of futures contracts, $\boldsymbol{b}$. If $b_k > 0$, we will say that the individual has purchased a contract for delivery of commodity k at time 2, subject to the provisions of the bankruptcy–default laws; conversely, for sales, $b_k < 0$, delivery is promised subject to these rules. Note that the institutional rules of the economy enter directly into the nature of the contracts themselves.

We denote an action by $z = (\boldsymbol{x}, \boldsymbol{b}) \in R^{\ell}$. We assume that *consumption sets* are the non-negative orthants of the appropriate commodity space; hence $\boldsymbol{x} \geq 0$. No restriction is placed on the domain of $\boldsymbol{b}$.

We next determine the *consequences* of an action. Consequences have three components: consumption in each of the two periods, and the extent of bankruptcy, if any, in period 2.

The second period has essentially the same structure as the first. At any prices–returns combination, the individuals' endowments are modified according to the same default rules. There will, of course, be only one consistent returns vector associated with each period 2 price vector. However, since the returns vector depends on the preexisting contracts of all individuals, any one individual cannot determine what it will be if his only information is prices and his own actions. We will use $\sim$ over any letter to denote the magnitude in date 2 that corresponds to the one in date 1 represented by this letter.

An individual having taken the action $z = (x, b)$ has preexisting contracts at date 2 equal to $\eta^2(p, r) + b \equiv \tilde{e}$. If prices are $\tilde{p}$ and returns are $\tilde{r}$ and

$$\tilde{p} \cdot (\omega^2 + \tilde{r}\tilde{e}_+ + \tilde{e}_-) \geq 0,$$

the individual selects $\tilde{x} \geq 0$ such that

$$\tilde{p} \cdot \tilde{x} \leq \tilde{p} \cdot (\omega^2 + \tilde{r}\tilde{e}_+ + \tilde{e}_-).$$

If the value of his wealth is negative, so that he must default on some of his debt, he is forced to consume $\tilde{x} = 0$. The *extent of bankruptcy y* associated with this position is defined as

$$y = \tilde{d}(\tilde{p}, \tilde{r})\tilde{p} \cdot \tilde{e}_-$$

which is the value of the contracts on which he is defaulting. We define $\tilde{d}(\cdot, \cdot)$ exactly as $d(\cdot, \cdot)$.

Thus the extent of bankruptcy is non-zero only when consumption is zero, and conversely, consumption in the second period can be positive only if the individual is solvent at that date.

We assume that the individuals' attitudes regarding a consequence $(x, \tilde{x}, y)$ can be described by a von Neumann–Morgenstern utility function u. The domain of u will be all triples $(x, \tilde{x}, y)$ satisfying $x \in R_+^{\ell_1}$, $x \in R_+^{\ell_2}$, $y \in R_-$, and $y\tilde{x} = 0$, in accordance with the remarks above. One should note that the domain of u is non-convex because of the last condition. This will lead to a non-convexity in the demand correspondence and hence to the fact that we will only be able to prove the existence of an approximate equilibrium, following the methods of Starr [15], with better approximations as the number of individuals becomes large.

With respect to u, we assume:

(u.1) $u(x, \tilde{x}, 0)$ is concave in its first two arguments;
(u.2) $u(x, 0, y)$ is concave in its third argument for each $x \in R_+^{\ell_2}$;
(u.3) u is strictly monotone in all its arguments throughout the domain; and
(u.4) u is bounded above by $\bar{u}$.

Thus a solvent individual at date 2 will choose $\tilde{x} \in R_+^{\ell_2}$ to maximize $u(x, \tilde{x}, 0)$ subject to

$$\tilde{p} \cdot \tilde{x} \leq \tilde{p} \cdot (\omega^2 + \tilde{r}\tilde{e}_+ + \tilde{e}_-).$$

The utility of a bankrupt individual is $u(x, 0, \tilde{d}(\tilde{p}, \tilde{r})\tilde{p}\tilde{e}_-)$.

For each action (x, b) at date 1, and every

$$(\tilde{p}, \tilde{r}) \in \text{int}\Delta^{\ell_2} \times [0, 1]^{\ell_2}$$

define $\phi(x, b, \tilde{p}, \tilde{r})$ as the value of the second period problem of an individual who has taken this action and is faced with this situation. Note that the function ϕ depends on the (p, r) faced at time 1 because $\tilde{e}$ varies with this through η.

At time 1, the individual is assumed to maximize his expected utility subject to the budget constraint that $p \cdot z \leq p^1 \cdot (\omega^1 + \eta^1(p, r))$. We describe his expectations as follows.

For each $(p, r) \in \Delta^\ell \times [0, 1]^\ell$ the expectations of the individual regarding $(\tilde{p}, \tilde{r})$ are given by the measure on $\Delta^{\ell_2} \times [0, 1]^{\ell_2}$ (with its Borel σ-field) denoted $\psi(p, r)$. Thus,

$$\psi : \Delta^\ell \times [0, 1]^\ell \to M_{\Delta^{\ell_2} \times [0,1]^{\ell_2}}.$$

For $p^2 \neq 0$, define

$$\pi(p) = (p^2 / \textstyle\sum_k p_k^2) \in \Delta^{\ell_2}.$$

We assume that ψ satisfies:

(ψ.1) $\psi(\cdot, \cdot)$ is continuous in the weak topology;

(ψ.2) for every $p \gg 0$ and $r \in [0, 1]^\ell$, $\pi(p) \in \text{int co supp}_\Delta\, \psi(p, r)$ where supp_Δ is the projection of the support on its first factor space, and int co is 'the interior of the convex hull of';

(ψ.3) there exists an open set $C \subseteq \Delta^{\ell_2}$ such that for every $p \in \Delta^\ell$, $C \subseteq \text{supp}_\Delta\, \psi(p, r)$; the set C is assumed to be the same for every individual; and

(ψ.4) for every $(p, r) \in \Delta^\ell \times [0, 1]^\ell$

(i) $\psi(p, r)(\{(\tilde{p}, \tilde{r}) \mid \tilde{p}_k = 0 \text{ for some } k\}) = 0$

(ii) $\psi(p, r)(\{(\tilde{p}, \tilde{r}) \mid \tilde{r}_k = 0 \text{ for some } k\}) < 1$.

We now discuss each of these conditions on ψ. Assumption (ψ.1) means that small changes in the environment of any individual will give rise to small changes in his beliefs. This appears to be a very reasonable condition; it has been used by all previous writers in this area. Nevertheless, we should remark in passing that it would be violated by an individual who behaves as a hypothesis-testing classical statistician. For Bayesians, however, it will follow from some continuity conditions on their underlying postulates.

Assumption (ψ.2) is the condition that no individual believes that any linear combination of futures contracts is *sure* to make an arbitrage profit.

Assumption (ψ.3) is a compatibility condition on the expectations of the various individuals. It asserts that there is some open set that is always given a positive, though perhaps very small, weight in everyone's beliefs. We will discuss this condition and the necessity for assuming it when it is used in the next section.

Assumption (ψ.4)(i) is the condition that no prices are ever expected to be zero. The reason for this is that, since all goods are strictly desired by the individual himself, and since the expectations concern his beliefs about future equilibria, he can be sure that no goods will be free in these equilibria. Assumption (ψ.4)(ii) asserts that futures contracts are never expected to be worthless because of default with certainty.

Returning to the individual's problem at date 1, he solves

$$\max_{(x, b)} \int \phi(x, b, \tilde{p}, \tilde{r}) \mathrm{d}\psi(\tilde{p}, \tilde{r})$$

subject to

$$p^1 \cdot x + p^2 \cdot b \leq p^1 \cdot (\omega^1 + \eta^1(p, r)).$$

We define the individual's *demand correspondence* as follows:

$$\xi(p, r) = \{(x, b) \mid \int \phi(x, b, \tilde{p}, \tilde{r}) \mathrm{d}\psi(\tilde{p}, \tilde{r}) \geq \int \phi(x', b', \tilde{p}, \tilde{r}) \mathrm{d}\psi(\tilde{p}, \tilde{r})$$
$$\text{for all } (x', b') \text{ satisfying}$$
$$p^1 \cdot x' + p^2 \cdot b' \leq p^1 \cdot (\omega^1 + \eta^1(p, r))\}.$$

The remainder of this section is concerned with various properties of $\xi(\cdot, \cdot)$. They are generally similar to standard results used in general equilibrium theory and we will only indicate methods or cite the work of others to conserve space.

LEMMA 10.3. *For all* $(p, r) \in \operatorname{int} \Delta^\ell \times [0, 1]^\ell$, $\xi(p, r) \neq \phi$.

PROOF. Using the boundedness of u from above, the concavity property (u.2) and the no-sure-thing property (ψ.2), one may appeal to the result of Leland [11, theorem III], on the existence of optimal policies under uncertainty.

LEMMA 10.4. $\xi(\cdot, \cdot)$ *has a closed graph.*

PROOF. Using the methods of Grandmont [5], Sondermann [4] or Green [3], this can be proven directly. The proof in our case will rely on the continuity of $\eta(\cdot, \cdot)$, as well as on continuity of expectations (ψ.1) and utility, which follows from (u.1) and (u.2).

LEMMA 10.5. *If $\langle p^j \rangle \in \text{int } \Delta^\ell$ and $p^j \to \bar{p} \in bdy\ \Delta^\ell$, then there exists an individual i such that if ${}^i z^j \in {}^i\xi(p^j, r^*(p^j))$, then $\|{}^i z^j\| \to \infty$.*

PROOF. We will treat two separate cases:
(I) $\bar{p}^2 \neq 0$, and
(II) $\bar{p}^2 = 0$.

In case (I) standard methods apply directly, since the good that becomes free in the limit can be purchased to a greater extent and the extra expenditure can be financed with sales of a futures contract with a positive limiting value. The required financing approaches zero in the limit, and hence continuity of the expected utility index suffices to contradict the optimality of any bounded sequence of actions for any individual. Case (II) requires separate treatment, since financing increased purchases of a futures contract may be impossible through decreased purchases of current commodities as current consumption may already be zero. We therefore offer a more detailed proof in this case.

For each j, aggregate wealth is given by $p^{j^1} \cdot \sum_i {}^i\omega^1$. Since $\sum_i {}^i\omega^1 \gg 0$ we have that, for j sufficiently large, there exists $\varepsilon > 0$ such that $p^{j_1} \cdot \sum_i {}^i\omega^1 = I\,\varepsilon > 0$.

Thus there must be some individual i and some subsequence (retain the index j) such that

$$p^{j^1} \cdot ({}^i\omega^1 + {}^i\eta^1(p^j, r^*(p^j))) \geq \varepsilon$$

for all j, since otherwise aggregate wealth would converge to zero, contradicting the statement above. Consider this individual to be fixed; we shall drop the index i for brevity. Let $z^j = (x^j, b^j)$ and $z^j \in \xi(p^j, r^*(p^j))$ and assume $\|z^j\|$ is bounded. Let a subsequence be selected converging to $(\bar{x}, \bar{b})$.

Since b is bounded, $p^{2j} \cdot b^j \to 0$; hence, for j sufficiently large, $p^{1j} \cdot x^j \geq \varepsilon/2$. In particular, for some further subsequence (retain the index j), and some $\delta > 0$, there exists a current commodity k with $\bar{p}_k^1 > 0$ such that $x_k^j \geq \delta$, for all j.

Thus, for some $\boldsymbol{\alpha} = (\alpha, \ldots, \alpha) \in \boldsymbol{R}^{\ell_2}$, $\boldsymbol{\alpha} > 0$, there exists a sequence $\boldsymbol{x}'^j \to \bar{\boldsymbol{x}}$ and $0 \leq \boldsymbol{x}'^j < \boldsymbol{x}^j$ such that

$$\boldsymbol{p}^j \cdot \boldsymbol{z}'^j = \boldsymbol{p}^j \cdot (\boldsymbol{x}'^j, \boldsymbol{b}^j + \boldsymbol{\alpha}) = \boldsymbol{p}^{j^1} \cdot (\omega^1 + \boldsymbol{\eta}^1(\boldsymbol{p}^j, \boldsymbol{r}^*(\boldsymbol{p}^j))).$$

By continuity of the expected utility index and strict monotonicity of the utility function we obtain a contradiction; hence $\| \boldsymbol{z}^j \|$ must be unbounded. *Q.E.D.*

10.6. Equilibrium

Let us define the aggregate excess demand, after convexification of the individual demand correspondences, as

$$\zeta(\boldsymbol{p}, \boldsymbol{r}) = \sum_{i \in S} {}^i\xi(\boldsymbol{p}, \boldsymbol{r}) - \left\{ \sum_{i \in S} {}^i\boldsymbol{\omega}^1, 0 \right\}.$$

That is, excess demand for current commodities is demand minus supply; excess demand for futures contracts is the sum of offers to buy such contracts minus offers to sell them.

Equilibrium is defined as a $(\boldsymbol{p}, \boldsymbol{r})$ such that $0 \in \zeta(\boldsymbol{p}, \boldsymbol{r})$ and $\boldsymbol{r}$ is consistent at $\boldsymbol{p}$.

The proof of the existence of an equilibrium for this convexified excess demand function will follow essentially standard lines, although at two points the usual techniques will have to be modified. It is at these instances that we will use assumptions ($\boldsymbol{u}$.4) (bounded utility), ($\boldsymbol{u}$.2) (concave subjective bankruptcy penalty) and ($\boldsymbol{\psi}$.3) (existence of a common open set that is always given positive weight).

It can be shown, given the existence of an equilibrium for the convexified demand, that an approximate equilibrium exists for the original economy. For this result, one can employ the methods of Starr [15]. An alternative approach would be to use an atomless measure space of agents, in which case an exact equilibrium exists which can be viewed as the limit of approximate equilibria for sequence of large economies converging to it in a suitably defined way; see Hildenbrand [8] and Hildenbrand, Schmeidler and Zamir [9].

10.6.1. A sketch of the proof of the existence theorem

We will now sketch the proof we shall use for the existence theorem and point out those instances at which further reasoning, of a non-standard variety, is required. Both problems arise through the fact that, when bankruptcy is allowed, the set of allowable actions is not bounded below.

First, we observe that, by virtue of the results of section 10.4, it suffices to consider, for each $\boldsymbol{p} \in \text{int } \Delta^{\ell}$, only $\boldsymbol{r}^*(\boldsymbol{p})$, the unique consistent returns vector. Equilibrium could equally well have been defined as a $\boldsymbol{p} \in \text{int } \Delta^{\ell}$ for which $0 \in \zeta(\boldsymbol{p}, \boldsymbol{r}^*(\boldsymbol{p}))$. Thus we let $\langle S^j \rangle$ be an increasing sequence of compact, convex subsets of int Δ such that $\bigcup_{j=0}^{\infty} S^j \supseteq \text{int } \Delta^{\ell}$. To each S^j we apply the fundamental lemma of Debreu [2] which, by virtue of lemma 10.4, implies the existence of $\boldsymbol{p}^j \in S^j$ and $z^j \in \zeta(\boldsymbol{p}^j, \boldsymbol{r}^*(\boldsymbol{p}^j))$, such that $\boldsymbol{p}^j \cdot z^j = 0$ and $\boldsymbol{p} \cdot z^j \leq 0$ for all $\boldsymbol{p} \in S^j$.

One then must show that $\{z^j\}$ is bounded. This is the first point at which some new arguments, based on the assumptions mentioned, must be advanced. In the Arrow–Debreu model, the boundedness of z^j followed directly from $\boldsymbol{p}^\circ \cdot z^j \leq 0$ for some $\boldsymbol{p}^\circ \gg 0$, $\boldsymbol{p}^\circ \in S^\circ$, and the boundedness of the actions z from below (consumption sets bounded below). One then must prove that if $\boldsymbol{p}^j \to \bar{\boldsymbol{p}} \in \boldsymbol{bdy}\, \Delta^{\ell}$, then $\{z^j\}$ is not bounded. Usually this follows directly from lemma 10.5 and boundedness of the consumption sets from below. But without this, it is possible that although excess demands for two individuals become large their sum remains bounded because their speculations on futures transactions 'cancel each other out'.

Combining these results, one has that $(\boldsymbol{p}^j, z^j)$ has a subsequence converging to $(\boldsymbol{p}^*, z^*)$ with $\boldsymbol{p}^* \in \text{int } \Delta^{\ell}$. Because the excess demand correspondence has a closed graph, $z^* \in \zeta(\boldsymbol{p}^*, \boldsymbol{r}^*(\boldsymbol{p}^*))$, and $z^* = 0$ follows from $\boldsymbol{p} \cdot z^* \leq 0$ for all $\boldsymbol{p} \in \text{int } \Delta^{\ell}$.

We now lay the groundwork for the two missing steps in the above argument with the following lemmas.

LEMMA 10.6. *Let C be an open subset of* int Δ^{ℓ_2}. *There exists $C^* \subseteq C$, C^* open and a real number $\delta > 0$ such that, for any partition, $\mathscr{P}$, of the indices $\{1, \ldots, \ell_2\}$ into two non-empty subsets K_+ and K_-, there exists $V_{\mathscr{P}} \subseteq C$, an open set with the property that*

$$\boldsymbol{p} \in V_{\mathscr{P}}, \boldsymbol{p}^* \in C^*$$

implies

$$p_k - p_k^* > \delta \qquad \text{for } k \in K_-$$

and

$$p_k - p_k^* < -\delta \qquad \textit{for } k \in K_+.$$

PROOF. Choose $p^\circ \in C$ arbitrarily. Let $\beta > 0$ be such that $|p - p'| < \beta$ for all $p \in C$. Let $\varepsilon = \beta/2\ell_2$.

Thus, for any partition $\mathscr{P} = (K_+, K_-)$ where $\#K_+ = a_+$ and $\#K_- = a_-$ are the number of members of K_+ and K_-, we have

$$p'_{\mathscr{P}} = \begin{cases} p_k^\circ + \varepsilon a_- & \text{for } k \in K_+ \\ p_k^\circ - \varepsilon a_+ & \text{for } k \in K_- \end{cases}$$

is in C. (Note that $p'_{\mathscr{P}} \in \Delta^{\ell_2}$ by construction.)

Let C^* be the $\varepsilon/4$-ball about p°, and $V_{\mathscr{P}}$ the $\varepsilon/4$-ball about $p'_{\mathscr{P}}$. Let $\delta = \varepsilon/2$. $V_{\mathscr{P}}$ is clearly open and $V_{\mathscr{P}} \subseteq C$ because $|p_k - p'_{\mathscr{P}k}| < \varepsilon/4$ and $|p_k^\circ - p'_{\mathscr{P}k}| < \beta/2$ imply $|p_k - p_k^\circ| < \beta$. To show the property stated in the lemma, choose $p \in V_{\mathscr{P}}$ and $p^* \in C^*$, drop the index $\mathscr{P}$, considering the partition to be fixed, and let $k \in K_+$ be fixed.

By definition of p' we have

$$\varepsilon \leq \varepsilon a_- = p_k^\circ - p'_k.$$

Since $p \in V$ and $p^* \in C^*$ we have

$$p_k^\circ < p_k^* + \varepsilon/4,$$
$$p'_k < p_k - \varepsilon/4.$$

Combining the last three relations, we have

$$p_k - p_k^* < -\varepsilon/2 = -\delta.$$

Exactly analogous reasoning can be used for $k \in K_-$. *Q.E.D.*

LEMMA 10.7. *Let T be open in C; then*

$$\psi_\Delta(p, r)(T) = \psi(p, r)(\{(\tilde{p}, \tilde{r}) \in \Delta^{\ell_2} \times [0, 1]^{\ell_2} \,|\, \tilde{p} \in T\})$$

is bounded away from zero as a function of (p, r).

PROOF. Let

$$\{(\tilde{p}, \tilde{r}) \in \Delta^{\ell_2} \times [0, 1]^{\ell_2} \,|\, \tilde{p} \in T\} \equiv X.$$

Suppose that $\psi(p, r)(X)$ is not bounded away from zero. Then there exists a sequence $\langle(p^j, r^j)\rangle$ such that $\psi(p^j, r^j)(X) \to 0$. We can assume, without loss of generality, that $(p^j, r^j) \to (\bar{p}, \bar{r})$. By the weak continuity of ψ (assumption (ψ.1)) we have

$$\lim_j \inf \psi(p^j, r^j)(X) \geq \psi(\bar{p}, \bar{r})(X),$$

since X is open in $\Delta^{\ell_2} \times [0, 1]^{\ell_2}$. (The equivalence of weak continuity to this statement may be found in Parthasarathy [12].) Therefore $\psi(\bar{p}, \bar{r})(X) = 0$. Then $\psi(\bar{p}, \bar{r})((\Delta^{\ell_2} \times [0, 1]^{\ell_2})\backslash X) = 1$. But this contradicts ($\psi$.3) since, if it holds, we have $\psi(p, r)(\{\tilde{p}, \tilde{r}) \in \Delta^{\ell_2} \times [0, 1]^{\ell_2} \mid \tilde{p} \in C\backslash T\})$ $= 1$ which implies $\mathrm{supp}_\Delta\, \psi(\bar{p}, \bar{r}) \subseteq C\backslash T$. This contradiction establishes the lemma. *Q.E.D.*

The index i of an individual is deleted throughout the following lemma in which the demand for futures contracts is characterized, to shorten the notation.

LEMMA 10.8. *Let* $z^j = (x^j, b^j)$ *and* $z^j \in \xi(p^j, r^*(p^j))$ *for* $j = 1, 2, \ldots$. *There exists a number B such that* $\| b^j \| > B$ *implies*

$$\tilde{p} \cdot (\omega + \eta(p^j, r^*(p^j)) + b^j) \geq 0$$

for all $\tilde{p} \in C^*$, *where* C^* *is the set whose existence is asserted in lemma* 10.6.

PROOF. Suppose not. Then one can extract a subsequence diverging to $+\infty$ in norm such that, for each j, there exists $\tilde{p}^j \in C^*$ and $\tilde{p}^j \cdot (\omega + \eta(p^j, r^*(p^j)) + b^j) < 0$. If $\{b^j\}$ is bounded above, then $\tilde{p} \cdot (\omega + \eta^j + b^j) < 0$ for all $\tilde{p} \in C^*$ (where we have written η^j for $\eta(p^j, r^*(p^j))$). Since the utility function diverges to $-\infty$ as its third argument becomes large and negative (follows from (u.2), (u.3) and (u.4)), and since $\psi(\cdot, \cdot)(\{(\tilde{p}, \tilde{r}) \mid \tilde{p} \in C^*\})$ is bounded away from zero by lemma 10.7 and since u is bounded above (u.4), the expected utility of these actions diverges to $-\infty$. Since $x = \tilde{x} = y = 0$ is always attainable with certainty, this violates $z^j \in \xi(p^j, r^*(p^j))$ for large j. Thus $\{b^j\}$ is not bounded above.

Hence, for a subsequence which we can assume to be the original, there exists a partition $\mathscr{P} = (K_+, K_-)$ of the indices $\{1, \ldots, \ell_2\}$ such that

$$b^j_k \to \infty \qquad \text{for some } k \in K_+$$

and $\{b^j_k\}$ is bounded below for all $k \in K_+$ and

$$b^j_k \to \infty \qquad \text{for some } k \in K_-$$

and $\{b^j_k\}$ is bounded above for all $k \in K_-$.

Let $V_{\mathscr{P}}$ and $\delta > 0$ be as asserted in Lemma 10.6. We have that, for all $p' \in V_{\mathscr{P}}$ and $\tilde{p} \in C^*$,

$$p' \cdot (\omega + \eta^j + b^j) \leq \hat{p} \cdot (\omega + \eta^j + b^j) - \delta \, || \omega + \eta^j + b^j ||$$

The first term on the right is non-positive and the second diverges to $-\infty$. Hence, using the same properties of utility and expectations appealed to in the case of $\{b^j\}$ bounded above, the expected utilities of $\{z^j\}$ must diverge to $-\infty$, contradicting $z^j \in \xi(p^j, r^*(p^j))$. This contradiction establishes the lemma. *Q.E.D.*

LEMMA 10.9. *Let* $\langle p^j \rangle \in \text{int}\, \Delta^\ell$ *and* $p^j \to \bar{p} \in \textbf{bdy}\, \Delta^\ell$; *let* $z^j \in \zeta(p^j, r^*(p^j))$; *then* $||z^j|| \to \infty$.

PROOF. Let

$$z^j = (x^j, b^j) = \sum_{i \in S} ({}^i x^j, {}^i b^j).$$

By lemma 10.5, there exists an individual i such that $||({}^i x^j, {}^i b^j)|| \to \infty$. If $||{}^i x^j| \to \infty$ for some i, then, because ${}^i x^j \geq 0$, we have $||x^j|| \to \infty$, and the result of the lemma holds.

Thus assume that $\{{}^i x^j\}$ is bounded for all i, hence that $||{}^i b^j|| \to \infty$ for some i. Let $S' = \{i \in S| \; ||{}^i b^j|| \to \infty\}$. By the last lemma,

$$\tilde{p} \cdot ({}^i\omega + {}^i\eta(p^j, r^*(p^j)) + {}^i b^j) \geq 0$$

for all $\tilde{p} \in C^*$, and all $i \in S'$. Since C^* is open we have that

$$\tilde{p} \cdot ({}^i\omega + {}^i\eta(p^j, r^*(p^j)) + {}^i b^j) \to \infty$$

for $\tilde{p} \in C^*$, $i \in S'$, as j becomes large. (It is at this point that the openness of C, and hence C^*, becomes crucial). Hence,

$$\tilde{p} \cdot (\sum_i {}^i\omega + \sum_i {}^i\eta(p^j, r^*(p^j)) + \sum_i {}^i b^j) \to \infty$$

for such i, j and $\tilde{p}$. But $\sum_i {}^i\eta(p^j, r^*(p^j)) = 0$ because $r^*(p^j)$ is consistent at p^j by definition. Since ${}^i\omega$ is constant, the limit above implies $\sum_i {}^i b^j_k \to \infty$ for some k. *Q.E.D.*

THEOREM. *There exists an equilibrium for the convexified economy.*

PROOF. Let D^j be an increasing sequence of compact convex sets in int Δ^ℓ such that int $\Delta^\ell \subseteq \bigcup_j D^j$. Applying the lemma of Debreu to the convexified aggregate excess demand $\zeta(\cdot, r^*(\cdot))$ restricted to each successive D^j, we obtain the existence of a sequence $\langle p^j, z^j \rangle$ in $\Delta^\ell \times R^\ell$ such that $p^j \cdot z^j = 0$, $z^j \in \zeta(p^j, r^*(p^j))$ and $p \cdot z^j \leq 0$ for all $p \in D^j$, for each j. We first show that $\{z^j\}$, so constructed, is bounded.

If $\|x^j\| \to \infty$, then $\|b^j\| \to \infty$, for otherwise $p \cdot z^j > 0$ for some $p \in D^1$. Hence it suffices to show that $\{b^j\}$ is bounded. If not, then, by lemma 10.8,

$$\tilde{p} \cdot (\sum_i {}^i\omega + b^j) \geq 0 \qquad \text{for all } \tilde{p} \in C^*.$$

Since C^* is open,

$$\tilde{p} \cdot (\sum_i {}^i\omega + b^j) \to \infty.$$

Let $\bar{p}$ and $\bar{j}$ be selected such that, for some $\alpha > 0$, $\bar{p} = (\bar{p}^1, \alpha\tilde{p}) \in D^{\bar{j}}$ for some $\tilde{p} \in C^*$. Thus $\bar{p} \cdot z^j \to \infty$ since $x^j \geq 0$. But this contradicts $\bar{p} \cdot z^j \leq 0$ for $j \geq \bar{j}$. Hence $\{z^j\}$ is bounded.

Extract a convergent subsequence $\langle p^j, z^j \rangle \to (p^*, z^*) \in \Delta^\ell \times R^\ell$ (retaining the index j). We have that $p^* \in \operatorname{int} \Delta^\ell$, for if not, $\{z^j\}$ would be unbounded according to lemma 10.9. It follows that $z^* \in \zeta(p^*, r^*(p^*))$, since ζ has a closed graph. If $z_k^* < 0$ for some k and $z^* \leq 0$, then $p^* \cdot z^* \neq 0$, contradicting $p^j \cdot z^j = 0$ for all j. If $z_k^* > 0$ for some k, we obtain a contradiction to $p \cdot z^* \leq 0$ for all $p \in \operatorname{int} \Delta$ by considering

$$p = (\varepsilon, \ldots, 1 - (\ell_2 - 1)\varepsilon, \varepsilon, \ldots, \varepsilon)$$

for $\varepsilon > 0$ sufficiently small, where the element $1 - (\ell_2 - 1)\varepsilon$ is in the kth place. *Q.E.D.*

COROLLARY. *There exists an approximate equilibrium for the original economy.*

The methods of Starr [15] can be used to prove this. We omit the demonstration; see also Arrow and Hahn [1], ch. 7.

10.7. Conclusion

In this paper we have explored a typical period in a model of pure exchange with the features that markets are known to reopen in every future period, and markets exist for all commodities, current and future, at the present date. In this way, time has been incorporated explicitly into a general equilibrium theory. The model has the property that the institutional structure of sequential trading is validated by the need and choice of the agents to trade at each date. Further, the system can be viewed as generating a sequence of such temporary competitive equilibria because the results of previous equilibria will give rise to economic environments satisfying the assumptions needed for existence of an equilibrium in any given time period.

These assumptions seem to be quite mild. Further investigation along these lines will have to make stronger qualitative assumption on the basic data of the system in order to deduce results in comparative statics or limiting results on the sequence of temporary equilibria. Other open questions concern the role of firms in models of uncertainty and the incorporation of expectations generating hypotheses that will be useful in studying properties of the model. It is hoped that models of this type will provide a framework for the study of monetary theory.

References

[1] Arrow, K. J. and Hahn, F. H. *General Competitive Analysis*. Holden Day: San Francisco (1971).
[2] Debreu, G. Market equilibrium. *Proceedings of the National Academy of Sciences of the USA*, **42** (1956), 876–878.
[3] Green, J. *Temporary general equilibrium in a sequential trading model with spot and futures transactions. Econometrica*, **41** (November, 1973), 1103–1124.
[4] Grandmont, J.-M. *On the temporary competitive equilibrium*. Working paper No. 305, Center for Research in Management Science, University of California at Berkeley (August 1970).
[5] Grandmont J.-M. *On the short-run equilibrium in a monetary economy*. CEPREMAP discussion paper, Paris (February 1971).
[6] Hahn, F. H. Equilibrium with transaction costs. *Econometrica*, **39** (1971), 417–439.
[7] Hicks, J. R. *Value and Capital*, second edition. Clarendon Press: Oxford (1946).
[8] Hildenbrand, W. On economies with many agents. *Journal of Economic Theory*, **2** (1970), 161–188.
[9] Hildenbrand, W., Schmeidler, D. and Zamir, S. Existence of approximate equilibria and core. *Econometrica*, **41** (November, 1973), 1159–1166.
[10] Kurz, M. *Equilibrium in a finite sequence of markets with social transaction cost*. Technical report No. 52, Institute for Mathematical Studies in the Sciences, Stanford University (February 1972).
[11] Leland, H. E. On the existence of optimal policies under uncertainty. *Journal of Economic Theory*, **4** (1972), 35–44.
[12] Parthasarathy, K. *Probability Measures on Metric Spaces*. Academic Press: New York (1967).

[13] Radner, R. *Existence of equilibrium of plans, prices, and price expectations in a sequence of markets. Econometrica,* **40** (March, 1972), 289–303.
[14] Sondermann, D. *Temporary competitive equilibrium under uncertainty.* Presented to the European Research Conference on Economic Theory, Bergen, Norway (July–August 1971).
[15] Starr, R. M. Quasi-equilibria in market with non-convex preferences. *Econometrica,* **37** (1969), 25–38.
[16] Stigum, B. Competitive equilibria under uncertainty. *Quaterly Journal of Economics,* **83** (1969), 533–561.

COMMENTS

*On preexisting contracts and temporary equilibria**

Bernt P. Stigum

C10.1.1. Discussion

In his paper 'Preexisting contracts and temporary general equilibrium', Green studies 'a model of general economic equilibrium over time in which the markets for trading commodities are open at every date' (cf. [3], p. 263). His main objective is to establish (1) codes of conduct, and (2) sufficient conditions for the existence of a temporary equilibrium in a model in which there is a real possibility that one or more consumers might go bankrupt.

Green's paper provides many interesting insights into the problems of modeling the behavior of individuals in a 'bankruptcy world'. But the paper also has several serious defects which should be pointed out. First, Green's model is theoretically weak for reasons detailed in sections C10.1.1.1–C10.1.1.4 below.

C10.1.1.1. Green considers a two-period exchange economy which operates in a world in which there exists one and only one state of

* Sponsored by the United States Army under Contract No. DA-31-124-ARO-D-462.

nature. This economy possesses in period 1 markets for current commodities and for claims on all (!) future (i.e. period 2) commodities. In period 2 it possesses markets for period 2 commodities only. Thus the institutional structure of the Green economy differs from the Arrow–Debreu economy [2, ch. 7] only in the fact that markets are allowed to open in period 2.

In the Green economy, individual consumers in period 1 entertain expectations about period 2 prices that are multi-valued with respect to the state of nature. These expectations depend both on the prices of current goods and on the futures prices of period 2 goods. Moreover they satisfy certain reasonable postulates (cf. [3], p. 276).

It is true and ought to be obvious to Green's consumers that by acting in accordance with these expectations they will end up with a suboptimal allocation of resources over time. If they instead all behaved as if the futures prices of period 2 goods equalled the present value of the true period 2 prices, then they would achieve a Pareto-optimal allocation of resources, they would not have to worry about the possibility of bankruptcy in period 2 and they could dispense with markets in period 2 altogether.

I can't help but believe that consumers in Green's economy would eventually decide to behave like ordinary Arrow–Debreu consumers rather than in the way predicted by Green.

The preceding criticism would not be a serious one if Green's objective were only to model the first-period behavior of an economy in which consumers in period 1 could not determine the price at which commodities were to be exchanged in period 2 because, say, of a lack of certain futures markets. Then Green's model could have been justified by saying that the allowance of a complete set of futures markets in period 1 enabled Green to simplify his notation without decreasing the generality of his results. However, Green is not interested just in the first-period behavior of his economy. He is in fact attempting to model the behavior of his economy over two periods. For that reason I consider it a serious criticism.

In this respect note also that if Green were to meet the criticism detailed at the beginning of the section, i.e. if he were to throw out one or more futures markets, he would simultaneously have to make drastic changes in other facets of his model. For one thing Green would have to modify his bankruptcy law since he would have no way of evaluating the vector $(\boldsymbol{\omega} + \boldsymbol{r}\boldsymbol{e}_+ + \boldsymbol{e}_-)$ [3, p. 270] and hence of determining $\boldsymbol{d}(\cdot)$ and $\boldsymbol{\eta}(\cdot)$ [3, p. 270]. Actually if Green were to modify his bankruptcy law, it would improve his model. I will now give the reasons why it would improve his model.

C10.1.1.2. Green treats the concepts of bankruptcy and negative net worth as synonymous. This is all right in an Arrow–Debreu world (cf. [2, ch. 7]). It is also all right as long as we discuss consumer behavior in Arrow and Hahn's one-period model (cf. [1, ch. 3]). However, it makes no sense at all when we discuss consumer behavior in period 1 in Green's model. There, as in the real world, a consumer should not have to declare bankruptcy as long as he is able to obtain enough funds (by borrowing or otherwise) to meet his currently maturing obligations.

Just how unreasonable Green's bankruptcy law is can best be seen by considering two simple examples.

EXAMPLE C10.1. Consider a consumer in period one in Green's economy with 'naturally occurring endowment' $\boldsymbol{\omega} = (\omega^1, \omega^2) > 0$, and with pre-existing contracts $\boldsymbol{e} = (e^1, e^2)$. Suppose that he faces the price vector $\boldsymbol{p} = (p^1, p^2)$ and the return vector $\boldsymbol{r} = (r^1, r^2)$, and that

$$p^1(r^1 e^1_+ + e^1_-) > 0, \tag{C10.1}$$

$$p^2(r^2 e^2_+ + e^2_-) < 0, \tag{C10.2}$$

$$\boldsymbol{p}(\boldsymbol{\omega} + \boldsymbol{r}\boldsymbol{e}_+ + \boldsymbol{e}_-) < 0. \tag{C10.3}$$

This consumer has no difficulty meeting his first-period obligations. Moreover, since p^2 is only a point forecast of period 2 prices, it is entirely possible that there are vectors $(\boldsymbol{x}^1, \boldsymbol{b})$ with $\boldsymbol{x}^1 \geq 0$ such that

$$p^1\boldsymbol{x}^1 + p^2\boldsymbol{b} = p^1(\omega^1 + r^1 e_+ + e^1_-), \tag{C10.4}$$

and such that in the minds of all consumers in the economy the probability is greater than 0.999 that

$$\tilde{\boldsymbol{p}}[\omega^2 + \tilde{\boldsymbol{r}}(\boldsymbol{b} + r^2 e^2_+ + e^2_-)_+ + (\boldsymbol{b} + r^2 e^2_+ + e^2_-)_-] \geq 0, \tag{C10.5}$$

where $\tilde{\boldsymbol{p}}$ and $\tilde{\boldsymbol{r}}$ denote the actual period 2 price and return vector respectively. Yet according to Green's bankruptcy law the consumer is bankrupt at $(\boldsymbol{p}, \boldsymbol{r})$ and his endowment must be modified to allow for this fact.

I think it is unreasonable to declare the consumer in example C10.1 bankrupt. He contracted $\boldsymbol{e}$ with the promise to meet his e^1 obligations in period 1 and his e^2 obligations in period 2. As an honorable man he should only have to care about whether there exist vectors $\boldsymbol{x}^1 \geq 0$ and $\boldsymbol{b}$ which satisfy eq. (C10.4), and which satisfy eq. (C10.5) with 'sufficiently large' probability. The signs of the left-hand sides of (C10.1)–(C10.3) are irrelevant.

EXAMPLE C10.2. Consider an economy in which $\ell_1 = \ell_2 = 1$ and assume that in this economy there is a consumer with naturally occurring endowment $\boldsymbol{\omega} = (\omega^1, \omega^2) = (3\tfrac{3}{8}, 9)$ and preexisting contract $\boldsymbol{e} = (e^1, e^2) = (0, -18)$. Assume also that this consumer faces the price vector $\boldsymbol{p} = (p^1, p^2) = (4/7, 3/7)$ and the return vector $\boldsymbol{r} = (r^1, r^2) > 0$, and that he orders $(\boldsymbol{x}, \tilde{\boldsymbol{x}}, \boldsymbol{y})$-vectors according to the values assumed by the function

$$U(\boldsymbol{x}, \tilde{\boldsymbol{x}}, \boldsymbol{y}) = \sqrt{(x)} + (1/6)\tilde{x} + (1/12)y, \ (x, \tilde{x}) \geq 0, y \leq 0. \quad \text{(C10.6)}$$

For this consumer the fraction d which satisfies

$$(4/7)\omega^1 + (3/7)(\omega^2 + (1 - d)e^2_-) = 0 \quad \text{(C10.7)}$$

is given by

$$d = 1/4. \quad \text{(C10.8)}$$

The fraction $\tilde{d}$ which satisfies

$$\min_{0 \leq \tilde{d} \leq 1} \tilde{p}[\omega^2 + \tilde{r}(b + (1 - d)e^2_-)_+ + (1 - \tilde{d})(b + (1 - d)e^2_-)_-] \geq 0 \quad \text{(C10.9)}$$

is given by

$$\tilde{d} = \begin{cases} 0 & \text{if } b \geq 9/2 \\ [(9 - 2b)/(27 - 2b)] & \text{for } b < 9/2. \end{cases} \quad \text{(C10.10)}$$

Moreover, if $E\{\cdot \mid \boldsymbol{p}, \boldsymbol{r}\}$ denotes the conditional expectation of $(\cdot)$ given $(\boldsymbol{p}, \boldsymbol{r})$, and if when $\boldsymbol{p}$ and $\boldsymbol{r}$ are as above,

$$E\{(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{r}}) \mid \boldsymbol{p}, \boldsymbol{r}\} = (1, a), \quad \text{(C10.11)}$$

where $0 < a < 1$, then it is easy to show that the consumer's indirect first-period utility function and its partial derivatives are given by

$$F(\boldsymbol{x}, \boldsymbol{b}, \boldsymbol{p}, \boldsymbol{r}) \equiv E(\phi(\boldsymbol{x}, \boldsymbol{b}, \tilde{\boldsymbol{p}}, \tilde{\boldsymbol{r}}) \mid \boldsymbol{p}, \boldsymbol{r}\}$$
$$= \begin{cases} \sqrt{(x)} - [(9 - 2\boldsymbol{b})/24] \text{ for } \boldsymbol{b} \leq 9/2,\ \boldsymbol{x} \geq 0 \\ \sqrt{(x)} + (1/6)(\boldsymbol{b} - 9/2) \text{ for } 9/2 < \boldsymbol{b} \leq 27/2,\ \boldsymbol{x} \geq 0 \\ \sqrt{(x)} + (1/6)(9 + a(\boldsymbol{b} - 27/2)) \text{ for } \boldsymbol{b} > 27/2,\ \boldsymbol{x} \geq 0 \end{cases} \quad \text{(C10.12)}$$

$$(\partial F/\partial \boldsymbol{x})(\boldsymbol{x}, \boldsymbol{b}, \boldsymbol{p}, \boldsymbol{r}) = 1/2\sqrt{(x)},\ \boldsymbol{x} \geq 0, \quad \text{(C10.13)}$$

$$(\partial F/\partial \boldsymbol{b})(\boldsymbol{x}, \boldsymbol{b}, \boldsymbol{p}, \boldsymbol{r}) = \begin{cases} 1/12 \text{ for } \boldsymbol{b} \leq 9/2,\ \boldsymbol{x} \geq 0 \\ 1/6 \quad \text{for } 9/2 < \boldsymbol{b} \leq 27/2 \\ \boldsymbol{a}/6 \quad \text{for } \boldsymbol{b} > 27/2. \end{cases} \quad \text{(C10.14)}$$

With a little algebra it follows easily from eqs. (C10.13) and (C10.14) that

$$(\boldsymbol{x}, \boldsymbol{b}) \equiv (81/4, -45/2) \quad \text{(C10.15)}$$

maximizes $F(\boldsymbol{x}, \boldsymbol{b}, \boldsymbol{p}, \boldsymbol{r})$ subject to the constraint

$$(4/7)\boldsymbol{x} + (3/7)\boldsymbol{b} = 27/14. \quad \text{(C10.16)}$$

The important point to notice about example C10.2 is that at *the beginning of period* 1 the consumer has a second-period debt of 18 units of e^2 which (since his net worth is negative) is considered extravagant and is promptly written down so that it equals 27/2 units of e^2. At *the end of period* 1 the consumer owes $45/2 + 27/2 = 36$ units of e^2! Moreover, he knows that he will be bankrupt in period 2 with probability 1. I think that a code of conduct and a model which allow such a situation to arise are unreasonable. Note that the model in example C10.2 does not violate any of Green's assumptions except the assumption that $U \leq \bar{U}$ [3, p. 275, assumption ($\boldsymbol{u}$. 4)] which is unimportant for the example.

The reader might think that I picked an unreasonable pair of prices (p^1, p^2) in example C10.2, one which could never represent equilibrium prices in Green's economy. That this is not so can be seen from example C10.3 below. But first one more serious criticism of Green's model.

C10.1.1.3. A temporary equilibrium in Green's economy is a vector

$$(\boldsymbol{p}, \boldsymbol{r}, (\boldsymbol{x}^1, \boldsymbol{b})^1, \ldots, (\boldsymbol{x}^1, \boldsymbol{b})^I),$$

where $\boldsymbol{p}$ is a strictly positive price vector and $\boldsymbol{r}$ is a return vector, and where for each $i(\boldsymbol{x}^1, \boldsymbol{b})^i$, $i = 1, \ldots, I$, is a vector of period 1 commodities and claims on period 2 commodities which consumer i demands at $(\boldsymbol{p}, \boldsymbol{r})$ and which satisfies the equations

$$\sum_{i=1}^{I} (\boldsymbol{x}^1, \boldsymbol{b})^i = \sum_{i=1}^{I} (\omega^1, 0)^i. \qquad \text{(C10.17)}$$

In such an equilibrium consumers can be partitioned into three groups N_1, N_2, N_3 with the following properties. Consumer i is in N_1 if and only if he believes that he will be solvent with probability 1 in period 2. Consumer j is N_2 if and only if he believes that he will go bankrupt with probability greater than 0 but less than 1. Consumer k is in N_3 if he believes that he will go bankrupt in period 2 with probability 1. One or two of these three groups may be empty. What is important to note is that Green's assumptions do not preclude the possibility that N_3 might contain one or more consumers.

It seems to me unlikely that institutional arrangements in practice would be such that a consumer would be allowed to buy a vector $(\boldsymbol{x}^1, \boldsymbol{b})$ if it entailed his going bankrupt with probability 1 in period 2. Thus Green's model which permits temporary equilibria that allocate such 'bankruptcy pairs $(\boldsymbol{x}^1, \boldsymbol{b})$' to one or more consumers appears unreasonable to me.

Here is an example which illustrates what I have in mind.

EXAMPLE C10.3. Consider an economy in which there are two consumers A and B and in which $\ell_1 = \ell_2 = 1$. Consumer A has 'naturally occurring endowment' $\omega_A = (5, 5)$, initial holdings of securities $\boldsymbol{e}_A = (-10, 0)$, and utility function

$$U_A(x, \tilde{x}, y) = \sqrt{(x)} + (1/16\sqrt{5})\tilde{x} + (1/16\sqrt{5})y, \; (x, \tilde{x}) \geq 0, \; y \leq 0. \qquad \text{(C10.18)}$$

Suppose consumer A believes that, regardless of the value of $(\boldsymbol{p}, \boldsymbol{r})$, $(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{r}}) = (1, 1)$ with probability 1. His indirect utility function is then given by

$$F_A(x, b, p, r) = \sqrt{(x)} + (1/16\sqrt{5})(5 + b), \; x \geq 0, \; -\infty < b < \infty. \qquad \text{(C10.19)}$$

B has 'naturally occurring endowment' $\omega_B = (15, 5)$, initial holdings of securities $\boldsymbol{e}_B = (10, 0)$, and utility function

$$U_B(x, \tilde{x}, y) = x + 10\sqrt{(\tilde{x})} - y^2, \qquad (x, \tilde{x}) \geq 0, \; y \leq 0. \qquad \text{(C10.20)}$$

Suppose, consumer B too believes that, regardless of the value of $(\boldsymbol{p}, \boldsymbol{r})$, $(\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{r}}) = (1, 1)$ with probability 1. Then his indirect utility function is given by

$$F_{\mathrm{B}}(\boldsymbol{x}, \boldsymbol{b}, \boldsymbol{p}, \boldsymbol{r}) = \begin{cases} \boldsymbol{x} + 10\sqrt{(5 + \boldsymbol{b})} & \text{for } \boldsymbol{b} \geq -5, \boldsymbol{x} \geq 0, \\ \boldsymbol{x} - (5 + \boldsymbol{b})^2 & \text{for } \boldsymbol{b} < -5, \boldsymbol{x} \geq 0. \end{cases} \qquad \text{(C10.21)}$$

It is easy to show that there is one and only one temporary equilibrium in the economy, namely

$$(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{x}_{\mathrm{A}}, \boldsymbol{b}_{\mathrm{A}}, \boldsymbol{x}_{\mathrm{B}}, \boldsymbol{b}_{\mathrm{B}}) = (4/5, 1/5, 5/8, 20, -85, 0, 85), \qquad \text{(C10.22)}$$

with $d = \frac{3}{8}$. In this temporary equilibrium consumer A is allocated $(\boldsymbol{x}_{\mathrm{A}}, \boldsymbol{b}_{\mathrm{A}}) = (20, -85)$. Since $\omega_{\mathrm{A}}^2 = 5$ he will start out period 2 with net worth equal to -80 and will be bankrupt with probability 1.

Example C10.3 displays an economy which has a unique temporary equilibrium during period 1. This equilibrium allocates a vector $(\boldsymbol{x}^1, \boldsymbol{b})$ to one consumer which entails that he will be bankrupt with probability 1 in period 2. The model describing this economy, therefore seems to me to be unreasonable. Yet it satisfies all of Green's assumptions except $U_i(\cdot) \leq \bar{U}$, $i = \mathrm{A, B}$.

Note also that, in the temporary equilibrium portrayed in eq. (C10.22), consumer A borrows (i.e. promises to supply) $8\frac{1}{2}$ times the available supply of the second-period commodity. Yet he believes that $\tilde{\boldsymbol{r}} = 1$ with probability 1. This too seems unreasonable. If a consumer takes a position in $(\boldsymbol{x}^1, \boldsymbol{b})$ which entails that he will go bankrupt with probability 1, he should also know that $\tilde{\boldsymbol{r}} < 1$ with probability 1.

If one wants to avoid temporary equilibria of the sort portrayed in eq. (C10.22) one must impose lower bounds on $\boldsymbol{b}$. In example C10.3 a natural lower bound would be $\boldsymbol{b} \geq -5$ for both A and B. In a more general economy with $\ell_2 > 1$ such bounds are not as easily derived. One possibility would be to insist that, for each and every consumer

$$p^2(\boldsymbol{b} + \eta^2(\boldsymbol{p}, \boldsymbol{r}) + \omega^2) \geq 0.$$

Such a bound would certainly make sense in Green's model since then Green's assumption (ψ.2)(cf. [3, p. 276]) would ensure that each consumer would choose $(\boldsymbol{x}^1, \boldsymbol{b})$ in such a way that he would believe that he would be solvent in period 2 with positive probability.

In my own work on temporary equilibria (which preceded Grandmont's and Arrow and Hahn's work by several years). I have often insisted that consumers in choosing $(\boldsymbol{x}^1, \boldsymbol{b})$ observe an inequality of the form $\boldsymbol{b} \geq \boldsymbol{a}$, where $\boldsymbol{a}$ is an exogeneously determined vector which may vary from one consumer to another. By choosing $\boldsymbol{a}$ judiciously, one should be able to avoid a situation in which people in one period can borrow so much that they will go bankrupt the next period with probability 1.

So much for temporary equilibria and bankruptcy. Next a comment on Green's first-period budget constraint.

C10.1.1.4. Green insists that a consumer is bankrupt if

$$\boldsymbol{p}(\boldsymbol{\omega} + \boldsymbol{r}\boldsymbol{e}_+ + \boldsymbol{e}_-) < 0. \tag{C10.23}$$

Here the left-hand side involves both the consumer's first and second-period 'naturally occurring endowments' $\boldsymbol{\omega} = (\omega^1, \omega^2)$. It also involves both his first and second-period holdings of securities $\boldsymbol{e} = (e^1, e^2)$. Yet the consumer's first-period budget constraint is given by

$$p^1\boldsymbol{x}^1 + p^2\boldsymbol{b} \leq p^1(\omega^1 + \eta^1(\boldsymbol{p}, \boldsymbol{r})), \tag{C10.24}$$

where the right-hand side gives the value of his first-period endowment after the bankruptcy law has been applied.

I think (C10.23) and (C10.24) are incongruous, and I am quite sure that no one consumer in Green's economy would accept (C10.24) as his first-period budget constraint. Certainly, since securities in Green's model are traded in an abstract market, consumers cannot distinguish between e^2 and $\boldsymbol{b}$. Therefore $p^2\eta^2(\boldsymbol{p}, \boldsymbol{r})$ should appear on the right-hand side of (C10.24). Moreover, as long as Green insists that (C10.23) provides the ultimate test of whether a consumer is bankrupt in period 1, $p^2\omega^2$ presumably ought to appear on the right-hand side of (C10.24). Consequently, the 'right' budget constraint faced by each consumer in period 1 is not (C10.24) but instead

$$p^1\boldsymbol{x}^1 + p^2\boldsymbol{b} \leq \boldsymbol{p}(\omega + \eta(\boldsymbol{p}, \boldsymbol{r})). \tag{C10.25}$$

If we were to modify Green's model by substituting (C10.25) for (C10.24), we would also have to change the consumer's second-period budget constraint to

$$\tilde{\boldsymbol{p}}\tilde{\boldsymbol{x}} \leq \tilde{\boldsymbol{p}}(\tilde{\boldsymbol{r}}\boldsymbol{b}_+ + \boldsymbol{b}_-).$$

Moreover, we would have to change the aggregate excess demand function to

$$\xi(p, r) = \sum_{i=1}^{I} ({}^i\bar{\xi}(p, r) - {}^i\omega).$$

Other than that the statements and proofs of Green's results would with only obvious modifications still hold.

The preceding criticisms explain my major reasons for thinking Green's model is unsound. In addition I have three minor comments concerning Green's paper:

(1) Consumers in Green's model are penalized for being bankrupt in period two, but not for being bankrupt in period 1. At least, so it seems to me since it is perfectly possible to be bankrupt in period 1 and be solvent in period 2. I do not understand why it is worse to be bankrupt in period 2 than in period 1.

(2) Green's discussion of previous work on temporary equilibria is misleading. The idea of a temporary equilibrium was first introduced in a 'formal general equilibrium model' (cf. [3], p. 263) in my paper on 'Competitive equilibria under uncertainty' [4], which was privately circulated in 1966, presented at the Winter Meetings of the Econometric Society in Chicago in 1968 and published in the *Quarterly Journal of Economics* in November 1969. Grandmont's model is a special case of my model which considers only one state of nature, one security, two periods, and no initial endowments of preexisting contracts. Moreover, sufficient conditions for the existence of a temporary equilibrium in a bankruptcy world were first given in my paper 'Resource allocation under uncertainty' [5] which was presented at the Winter Meetings of the Econometric Society in New York in 1969 and published in the *International Economic Review* in October 1972.

(3) Green's statement about the fulfilment of contracts in an Arrow–Debreu world (cf. [3], p. 266) seems incorrect to me. Since Debreu does not make any assumption about how consumers order commodities in future periods, Green's assertion that 'even if it were possible to reopen' markets in future periods, 'no-one would want to engage in trade at the equilibrium prices that would emerge' cannot be verified.

To conclude my discussion of Green's paper, I will give a simple example of an economy which satisfies Green's assumptions and yet does not possess a temporary equilibrium. The example shows why Green's seemingly weak result about the existence of ε-equilibria cannot be improved upon.

EXAMPLE C10.4. Consider an economy in which there are two consumers A and B and in which $\ell_1 = \ell_2 = 1$. Assume that ω_A, e_A, and $U_A(\cdot)$ are as in example C10.3. Moreover, assume that

$$U_B(x, \tilde{x}, y) = x + (1/4)\tilde{x} - y^2, (x, \tilde{x}) \geq 0, y \leq 0, \qquad \text{(C10.26)}$$

and that ω_B and e_B are as specified in example C10.3. Finally, assume that both A and B expect that $(\tilde{p}, \tilde{r}) = 1$ with probability 1 regardless of the observed value of (p, r). Then A's indirect utility function is given by eq. (C10.19), and B's is given by

$$F_B(x, b, p, r) = \begin{cases} x + (1/4)(5 + b) & \text{for } b \geq -5, x \geq 0, \\ x - (5 + b)^2 & \text{for } b < -5, x \geq 0. \end{cases} \qquad \text{(C10.27)}$$

Using Green's bankruptcy law it is easy to see that with $r = \min(\frac{1}{2}(1 + (p^2/p^1)), 1)$ A's optimal choice of (x, b) is given by

$$(x_A, b_A)(p^1, p^2) = \begin{cases} (320(p^2/p^1)^2, -320(p^2/p^1) - 5) \quad \text{for } (p^2/p^1) \leq 1, \\ \text{and} \\ (320(p^2/p^1)^2, -320(p^2/p^1) - 5(p^1/p^2)) \\ \text{for } (p^2/p^1) > 1. \end{cases} \qquad \text{(C10.28)}$$

Moreover B's optimal choice of (x, b) is given by

$$(x_B, b_B)(p^1, p^2) = \begin{cases} (0, 20(p^1/p^2) + 5) \text{ if } (p^2/p^1) \leq \alpha \equiv 0.249827 \\ ((p^2/p^1)^2/2 + 10(p^2/p^1) + 20, -5 - (p^2/p^1)/2) \\ \text{if } \alpha \leq (p^2/p^1) \leq 1, \text{ and} \\ ((p^2/p^1)^2/2 + 5(p^2/p^1) + 25, -5 - (p^2/p^1)/2) \\ \text{if } (p^2/p^1) > 1. \end{cases} \qquad \text{(C10.29)}$$

From eqs. (C10.28) and (C10.29) it follows that this economy does not have a temporary equilibrium in period 1.

One more look at (C10.28)–(C.10.29) will show that A's choice of (x, b) is uniquely determined for all pairs (p^1, p^2). B's choice of (x, b) is uniquely determined for all pairs (p^1, p^2) such that $(p^2/p^1) \neq \alpha$. When $(p^2/p^1) = \alpha$, B is indifferent between $(x, b) = (0, 85.05)$ and $(x, b) = (22.53, -5.12)$. Thus if we were to convexify A's and B's demand correspondences as Green does, we would find that the 'convexified' economy possessed one and only one temporary equilibrium in period 1, namely

$$(\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{x}_A, \boldsymbol{b}_A, \boldsymbol{x}_B, \boldsymbol{b}_B) = ((1 + \alpha)^{-1}, \alpha/(1 + \alpha), (1 + \alpha)/2, 19.97, -84.94, 0.03, 84.94). \quad \text{(C10.30)}$$

References

[1] Arrow, K. J. and Hahn, F. H. *General Competitive Analysis.* Holden Day: San Francisco (1971).
[2] Debreu, G. *Theory of Value.* Wiley: New York (1959).
[3] Green, J. Preexisting contracts and temporary general equilibrium. In this volume, pp. 263–286.
[4] Stigum, B. P. Competitive equilibria under uncertainty. *Quarterly Journal of Economics,* **83** (1969), 533–561.
[5] Stigum, B. P. Resource allocation under uncertainty. *International Economic Review* (October 1972).

REPLY TO COMMENTS

Jerry Green

The remarks by Professor Stigum can be classified into three broad areas: methodological criticisms concerning the structure of a model in which there are future markets in the present that reopen at a later date, objections to the definition of bankruptcy, and criticisms of the budget constraint. I will take issue with each of these and will try to indicate the reasons for the path I have followed in the paper.

At the top of p. 287, the discussant states that '... thus the institutional structure of the Green economy differs from the Arrow–Debreu economy *only* in the fact that markets are allowed to be open in period 2' [my italics]. This is correct, but the discussant has not perceived that many of the generalizations of the Arrow–Debreu model which are currently being studied lead to the institution of sequential trading. My paper, while not addressing any of these generalizations directly, has tried to recast the institutional structure of general equilibrium theory in such a way that it will be able to handle, in a consistent fashion, some of the more complex phenomena which cannot be modeled in the Arrow–Debreu framework. Among the motivations for adopting the alternative structure, we can cite the work of Hahn on transactions costs, Radner on differential information, and the numerous contributions to the overlapping generations' literature. If any of these are present in the Arrow–Debreu system, there may be a mutual incentive for agents to reopen markets at later dates, as the economy evolves in time. This reopening will cause them to recast their decision problem at the initial date and it is on this re-evaluation of the equilibrium at the initial date that I have tried to focus.

Professor Stigum's comments on *ex post* efficiency and his insistence that individuals would realize that they would be better off with the Arrow–Debreu framework indicate that he believes my model to be a substitute for the Arrow–Debreu model in the same economic environment. I do not view it in this way; and it is in fact only the differences in the economic environment that can justify the difference in the institutional structure I have chosen. However, rather than introduce many things at once, I tried to pose a model as close as possible to the Arrow–Debreu pure-exchange model, but with sequential trading possible. This would focus on the differences in the institutional framework as distinct from those introduced by additional complexities in the systems.

The discussant thinks that using a different definition of bankruptcy would improve the model. The proposal is to say that an individual is bankrupt if he cannot meet his currently maturing obligations 'by borrowing or otherwise'. But this will surely lead to inconsistencies, for, *since the market is anonymous*, everyone would carry vast amounts of debt throughout their lifetimes, consuming at unrealistically high levels. The proposed definition might be a good one in the context of a bilateral trading model where credit can be rationed according to the individual's asset position, as I suggest in section 10.3 of my paper. This criticism of the bankruptcy law should really be a criticism of the anonymity of the market, and one with which I would certainly agree. We *are* in need of a good bilateral general equilibrium model, but let us not put the burden of the inadequacy of a market equilibrium model where it does not belong.

One should also note that adopting Professor Stigum's suggestion in this regard would force us to declare individuals bankrupt if they could not meet current obligations regardless of the fact that they might be able to borrow against their (large) future naturally occurring endowments to regain solvency – and they would be able to honor these commitments with certainty.

In his example C10.1, the discussant thinks that p^2 is a 'forecast' of $\tilde{p}$, but this is clearly incompatible with '... all consumers in the economy [believing] the probability is greater than 0.999 that

$$\tilde{p}[\omega^2 + \tilde{r}(b + r^2e_+^2 + e_-^2)_+ + (b + r^2e_+^2 + e_-^2)_-] \geq 0.\text{'}$$

If this were the case, speculative forces would have lead to p^2 being a disequilibrium. I believe, therefore, that the argument of the example is inconsistent.

I see nothing pathological with example C10.2. The consumer believes that he will certainly be bankrupt in period 2 and takes this into account through the subjective (non-economic) bankruptcy penalty. The concavity of this penalty function insures us that his actions will be determinate at any p, regardless of his expectations (if they fulfil the assumptions stated in my paper). They are *not a priori* bounded from below, as the discussant suggests I assume – and on which he states that he has 'often insisted' (p. 292). It is just such an assumption, grounded on neither rationality nor institutional restrictions, that I tried so hard to avoid. Actions in my model are *not* bounded below *on purpose.* Overcoming the technical difficulties created by this desire for a lack of *a priori* non-economic bounds, unlike the corresponding conditions in the Arrow–Debreu model which are well motivated, was one of the major goals of the paper.

In discussing his example C10.3, the discussant has lost sight of the competitive character of the model. If an individual knows that he will certainly be bankrupt, then, it is true that he can infer that $\tilde{r}_k < 1$ for all commodities in which he is a net debtor. But this gives him no reason to believe that $\tilde{r}_k$ will be any different for the *other* commodities and it is only these that are relevant to him. Thus, the discussant's comments in the second paragraph following this example seem to me completely irrelevant and misleading.

Professor Stigum says that I use a bankruptcy constraint that is inconsistent with the first-period budget constraint. This view is mistaken. There are two markets in period 1 – the market for period 1 commodities and that for contracts for claims on period 2 goods (with a bankruptcy clause in the contract). The consumer's excess demand in these two markets at date 1 is simply, $\boldsymbol{\omega}^1 + \boldsymbol{\eta}^1(\boldsymbol{p}, \boldsymbol{r}) - \boldsymbol{x}^1$ and $\boldsymbol{b}$, respectively. Hence the budget constraint, $\boldsymbol{p}^1\boldsymbol{x}^1 + \boldsymbol{p}^2\boldsymbol{b} \leq \boldsymbol{p}^1(\boldsymbol{\omega}^1 + \boldsymbol{\eta}^1(\boldsymbol{p}, \boldsymbol{r}))$. The quantities $\boldsymbol{\omega}^2$ and $\boldsymbol{\eta}^2(\boldsymbol{p}, \boldsymbol{r})$ have nothing to do with *market excess demands at date* 1 and therefore do not enter into the budget equation. They do play a role in determining the individual's wealth position and hence bankruptcy status, which we have argued above should be related in the indicated manner in an anonymous market model. Hence they appear in the bankruptcy relation.

If we were to follow Professor Stigum's suggestions at the foot of p. 293, we would be led to an inconsistency. Consider a consumer who desires simply to consume his endowment at all prices and for whom $\boldsymbol{\eta}(\boldsymbol{p}, \boldsymbol{r}) = 0$. If his budget constraints were (i) $\boldsymbol{p}^1\boldsymbol{x}^1 + \boldsymbol{p}^2\boldsymbol{b} \leq \boldsymbol{p} \cdot (\boldsymbol{\omega} + \boldsymbol{\eta}(\boldsymbol{p}, \boldsymbol{r}))$ in period 1, and (ii) $\tilde{\boldsymbol{p}}\tilde{\boldsymbol{x}} \leq \boldsymbol{p}(\tilde{\boldsymbol{r}}\boldsymbol{b}_+ + \boldsymbol{b}_-)$ in period 2, he would have to choose $\boldsymbol{x}^1 = \boldsymbol{\omega}^1$ and $\boldsymbol{b} = \boldsymbol{\omega}^2$, according to (i). But then his contracts held in period 2, $\boldsymbol{b}_+$, would be subject to the market rate default even though he traded with no-one!

My model tried to make a distinction between preexisting contracts with others, which are subject to default, and naturally occurring endowment which is certain. Adopting Professor Stigum's modifications would necessitate destroying this feature of the model, which I believe to be a desirable one.

I do not include a 'bankruptcy penalty' in period 1 because, from the individual's point of view, this is a datum and not a choice variable at the equilibrium prices. I never meant to imply that bankruptcy in period 1 is better than in period 2 – but only that, if it does not affect the period 1 excess demand function, we need not consider it explicitly. I did not want to crowd an already messy set of notations and hence did not make explicit reference to this utility loss.

Although I realize that alternative concepts of bankruptcy are possible, I have not been successful in finding any others which are internally consistent and preserve the economic phenomena that I tried to model.

CHAPTER 11

COMPETITIVE RESOURCE ALLOCATION OVER TIME UNDER UNCERTAINTY*

Bernt P. Stigum

11.1. Introduction

In this paper we study competitive resource allocation over time in an exchange economy which operates under uncertainty over infinitely many periods and which faces a tree structure of events with the property that, at each point in time, one of at most a finite number of events can occur. Like most real life economies, this one possesses in each period markets for current goods and for currently available securities (i.e. for contingent claims on future purchasing power) but not for contingent claims on future commodities. Our object is twofold:

(1) to ascertain whether there exists a family of temporary competitive equilibrium allocations of goods and securities – one allocation for each possible event – such that, as events occur, the economy can move along the resulting time path of equilibrium allocations without having to redistribute purchasing power in any period; and

(2) to determine whether or not the economy can achieve a Pareto-optimal allocation of resources over time.

The answers to these two queries have practical bearing on public policy. For example, pursuit of a vigorous anti-trust policy makes good sense in an economy only if the scale economies are small *and* if the

* The ideas of this paper were presented first in 'Optimal allocation of risk-bearing over time', Evanston, in April 1970, and in 'Competitive resource allocation over time under uncertainty', the Second World Congress of the Econometric Society, Cambridge, England, in September 1970. In writing this version of the paper we have benefited from discussions with members of the Mathematical Social Science Board Workshop on Uncertainty at Berkeley in Summer 1971, from helpful comments by members of the department of economics at the University of Texas, Austin, and most of all from constructive criticism of an earlier draft by Professor Michael Balch.

resulting economy of price-takers could function smoothly over time. In an economy like ours which possesses in each period only markets for current goods and securities, one obstacle to such smooth functioning would be the occurrence of bankruptcies. If consumers borrow and lend, chances are that in some periods no temporary equilibrium will exist at which all consumers would be solvent. Therefore, even though an imperfectly competitive economy need not perform well in this respect either, it is important to investigate whether or not a perfectly competitive economy could overcome the bankruptcy problem. In our answer to query one above, we explore such a possibility for an economy of the sort studied in this paper.

Even if a perfectly competitive economy could function smoothly without government intervention, this alone would not justify pursuing an activist anti-trust policy. In addition it must be shown that the pattern of resource allocation which such an economy achieves is in some sense optimal. It is well known that, for an economy which operates under uncertainty over denumerably many periods and which possesses a complete set of markets for current goods and for contingent claims on future goods, a competitive equilibrium is Pareto-optimal (cf. refs. [4, p. 102] and [16, p. 226]). The answer to our second query above concerns whether this result also holds for economies in which markets for contingent claims on future goods are almost non-existent.

Even if the economy can be shown to achieve a Pareto-optimal allocation of resources, this allocation might be judged undesirable by a quite reasonable social preference function; for example, because it allocated most resources to a few consumers. However, theory tells us that, for economies which operate under uncertainty and possess complete sets of markets for contingent claims on future commodities, any inequality in the distribution of resources that might arise at a given competitive equilibrium could be eliminated by a once-and-for-all lump-sum tax-subsidy scheme that would move the economy from the socially undesirable efficient point to a socially optimal one, a point moreover that could be sustained as a competitive equilibrium. Thus, for economies with enough contingent claims markets and only small economies of scale in production, an active anti-trust policy combined with an appropriate lump-sum tax-subsidy scheme could be used to establish the socially optimal allocation of resources in the economy. In answering our second query we will investigate whether such a combination of government policies could be counted on to be as effective in economies like the one studied in this paper.

In discussing the practical bearing of our model, we have focused on antitrust policy. However, we could just as well have talked about the desirability of free trade (cf. ref. [17]), or of using 'best' possible schemes for sharing the burden of providing the necessary public goods in a Pareto-optimal allocation of private and public goods (cf. refs. [6] and [10, pp. 123–140]). The two queries posed at the beginning of this paper thus have an important bearing on many aspects of public policy in a free enterprise economy.

Some of the results obtained in this paper are extensions of results obtained elsewhere.

(1) Sufficient conditions for the existence of temporary equilibria in a production economy in which each decision maker has a finite planning horizon and multi-valued price expectations were established in refs. [2], [14] and [15][1]. Theorem 11.1 below gives sufficient conditions that a temporary equilibrium exists when consumers have an infinite planning horizon and univalued price expectations with respect to the state of nature.

(2) Radner (cf. ref. [12]) has established for an exchange economy which operates under uncertainty over finitely many periods sufficient conditions that in the first period a temporary equilibrium exists at which consumers' price expectations and plans for future purchases of goods are mutually consistent in the sense that consumers' price expectations agree and are such that, at these prices, the planned supply of goods equals the planned demand for goods at each and every relevant future event. In the proof of theorem 11.2 we construct a first-period temporary equilibrium for our economy at which consumers' price expectations and plans for future purchases of goods and securities are mutually consistent.

(3) Arrow has shown in ref. [1] that a perfectly competitive economy which operates under uncertainty over a single period can achieve an optimal allocation of risk-bearing by trading in 'current' goods and

---

[1] The idea of a temporary equilibrium is due to J. R. Hicks. A discussion of its comparative statics properties under conditions of certainty can be found in ref. [9, chapters IX–XXII]. A very interesting discussion of the existence of temporary equilibria in a two-period exchange economy is given in ref. [7].

securities only. Theorem 11.3 extends Arrow's result to economies that operate under uncertainty over infinitely many periods.

Finally, we should point out that many authors (cf. for instance Borch [3], Diamond [5] and Sandmo [13]) have set out to answer questions similar to query (2) above for economies which operate under uncertainty over one or two periods and have markets for 'current' goods and securities only. These authors found that, if the economy did not possess 'enough' securities markets, it would, in general, not allocate resources Pareto-optimally. It is, therefore, important that the reader note that the economy we study in this paper, like Arrow's economy, has 'enough' securities markets! When we find that it still might not allocate resources Pareto-optimally, the reason is that the competitive resource allocation mechanism in an economy such as ours is essentially myopic.

11.2. Statement of Results

In this section we state our results concerning the way a competitive exchange economy in which there are a finite number of consumers would allocate resources over time under uncertainty. In the interests of brevity all proofs are relegated to the appendix.

We begin by specifying what we mean when we use the words 'time', 'states of nature' and 'current events'. We assume throughout that the world evolves in discrete epochs or periods $t \in \{\ldots, -1, 0, 1, \ldots\}$ and that we begin observing our model economy at time $t = 0$. How the world evolves can be described in many ways. For the purposes of this paper we can couch this description in terms of two concepts, a state of nature and a current event.

A *state of nature* is a description of the world that specifies for each and every period 'atmospheric conditions, natural disasters, technical possibilities, ...' [4, p. 98]. This definition is imprecise but suffices for our purpose. The important thing to note about it is that a state of nature prescribes consumers' initial endowments of goods (for example, 'technical possibilities') in each and every period, but gives no information about prices in different periods.

If S denotes the (mutually exclusive and collectively exhaustive) set of all states of nature, then an *event* is a subset of S, and a *t-current event*

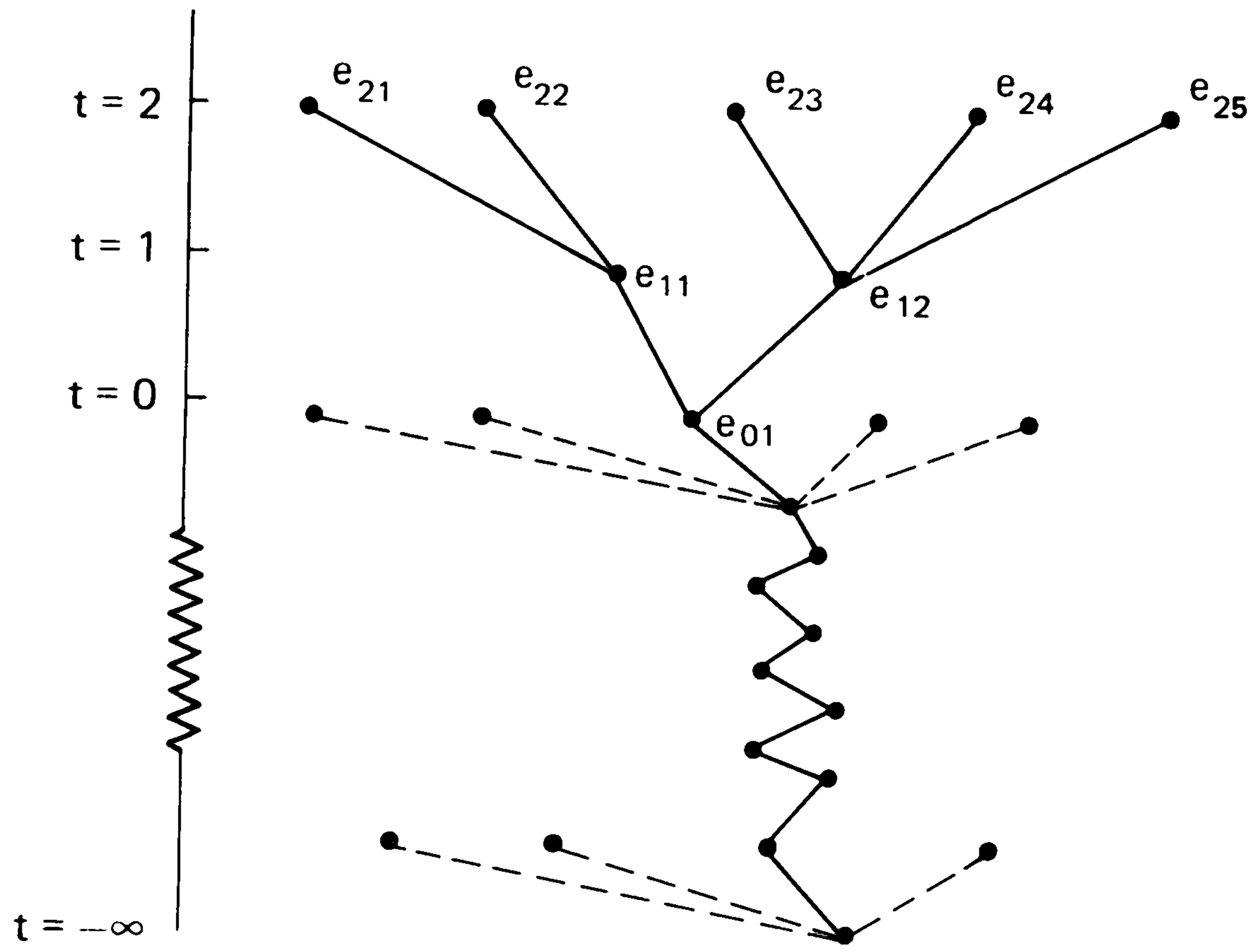

Fig. 11.1.

e_{ti} is a maximal subset of S whose member states share a common history up to and including epoch t. Alternatively, e_{ti} may be characterized as a description of the world that specifies for each and every period in $\{\ldots, t-1, t\}$ 'atmospheric conditions, natural disasters, technical possibilities, ...'.

We assume that our model economy at $t = 0$ has observed the 0-current event e_{01}, and that it faces a tree-structure of events which, like the tree in fig. 11.1, has the temporal 'successor' property that only a finite number of $(t+1)$-current events $e_{(t+1)j}$ are contained in the t-current event e_{ti}. The total number of events that could occur in period t given that e_{01} has occurred will be denoted by r_t.

Next a word about the institutions. Let $\boldsymbol{q}(e_{ti}) = (q_1, \ldots, q_m)(e_{ti})$ denote a vector each component of which represents the quantity of a commodity that would be available to consumers in period t if the event e_{ti} were to occur. Moreover, let $V(e_{ti})$ be the quantity of a one-period security issued in period $(t-1)$ each unit of which will pay one 'dollar' (the unit of account) in period t if the event e_{ti} occurs and nothing otherwise. Finally, for each pair (t, i), let

$$N[ti] \equiv \{j : e_{(t+1)j} \subset e_{ti}\}. \tag{11.1}$$

Then we assume:

(1) In each period t, if the event e_{ti} occurs, markets will exist for $\boldsymbol{q}(e_{ti})$ and for every $V(e_{(t+1)j})$ such that $j \in N[ti]$. Moreover, each con-

sumer will enter into irrevocable contracts with respect to $\boldsymbol{q}(e_{ti})$ and the $V(e_{(t+1)j})$s while he simultaneously makes *plans* for buying and selling goods and securities in future periods.

Now let us consider consumer behavior. To state our assumptions concerning it, we need still more notation. Let $\boldsymbol{p}(e_{ti})$ denote the price (vector) of $\boldsymbol{q}(e_{ti})$, let $\boldsymbol{\beta}(e_{ti})$ denote the price of $V(e_{ti})$, and let

$$(\boldsymbol{q}, V)(e_{ti}) \equiv (\boldsymbol{q}(e_{ti}), V(e_{(t+1)j}) : j \in N[ti]),^2 \tag{11.2}$$

$$(\boldsymbol{p}, \boldsymbol{\beta})(e_{ti}) \equiv (\boldsymbol{p}(e_{ti}), \boldsymbol{\beta}(e_{(t+1)j}) : j \in N[ti]), \tag{11.3}$$

$$\boldsymbol{q}(e^{ti}) \equiv (\boldsymbol{q}(e_{sj}) : s > t,\ e_{sj} \subset e_{ti}),^3 \tag{11.4}$$

$$(\boldsymbol{q}, V)(e^{ti}) \equiv ((\boldsymbol{q}, V)(e_{sj}) : s > t,\ e_{sj} \subset e_{ti}), \tag{11.5}$$

$$(\boldsymbol{p}, \boldsymbol{\beta})(e^{ti}) \equiv ((\boldsymbol{p}, \boldsymbol{\beta})(e_{sj}) : s > t,\ e_{sj} \subset e_{ti}). \tag{11.6}$$

In reading these definitions observe the notational 'time shift' difference between $V(e_{ti})$ and $(\cdot, V)(e_{ti})$. $V(e_{ti})$ denotes a certain quantity of a security which was sold in period $(t-1)$ and which matures in period t. The components of $(\boldsymbol{q}, V)(e_{ti})$ represent the commodities and securities that can be bought in period t if e_{ti} occurs. The security-components of this vector mature in period $(t+1)$. Then we assume:

(2) There are $r \geq 1$ consumers in the economy. Each consumer k possesses in the first period initial quantities of commodities $\bar{\boldsymbol{q}}^k(e_{01})$ and – contingent on the events that occur – is certain to receive initial endowments $\bar{\boldsymbol{q}}^k(e^{01})$ in future periods. Moreover,

$$(\bar{\boldsymbol{q}}^k(e_{01}), \bar{\boldsymbol{q}}^k(e^{01})) \geq 0,\ (\bar{\boldsymbol{q}}^k(e_{01}), \bar{\boldsymbol{q}}^k(e^{01})) \neq 0,\ k = 1, \ldots, r, \tag{11.7}$$

and

$$\sum_{k=1}^{r} (\bar{\boldsymbol{q}}^k(e_{01}), \bar{\boldsymbol{q}}^k(e^{01})) > 0. \tag{11.8}$$

---

2 The expression $(\boldsymbol{q}(e_{ti}), V(e_{(t+1)j}) : j \in N[ti])$ is shorthand for the vector whose components represent *all* the commodities and securities that could be bought in period t if e_{ti} occurred.

3 The expression $(\boldsymbol{q}(e_{sj}) : s > t, e_{sj} \subset e_{ti})$ is shorthand for the infinite-dimensional commodity-vector corresponding to the branch of the tree-structure of events which starts at e_{ti}.

Finally, each consumer enters the first period debt-free and (*a fortiori*) owns no units of $V(e_{01})$.

(3) (*a*) In every period t, regardless of which event e_{ti} occurs, every consumer possesses well-defined, univalued expectations as to future commodity and security prices. These we denote in the following way. Let $(\boldsymbol{p}, \boldsymbol{\beta})^{e[\tau]}(e_{sj})$ be the expected price vector, contingent on e_{sj}, as perceived by the individual consumer looking forward from $\tau < s$. Then the consumer's price expectations at e_{ti} are given by

$$(\boldsymbol{p}, \boldsymbol{\beta})^{e[t]}(e^{ti}) = ((\boldsymbol{p}, \boldsymbol{\beta})^{e[t]}(e_{sj})\colon s > t,\ e_{sj} \subset e_{ti}). \tag{11.9}$$

(*b*) So long as expectations have always been historically 'validated', looking backward from t, they will continue invariant into the future; i.e. if the observed prices

$$(\boldsymbol{p}, \boldsymbol{\beta})(e_{si}) = (\boldsymbol{p}, \boldsymbol{\beta})^{e[s-1]}(e_{si}),$$

for all $s \leq t$, then for any $v < t < u$ and $e_{vj} \supset e_{ti} \supset e_{uk}$, the $(\boldsymbol{p}, \boldsymbol{\beta})^{e[\cdot]}(e_{uk})$ components of $(\boldsymbol{p}, \boldsymbol{\beta})^{e[v]}(e^{vj})$ and $(\boldsymbol{p}, \boldsymbol{\beta})^{e[t]}(e^{ti})$ are the same.

In the sequel we will suppress the 'bracket' in $e[\cdot]$ and write $(\boldsymbol{p}, \boldsymbol{\beta})^{e}(\cdot)$ for $(\boldsymbol{p}, \boldsymbol{\beta})^{e[\cdot]}(\cdot)$. This will simplify our notation and should not cause any unnecessary confusion.

(4) In each period t if the event e_{ti} occurs, consumer k will order $(\boldsymbol{q}(e_{ti}), \boldsymbol{q}(e^{ti}))$-vectors according to the values assumed by a function $U_k^{ti}({}^{ti}\boldsymbol{q}, \cdot)$, where ${}^{ti}\boldsymbol{q}$ is the vector he *actually* consumed during the first t periods[4].

The utility functions $U_k^{ti}(\cdot)$ have certain structural properties and are (for each k) related in a definite way, which can be described as follows. Let

$$C \equiv \{\boldsymbol{q} = (\boldsymbol{q}(e_{01}), \boldsymbol{q}(e^{01}))\colon 0 \leq \boldsymbol{q} < \infty\};^{5} \tag{11.10}$$

---

[4] Strictly speaking the ${}^{ti}\boldsymbol{q}$-vector should contain information about consumption in periods prior to $t = 0$. Since this information would not play a part in the discussion which follows, we can without loss in generality ignore it here and in the remainder of the paper.

[5] Throughout this paper the symbols $\geq$ and $>$ are used as follows: if $\boldsymbol{x}$ and $\boldsymbol{y}$ are n-dimensional vectors, $\boldsymbol{x} \geq \boldsymbol{y}$ if and only if $x_i \geq y_i$ for all $i = 1, \ldots, n$. Similarly $\boldsymbol{x} > \boldsymbol{y}$ if and only if $x_i > y_i$ for all $i = 1, \ldots, n$. Finally, if A is a number, then $0 \leq \boldsymbol{x} \leq A$ if and only if $0 \leq x_i \leq A$ for all $i = 1, \ldots, n$. The same rules hold for infinite-dimensional vectors as well. Note that the symbol 0 is used both to denote zero and a vector whose components all equal zero.

let $\boldsymbol{x}_t \equiv (x_{t1}, \ldots, x_{tm})$; let $\boldsymbol{x} \equiv (\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots)$; let $||\boldsymbol{x}_t|| \equiv \sum_{i=1}^m |x_{ti}|$; and let

$$X \equiv \{\boldsymbol{x}: 0 \leq \boldsymbol{x} < \infty\}. \tag{11.11}$$

Moreover, let X be endowed with the product topology, and let $\boldsymbol{x}(\cdot) \equiv (\boldsymbol{x}_0(\cdot), \boldsymbol{x}_1(\cdot), \ldots)$ denote a function on S to X. We say that $\boldsymbol{x}(\cdot)$ is *feasible* (or *historically consistent*) if and only if, for each and every pair (t, i), $(x_0(\cdot), \ldots, x_t(\cdot))$ is constant on e_{ti}, and if $\boldsymbol{x}(s) = 0$ for all s in the complement of e_{01}. Finally for each $\boldsymbol{q} \in C$, let $\boldsymbol{x}_q(\cdot)$ be the feasible function on S to X whose tth component vector $(\boldsymbol{x}_q)_t(\cdot)$ takes the value $\boldsymbol{q}(e_{ti})$ on e_{ti}, for all $t \geq 0$ and all $i = 1, \ldots, r_t$. Then we assume:

(5) For each consumer k there exists a bounded, continuous, strictly concave, increasing function on X, $W_k(\cdot)$, and a subjective probability measure $\boldsymbol{Q}_k(\mathrm{d}s)$ on subsets of S such that $W_k(0) = 0$ and such that

$$\boldsymbol{Q}_k(e_{01}) = 1, \tag{11.12}$$

$$\boldsymbol{Q}_k(e_{ti}) > 0,\ t > 0,\ i = 1, \ldots, r_t, \tag{11.13}$$

$$U_k^{01}(q) = \int_S W_k(\boldsymbol{x}_q(s))\boldsymbol{Q}_k(\mathrm{d}s). \tag{11.14}$$

Moreover, for each pair (t, i) and actually chosen vector ${}^{ti}\boldsymbol{q}^*$,

$$U_k^{ti}({}^{ti}\boldsymbol{q}^*, \cdot) = (1/\boldsymbol{Q}_k(e_{ti}))\int_{e_{ti}} W_k(\boldsymbol{x}_{({}^{ti}q^*,\cdot)}(s))\boldsymbol{Q}_k(\mathrm{d}s).^6 \tag{11.15}$$

Of the preceding assumptions, assumption (2) is standard (except for $V(e_{01}) = 0$, about which we shall have more to say below), and assumptions (4) and (5) – if not exactly standard – ought to be acceptable to most economists. Assumption (3) is unrealistic since most consumers in real life probably entertain price expectations that are multivalued with respect to the state of nature. Note, therefore, that we make this assumption *not* because it is essential in establishing our results since the existence of temporary equilibria can be established without it (cf. ref. [14], p. 551) and since it is not really used in theorems 11.2 and 11.3. We make this assumption to simplify notation and the statement of results and to thereby facilitate the reader's task in singling out the important ideas of the paper.

---

[6] Here $x_{(ti,*,\cdot)}(\cdot)$ is a function on e_{ti} to X such that

$$[(x_{(ti,*,\cdot)})_0(s), \ldots, (x_{(ti,*,\cdot)})_{t-1}(s)] = {}^{ti}\boldsymbol{q}^* \qquad \text{for all } s \in e_{ti}.$$

Besides assumptions (1)–(5) we also make several implicit assumptions: (a) each consumer has an infinitely long planning horizon; (b) each consumer lives forever; and (c) the number of consumers in the economy is constant over time. The last two assumptions can be made realistic by interpreting a consumer as a family spending unit that remains intact over time but whose members may vary in number from period to period.

Before stating our results, we must introduce the budget correspondence faced by each consumer in period t, and also state precisely what we mean by a temporary equilibrium. If consumer k in period $(t-1)$ purchased (or sold) $V^k(e_{ti})$ units of $V(e_{ti})$ and observes the event e_{ti} and prices $(\boldsymbol{p}, \boldsymbol{\beta})(e_{ti})$, then his t-period budget constraint is

$$(\boldsymbol{p}, \boldsymbol{\beta})(e_{ti})(\boldsymbol{q}, V)(e_{ti}) \leq \boldsymbol{p}(e_{ti})\bar{\boldsymbol{q}}^k(e_{ti}) + V^k(e_{ti}),\ \boldsymbol{q}(e_{ti}) \geq 0.^7 \quad (11.16)$$

If consumer k in period t plans to buy (or sell) $V^k(e_{sj})$ units of $V(e_{sj})$ in period $(s-1) \geq t$, then the budget constraint which he expects to face if e_{sj} occurs is

$$(\boldsymbol{p}, \boldsymbol{\beta})^e(e_{sj})(\boldsymbol{q}, V)(e_{sj}) \leq \boldsymbol{p}^e(e_{sj})\bar{\boldsymbol{q}}^k(e_{sj}) + V^k(e_{sj}),\ \boldsymbol{q}(e_{sj}) \geq 0.^8 \quad (11.17)$$

Consequently, if we define $\Gamma(\cdot)$ by

$$\begin{aligned}&\Gamma((\boldsymbol{p}, \boldsymbol{\beta})(e_{ti}), (\boldsymbol{p}, \boldsymbol{\beta})^e(e^{ti}), \bar{\boldsymbol{q}}(e_{ti}), \bar{\boldsymbol{q}}(e^{ti}), V(e_{ti})) \equiv \\ &\{((\boldsymbol{q}, V)(e_{ti}), (\boldsymbol{q}, V)(e^{ti})): (\boldsymbol{q}(e_{ti}), \boldsymbol{q}(e^{ti})) \geq 0, (\boldsymbol{p}, \boldsymbol{\beta})(e_{ti})(\boldsymbol{q}, V)(e_{ti}) \leq \\ &\qquad \boldsymbol{p}(e_{ti})\bar{\boldsymbol{q}}(e_{ti}) + V(e_{ti}); (\boldsymbol{p}, \boldsymbol{\beta})^e(e_{sj})(\boldsymbol{q}, V)(e_{sj}) \leq \\ &\qquad \boldsymbol{p}^e(e_{sj})\bar{\boldsymbol{q}}(e_{sj}) + V(e_{sj}),\ s > t,\ e_{sj} \subset e_{ti}\}, \quad (11.18)\end{aligned}$$

then $\Gamma(\cdot)$ is the budget correspondence faced by each and every consumer in the economy in period t if e_{ti} obtains.

DEFINITION 11.1. Suppose that e_{ti} has occurred, and that consumer k, $k = 1, \ldots, r$, consumed ${}^{ti}\boldsymbol{q}^{*k}$ in the past periods, 'owns' $V^k(e_{ti})$ units of $V(e_{ti})$, and entertains the price expectations $(\boldsymbol{p}, \boldsymbol{\beta})^e(e^{ti})$. Then

$$((\boldsymbol{p}, \boldsymbol{\beta})(e_{ti}), (\boldsymbol{q}^*, V^*)^1(e_{ti}), \ldots, (\boldsymbol{q}^*, V^*)^r(e_{ti}))$$

is a temporary equilibrium in period t if

$$(\boldsymbol{p}, \boldsymbol{\beta})(e_{ti}) > 0; \quad (11.19)$$

---

7 If $\boldsymbol{p}$ and $\boldsymbol{q}$ are vectors, then $\boldsymbol{pq}$ denotes the inner product of $\boldsymbol{p}$ and $\boldsymbol{q}$.

8 Here $(\boldsymbol{p}, \boldsymbol{\beta})^e(e_{sj})$ and $\boldsymbol{p}^e(e_{sj})$ are short for $(\boldsymbol{p}, \boldsymbol{\beta})^{e[t]}(e_{sj})$ and $\boldsymbol{p}^{e[t]}(e_{sj})$ respectively.

$$\sum_{k=1}^{r} (\boldsymbol{q}^{*k}(e_{ti}) - \bar{\boldsymbol{q}}^k(e_{ti})) = 0; \tag{11.20}$$

$$\sum_{k=1}^{r} V^{*k}(e_{(t+1)j}) = 0, j \in N[ti]; \tag{11.21}$$

and if there exist feasible plans $(\boldsymbol{q}^*, V^*)^k(e^{ti})$, $k = 1, \ldots, r$, so that for all k

$$((\boldsymbol{q}^*, V^*)^k(e_{ti}), (\boldsymbol{q}^*, V^*)^k(e^{ti})) \in \Gamma_k, \tag{11.22}$$

and

$$U_k^{ti}({}^{ti}\boldsymbol{q}^{*k}, \boldsymbol{q}^*(e_{ti}), \boldsymbol{q}^*(e^{ti})) \geq U_k^{ti}({}^{ti}\boldsymbol{q}^{*k}, \boldsymbol{q}(e_{ti}), \boldsymbol{q}(e^{ti})) \tag{11.23}$$

for all pairs $(\boldsymbol{q}(e_{ti}), \boldsymbol{q}(e^{ti}))$ for which there exists a sequence of investments $V(e_{sj})$, $s > t$, $e_{sj} \subset e_{ti}$ such that

$$((\boldsymbol{q}, V)(e_{ti}), (\boldsymbol{q}, V)(e^{ti})) \in \Gamma_k, \tag{11.24}$$

where

$$\Gamma_k \equiv \Gamma((\boldsymbol{p}, \boldsymbol{\beta})(e_{ti}), (\boldsymbol{p}, \boldsymbol{\beta})^e(e^{ti}), \bar{\boldsymbol{q}}^k(e_{ti}), \bar{\boldsymbol{q}}^k(e^{ti}), V^k(e_{ti})). \tag{11.25}$$

There are several things to note about the preceding definition. First, we have omitted a superscript k in $(\boldsymbol{p}, \boldsymbol{\beta})^e(e^{ti})$ to simplify notation. Therefore, the definition should not be taken to mean that in a temporary equilibrium consumers necessarily entertain identical price expectations.

Second, the definition requires that the plans $(\boldsymbol{q}^*, V^*)^k(e^{ti})$, $k = 1, \ldots, r$, satisfy eqs. (11.22) and (11.23). However, it does not require that these plans be mutually consistent in the sense that

$$\sum_{k=1}^{r} (\boldsymbol{q}^*, V^*)^k(e^{ti}) = \left(\sum_{k=1}^{r} \bar{\boldsymbol{q}}^k(e^{ti}), 0\right). \tag{11.26}$$

Thus while supply and demand for commodities and securities must be equal in period t (cf. eqs. (11.20)–(11.21)) for a temporary equilibrium to occur, consumers are free to plan for the future without any thought as to whether these plans could actually be carried out (i.e. whether their plans are mutually consistent). This is the reason why we say that the competitive resource-allocation mechanism in our economy is essentially myopic[9].

[9] A discussion of this kind of myopic behavior is given in ref. [9, chapter X, pp. 130–140].

Whether or not a temporary equilibrium exists in any given period depends on consumer preferences and price expectations, and on consumers' initial endowments of goods and securities. Theorem 11.1 gives sufficient conditions on the consumers' price expectations that a competitive equilibrium exists in the first period.

THEOREM 11.1. *Suppose that assumptions* (1)–(5) *hold, and for all* $t = 2, 3, \ldots$, *and* $i = 1, \ldots, r_t$, *let*

$$\tilde{\boldsymbol{p}}(e_{ti}) = [\prod_{\substack{e_{ti} \subset e_{sj} \\ s = 2, \ldots, t}} \beta^e(e_{sj})]\boldsymbol{p}^e(e_{ti}), \tag{11.27}$$

and

$$\tilde{\boldsymbol{p}}(e_{1j}) = \boldsymbol{p}^e(e_{1j}), j \in N[01]. \tag{11.28}$$

If

$$\tilde{\boldsymbol{p}}(e^{01})\bar{\boldsymbol{q}}^k(e^{01}) < \infty, k = 1, \ldots, r, \tag{11.29}$$

and

$$\tilde{\boldsymbol{p}}(e_{ti}) > 0 \text{ for all } t = 1, 2, \ldots; i = 1, \ldots, r_t, \tag{11.30}$$

then there exists a temporary equilibrium

$$((\boldsymbol{p}, \boldsymbol{\beta})(e_{01}), (\boldsymbol{q}, \boldsymbol{V})^1(e_{01}), \ldots, (\boldsymbol{q}, \boldsymbol{V})^r(e_{01}))$$

in period 0.

In theorem 11.1, as in definition 11.1, we do not require consumers to share the price expectations $(\boldsymbol{p}, \boldsymbol{\beta})^e(e^{ti})$. We have omitted the superscript k for simplicity only. Note, however, that we do assume that $\boldsymbol{V}^k(e_{01}) = 0$ for all $k = 1, \ldots, r$. Thus there is no possibility that one or more consumers might go bankrupt in period 0. Finally, note that the assumptions (11.29)–(11.30) imply that the set of future allocations which satisfy consumer k's budget constraint is compact.

The importance of the assumption $\boldsymbol{V}^k(e_{01}) = 0$, $k = 1, \ldots, r$ can be seen from the following simple example.

EXAMPLE 11.1. Consider a two-period economy in a world in which there is one and only one state of nature. There are two consumers, A and B, one first-period commodity x, and two second-period commodities y and z. Consumer A orders triples (x, y, z) according to the values assumed by the function

$$U_A(x, y, z) \equiv \sqrt{(xy)} + z, (x, y, z) \geq 0, \tag{11.31}$$

and possesses the initial endowments $(\bar{x}_A, \bar{y}, 0)$. Consumer B orders triples (x, y, z) according the values assumed by the function

$$U_B(x, y, z) \equiv \sqrt{(x)} - e^{-\sqrt{(yz)}},\ (x, y, z) \geq 0, \tag{11.32}$$

and possesses the initial endowments $(\bar{x}_B, 0, \bar{z})$. Both A and B expect that prices of y and z in period 2 will be as follows:

$$(p_y^e, p_z^e) = (1, 1). \tag{11.33}$$

Finally, A starts out in period 1 with a debt of $\tilde{x}_A$ units of x to B, where $\tilde{x}_A > \bar{x}_A$.

During period 1 there are markets for x and for a security V each unit of which will pay one unit of the unit of account in period 2. During period 2 there are markets for y and z. To determine how A and B behave in the first-period market, we define two functions $F_A(\cdot)$ and $F_B(\cdot)$ as follows:

$$F_A(x, V) \equiv \max_{\{(y,z) \geq 0, y+z \leq \bar{y}+V\}} \sqrt{(xy)} + z$$
$$= \begin{cases} (x/4) + \bar{y} + V \text{ for } V \geq -\bar{y},\ x \leq 4(\bar{y} + V), \\ \sqrt{(x)}\sqrt{(\bar{y} + V)} \text{ for } V \geq -\bar{y},\ x > 4(\bar{y} + V), \end{cases} \tag{11.34}$$

$$F_B(x, V) \equiv \max_{\{(y,z) \geq 0, y+z \leq \bar{z}+V\}} \sqrt{(x)} - e^{-\sqrt{(yz)}} = \sqrt{(x)} - e^{-(\bar{z}+V)/2},$$
$$x \geq 0,\ V \geq -\bar{z}. \tag{11.35}$$

It is easy to verify that, to maximize their utility over two periods, A and B must in period 1 choose pairs (x_A, V_A) and (x_B, V_B) which respectively maximize $F_A(\cdot)$ and $F_B(\cdot)$ subject to the first-period budget constraint faced by each consumer.

Can consumers A and B arrive at a trading position that can be sustained as a temporary equilibrium? This depends on the values assumed by $(\bar{x}_A, \bar{y})$, $(\bar{x}_B, \bar{z})$, and $\tilde{x}_A$. If

$$\bar{x}_A = \bar{x}_B = \bar{y} = \bar{z} = 2, \tag{11.36}$$

the following assertions can be easily verified:

(1) If $\tilde{x}_A = 2.52$, then the vector

$$(p_x, \boldsymbol{\beta}, x_A, V_A, x_B, V_B) = (1, 0.351, 0, -1.48, 4, 1.48) \tag{11.37}$$

is a temporary equilibrium in period 1.

(2) If $\tilde{x}_A = 2.98$, then there is no set of prices $(\boldsymbol{p}_x, \boldsymbol{\beta})$ at which B would be willing to lend A enough so that he could settle his first period debt. This is true moreover regardless of the value of $\bar{y}$, i.e. regardless of A's future financial resources.

(3) If $\tilde{x}_A = 0$, and if instead B owes A 10.4 units of x, then at each and every set of prices $(\boldsymbol{p}, \boldsymbol{\beta})$ at which A would be willing to lend B enough to settle his current debt, B could not accept A's offer because it would mean that he would be unable to pay back the 'new' debt during period 2. For instance, at $(\boldsymbol{p}_x, \boldsymbol{\beta}) = (0.3, 1)$ A would offer 12.4 units of x in exchange for 3.72 units of V. But $-3.72 < -2$ so B would not be able to accept the offer.

From the preceding example it follows that, if the economy in one period should arrive at a temporary equilibrium at which one consumer borrows funds from another, the next period it may fail to establish a temporary equilibrium. The example also shows that whether the economy in a given period will succeed in establishing an equilibrium without anyone being bankrupt depends not just on the period's distribution of initial endowments of commodities and securities and the future earning power of the individual debtors, but also on the risk preferences of the individual creditors[10]. Finally, as the example indicates, in our economy as in real life a consumer will not be forced into bankruptcy in a given period just because his net worth is negative. For instance, in the temporary equilibrium defined by eq. (11.37), consumer A's first-period net worth is negative ($= -0.52$). Yet he is able to borrow enough to settle his currently maturing debt. And, if we adjust the price of y so that it equals one unit of the unit of account, then A will most certainly be able to pay back his newly acquired debt next period.

Theorem 11.1 establishes the existence of a temporary equilibrium in the first period if consumer price expectations satisfy conditions (11.29)–(11.30). However, it does not imply that the economy will necessarily function smoothly over time. In theorem 11.2 we will show that, if the price expectations of consumers are 'just right', then (at least in theory) the economy can function smoothly. But first a definition.

---

[10] In ref. [15, theorem 2] we give sufficient conditions on consumer and entrepreneurial preferences that a production economy in a given period would possess a temporary equilibrium even if the initial distribution of resources were such that some spending units would be insolvent at some sets of positive prices that might occur.

DEFINITION 11.2. We say that the family of vectors

$$[((\boldsymbol{p}, \boldsymbol{\beta})(e_{ti}), (\boldsymbol{q}, V)^1(e_{ti}), \ldots, (\boldsymbol{q}, V)^r(e_{ti})), t \geq 0, i = 1, \ldots, r_t]$$

is a feasible tree-structure of temporary equilibria if and only if for all pairs (t, i),

$$((\boldsymbol{p}, \boldsymbol{\beta})(e_{ti}), (\boldsymbol{q}, V)^1(e_{ti}), \ldots, (\boldsymbol{q}, V)^r(e_{ti}))$$

is a temporary equilibrium relative to the distribution of purchasing power

$$(\boldsymbol{p}(e_{ti})\bar{\boldsymbol{q}}^1(e_{ti}) + V^1(e_{ti}), \ldots, \boldsymbol{p}(e_{ti})\bar{\boldsymbol{q}}^r(e_{ti}) + V^r(e_{ti})), \qquad k = 1, \ldots, r.$$

The important thing to notice about this definition is that a tree structure of competitive equilibria will be feasible only if the economy can move along each 'branch' of it, from one equilibrium to the next, without purchasing power being redistributed in any period.

Now the theorem.

THEOREM 11.2. *If assumptions* (1)–(5) *hold, we can find a set of price expectations* $(\boldsymbol{p}, \boldsymbol{\beta})^e(e^{01})$ *which, if shared by all consumers, ensures the existence of a feasible tree-structure of temporary equilibria.*

This theorem requires several comments. First note that the validity of the theorem depends on the assumption

$$V^k(e_{01}) = 0, k = 1, \ldots, r. \tag{11.38}$$

This assumption, however, can be weakened considerably. For instance, suppose that the unit of account in period 1 is a fictitious commodity which does not correspond to any one of the components of $(\boldsymbol{q}(e_{01}), \boldsymbol{q}(e^{01}))$. Then eq. (11.38) can be replaced by the assumption: there exists a triple (t, i, j) such that $e_{ti} \subset e_{01}$, and such that

$$\bar{\boldsymbol{q}}_j^k(e_{ti}) > 0, k = 1, \ldots, r. \tag{11.39}$$

Next, note that it will be seen that the price expectations constructed in the proof of theorem 11.2 satisfy the conditions (11.29)–(11.30). In addition they have the following two properties: (1) the plans of consumers for future acquisitions of goods and securities, associated with each and every temporary equilibrium of the tree structure, are mutually consistent; and (2) at each event that might occur, consumers' price expectations are 'validated' (and their plans are actually carried out). These two properties are not intrinsic to feasible tree-structures of temporary equilibria. It is perfectly possible to have a feasible tree-structure of temporary equilibria with the property that consumers' price expectations agree but consumers' plans are inconsistent at each and every event that might occur. Along such a tree structure plans must be revised in each period, and consumers' price expectations are rarely validated.

Finally, note that theorem 11.2 shows that the answer to our first query is yes. Thus our economy can in theory at least function smoothly. Unfortunately such 'smooth sailing' is not a characteristic of the economy that can be taken for granted. Once on a feasible tree structure, the economy might not be able to stay on it for long. And if it veers off, it might not be able to move on to another feasible tree-structure. Certainly, if it does not, the 'smooth sailing' will halt abruptly.

There is a simple reason for the instability of a feasible tree-structure of competitive equilibria: for it to be stable, each and every temporary equilibrium on it must be globally stable – an obviously impossible requirement. Here is an example to help the reader's intuition.

EXAMPLE 11.2. Consider a two-period economy in a world with one and only one state of nature. There are two consumers A and B, one first-period commodity x and two second-period commodities y and z. A orders triples (x, y, z) according to the values assumed by the function

$$U_A(x, y, z) \equiv -(1/3)e^{-3x} - e^{-(5+2\sqrt{(yz)})}, (x, y, z) \geq 0, \quad (11.40)$$

and possesses the initial endowments $(\bar{x}_A, \bar{y}_A, \bar{z}_A) = (2.14/3, 1.14, 1)$. B orders triples (x, y, z) according to the values assumed by the function

$$U_B(x, y, z) \equiv -e^{-(5+x)} - (1/3)e^{-6\sqrt{(yz)}}, (x, y, z) \geq 0, \quad (11.41)$$

and possesses the initial endowments $(\bar{x}_B, \bar{y}_B, \bar{z}_B) = (3, 1, 0)$. Both A and B expect that prices of y and z in the next period will be as follows:

$$(p_y^e, p_z^e) = (1, 1). \quad (11.42)$$

Finally, neither A nor B has any outstanding debt in the first period.

During the first period there are markets for x and for a security V, each unit of which will pay one unit of z the next period. During the second period there are markets for y and z. The 'indirect utility functions' of A and B during the first period are defined by

$$F_A(x, V) \equiv \max_{\{(y,z)\geq 0, y+z\leq V+2.14\}} [-(1/3)e^{-3x} - e^{-(5+2\sqrt{(yz)})}] =$$
$$= -(1/3)e^{-3x} - e^{-(7.14+V)}, \; x \geq 0, \; V \geq -2.14, \quad (11.43)$$

and

$$F_B(x, V) \equiv \max_{\{(y,z)\geq 0, y+z\leq V+1\}} [-e^{-(5+x)} - (1/3)e^{-6\sqrt{(yz)}}]$$
$$= -e^{-(5+x)} - (1/3)e^{-3(1+V)}, \; x \geq 0, \; V \geq -1. \quad (11.44)$$

To maximize their utility over two periods, A and B must in the first period choose pairs (x_A, V_A) and (x_B, V_B) which respectively maximize $F_A(\cdot)$ and $F_B(\cdot)$ subject to the first-period budget constraint faced by each consumer.

For this economy, the following assertions are true:

(1) There are three possible temporary equilibria in the first period:

$$TE_1: (p_x, \beta, x_A, V_A, x_B, V_B) = (1, 0.169, 1.075, -2.14, 2.64, 2.14), \quad (11.45)$$

$$TE_2: (p_x, \beta, x_A, V_A, x_B, V_B) = (1, 1, 1.963, -1.25, 1.75, 1.25), \quad (11.46)$$

and

$$TE_3: (p_x, \beta, x_A, V_A, x_B, V_B) = (1, 5.944, 2.854, -0.36, 0.86, 0.36). \quad (11.47)$$

Of these TE_1 and TE_3 are locally stable, while TE_2 is unstable in the usual sense of these terms[11].

(2) If TE_1 obtained during the first period, the next period A and B would not be able to arrive at a temporary equilibrium where A could repay his debt to B. Consequently if TE_1 obtained, A would be forced to declare bankruptcy during the second period.

(3) If TE_2 obtained in the first period, then in period 2,

$$TE_2^*: (p_y, p_z, y_A, z_A, y_B, z_B) = ((1/2.14), 1, 0.30, 0.14, 1.84, 0.86) \quad (11.48)$$

would be the one and only one temporary equilibrium relative to the distribution of purchasing power determined by $(\bar{y}_A, \bar{z}_A + V_A) = (1.14,$

[11] A similar example for a one-period exchange economy which operates under conditions of certainty with respect to the state of nature has been constructed by L. Hurwicz (cf. ref. [11], pp. 45–48).

-0.25) and $(\bar{y}_B, \bar{z}_B + V_B) = (1, 1.25)$. The pair ($TE_2$, TE_2^*) can be characterized as an unstable feasible tree-structure of temporary equilibria.

(4) If TE_3 obtained during the first period, then in period 2,

$$TE_3^*: (p_y, p_z, y_A, z_A, y_B, z_B) = ((1/2.14), 1, 1.25, 0.59, 0.89, 0.41) \quad (11.49)$$

would be the one and only one temporary equilibrium relative to the distribution of purchasing power determined by $(\bar{y}_A, \bar{z}_A + V_A) = (1.14, 0.64)$ and $(\bar{y}_B, \bar{z}_B + V_B) = (1, 0.36)$. The pair ($TE_3$, TE_3^*) is a locally stable feasible tree-structure of temporary equilibria.

To continue our discussion of feasible tree-structures of temporary equilibria, we must first define what we mean by a Pareto-optimal allocation of resources over time.

DEFINITION 11.3. Let

$$A = \{(\boldsymbol{q}^1, \ldots, \boldsymbol{q}^r): \boldsymbol{q}^k \in C, k = 1, \ldots, r, \sum_{k=1}^{r} (\boldsymbol{q}^k - \bar{\boldsymbol{q}}^k) = 0\}. \quad (11.50)$$

Then the set

$$P = \{(\boldsymbol{q}^1, \ldots, \boldsymbol{q}^r) \in A: \exists \text{ no } z \in A \text{ such that } U_k^{01}(z^k) \geq U_k^{01}(\boldsymbol{q}^k) \text{ for all } k \text{ with } U_i^{01}(z^i) > U_i^{01}(\boldsymbol{q}^i) \text{ for some } i\} \quad (11.51)$$

is the set of all Pareto-optimal allocations of commodities over time.

In interpreting this definition, note that $U_k^{01}(\cdot)$ describes how the kth consumer in the *first* period orders bundles of current and future goods. Hence P, strictly speaking, represents the set of all Pareto-optimal *first-period* allocations of current goods and contingent claims on future goods. But if that is so, then one might ask: suppose that in the first period the economy's resources (current and future) have been allocated according to a plan prescribed by a vector in P. Is it still possible that in some future period consumers might find that their preferences have changed and there exist trades that would benefit some without hurting others? The answer is no! Specifically, if $(\boldsymbol{q}^{*1}, \ldots, \boldsymbol{q}^{*r}) \in P$, if the event e_{ti} occurs, and if the vector $({}^{ti}\boldsymbol{q}^{*1}, \ldots, {}^{ti}\boldsymbol{q}^{*r})$ was consumed in periods $s = 0, 1, \ldots, t-1$, then there exists no vector $((\boldsymbol{q}^1(e_{ti}), \boldsymbol{q}^1(e^{ti})), \ldots, (\boldsymbol{q}^r(e_{ti}), \boldsymbol{q}^r(e^{ti})))$ such that

$$\sum_{k=1}^{r} [(\boldsymbol{q}^k(e_{ti}), \boldsymbol{q}^k(e^{ti})) - (\bar{\boldsymbol{q}}^k(e_{ti}), \bar{\boldsymbol{q}}^k(e^{ti}))] = 0, \quad (11.52)$$

and such that

$$U_k^{ti}({}^{ti}\boldsymbol{q}^{*k}, \boldsymbol{q}^k(e_{ti}), \boldsymbol{q}^k(e^{ti})) \geq U_k^{ti}({}^{ti}\boldsymbol{q}^{*k}, \boldsymbol{q}^{*k}(e_{ti}), \boldsymbol{q}^{*k}(e^{ti})) \qquad (11.53)$$

for all $k = 1, \ldots, r$ with strict inequality holding for at least one k. Consequently, the allocations of current and future goods prescribed by the vectors in P are Pareto-optimal in an *ex post* sense as well as in an *ex ante* sense.

A feasible tree-structure of temporary equilibria need not allocate resources Pareto-optimally. In example 11.2, for instance, neither (TE_2, TE_2^*) nor (TE_3, TE_3^*) allocates resources Pareto-optimally, as can be seen from the following equations:

$$\begin{aligned}[(\partial U_A/\partial x)/(\partial U_A/\partial y)](1.96, 0.30, 0.14) &= 0.91 \neq 3.25 \\ &= [(\partial U_B/\partial x)/(\partial U_B/\partial y)](1.75, 1.84, 0.86), \end{aligned} \qquad (11.54)$$

$$\begin{aligned}[(\partial U_A/\partial x)/(\partial U_A/\partial y)](2.85, 1.25, 0.59) &= 0.23 \neq 0.16 \\ &= [(\partial U_B/\partial x)/(\partial U_B/\partial y)](0.86, 0.89, 0.41). \end{aligned} \qquad (11.55)$$

However, we can show that any Pareto-optimal allocation of commodities can be achieved by some feasible tree-structure of competitive equilibria if the first-period distribution of purchasing power is adjusted appropriately. This result is stated below. In reading it, note that we say that a feasible tree-structure of competitive equilibria is equivalent to a Pareto-optimal allocation if the allocation of commodities prescribed by the tree-structure is identical to that prescribed by the Pareto-optimal allocation.

THEOREM 11.3. *Suppose that assumptions (1)–(5) hold. Then to each Pareto-optimal allocation of commodities corresponds a set of expected prices* $(\boldsymbol{p}, \boldsymbol{\beta})^e(e^{01})$ *shared by all consumers, a first-period distribution of purchasing power, and an equivalent feasible tree-structure of temporary equilibria.*

This theorem shows that K. Arrow's classic theorem, 'Any optimal allocation of risk-bearing can be achieved by perfect competition on the securities and commodity markets, where securities are payable in money' [1, theorem 2, p. 94], is valid not only for an exchange economy which operates under uncertainty over a single period and in which each consumer faces a finite number of possible events (i.e. Arrow's case), but also for the economy studied in this paper. Thus the answer to query number 2 is Yes.

Unfortunately, theorem 11.3 is only an existence theorem. The feasible tree-structure of competitive equilibria which is equivalent to some Pareto-optimal allocation $(q*^1, \ldots, q*^r)$ is no more stable than any other feasible tree-structure; e.g. if the economy ever gets on to a Pareto-optimal tree-structure, it is unlikely to stay on it for long.

To see why a Pareto-optimal tree-structure of temporary equilibria may be as unstable as any other feasible tree-structure of such equilibria, we will present the following example.

EXAMPLE 11.3. Consider a two-period economy with two consumers A and B, one first-period commodity x, and two second-period commodities y and z. The utility functions and price expectations of A and B are as defined in (11.40)–(11.42), and their initial endowments are given by

$$(\bar{x}_A, \bar{y}_A, \bar{z}_A) = ((2.14/3), 1.30, 0.84), \tag{11.56}$$

and

$$(\bar{x}_B, \bar{y}_B, \bar{z}_B) = (3, 0.27, 0.73). \tag{11.57}$$

Finally, neither A nor B has any outstanding debt in period one.

During period 1 there are markets for x and for a security V each unit of which will pay one unit of z in period 2. During period 2 there are markets for y and z. The 'indirect utility functions' of A and B are as defined in (11.43)–(11.44). Consequently, since the first-period endowments of this economy are the same as the first-period endowments of the economy in example 11.2, we conclude that the three temporary equilibria defined in eqs. (11.45)–(11.47) are temporary equilibria in the economy considered here as well. Moreover, there are no others. Again TE_1 and TE_3 are locally stable while TE_2 is unstable.

If TE_1 obtained in period 1, then in period 2,

$$TE_1^{**}: (p_y, p_z, y_A, z_A, y_B, z_B) = (1, 1, 0, 0, 1.57, 1.57) \tag{11.58}$$

would be the one and only one temporary equilibrium relative to the distribution of purchasing power determined by $(\bar{y}_A, \bar{z}_A + V_A) = (1.30, -1.30)$ and $(\bar{y}_B, \bar{z}_B + V_B) = (0.27, 2.87)$. The pair (TE_1, TE_1^{**}) can be characterized as a locally stable Pareto-optimal tree-structure of temporary equilibria.

If TE_2 obtained in period 1, then in period 2,

$$TE_2^{**}: (p_y, p_z, y_A, z_A, y_B, z_B) = (1, 1, 0.445, 0.445, 1.125, 1.125) \tag{11.59}$$

would be the one and only one temporary equilibrium relative to the distribution purchasing power determined by $(\bar{y}_A, \bar{z}_A + V_A) = (1.30, -0.41)$ and $(\bar{y}_B, \bar{z}_B + V_B) = (0.27, 1.98)$. The pair (TE_2, TE_2^{**}) can be characterized as an unstable Pareto-optimal tree-structure of temporary equilibria.

If TE_3 obtained in period 1, then in period 2,

$$TE_3^{**}: (p_y, p_z, y_A, z_A, y_B, z_B) = (1, 1, 0.89, 0.89, 0.68, 0.68) \quad (11.60)$$

would be the one and only one temporary equilibrium relative to the distribution of purchasing power determined by $(\bar{y}_A, \bar{z}_A + V_A) = (1.30, 0.48)$ and $(\bar{y}_B, \bar{z}_B + V_B) = (0.27, 1.09)$. The pair (TE_3, TE_3^{**}) can be characterized as a locally stable Pareto-optimal tree-structure of temporary equilibria.

The Pareto-optimal tree-structure of temporary equilibria constructed in the proof of theorem 11.3 and the Pareto-optimal tree-structures in example 11.3 have the following three properties: (1) consumer price expectations are mutually consistent; (2) consumers' plans for future acquisitions of goods and securities, associated with each and every temporary equilibrium of the tree-structure, are mutually consistent; and (3) at each event that might occur, consumers' price expectations are 'validated' and their plans are actually carried out. It is fairly easy to construct examples of Pareto-optimal tree-structures of temporary equilibria which satisfy condition (1) but neither (2) nor (3). It is also possible to construct Pareto-optimal tree-structures of temporary equilibria which satisfy condition (2) but not (1) and (3). Thus these conditions are not necessary in order that a tree-structure of temporary equilibria be Pareto-optimal. On the other hand it is easy to show that, if assumptions (1)–(5) hold, and if a tree-structure of temporary equilibria satisfy conditions (1)–(3), then it must be Pareto-optimal. For brevity we omit the proof of this assertion here.

If our model economy were moving along one of the Pareto-optimal tree-structures of temporary equilibria prescribed in theorem 11.3 and by some miscalculation of the 'auctioneer' veered off it in period t, the economy would in most cases end up at a temporary equilibrium at which consumers' plans for future acquisitions of goods and securities would be inconsistent[12]. Since these plans obviously could not be

[12] 'The economy would in most cases...' should be taken to mean 'unless consumers' utility functions are of a special nature (such as in example 11.3) the economy would end up...'.

carried out in subsequent periods, one might wonder if the consumers could be persuaded to move back on to the original Pareto-optimal tree-structure by informing them about the inconsistency of their plans. The answer is an emphatic no! Reason: it is impossible to move from one temporary equilibrium to another in the same period without decreasing the expected utility of somebody, and this somebody will not move voluntarily.

The last observation can be stated as a theorem.

THEOREM 11.4. *Suppose that assumptions* (1)–(5) *hold. Then a temporary equilibrium* $((\boldsymbol{p}, \boldsymbol{\beta})(e_{ti}), (\boldsymbol{q}, \boldsymbol{V})^1(e_{ti}), \ldots, (\boldsymbol{q}, \boldsymbol{V})^r(e_{ti}))$ *is* admissible *in the sense that, given people's expectations about the future and the fact that* e_{ti} *occurred, there exists in period t no other feasible allocation of current goods and securities at which some person would experience a higher expected utility and nobody would experience a lower expected utility.*

The last theorem and the comments that preceded it allow the following comment on lump-sum tax-subsidy schemes: It is not true – as commonly believed – that schemes of lump-sum taxes and subsidies to redistribute current purchasing power would never interfere with the efficiency of resource allocation in a competitive economy. In an exchange economy that operates under uncertainty and that contains markets for current goods and securities only, such schemes might move the economy from a tree-structure of competitive equilibria that was equivalent to a Pareto-optimal allocation to one that was not. Moreover, there is no guarantee that such an economy after a redistribution in purchasing power would end up at a feasible tree-structure of competitive equilibria; and if not, it would be plagued in subsequent periods by bankruptcies which would force further redistributions of purchasing power.

11.3. Conclusion

We have shown that there exist feasible tree-structures of competitive equilibria. So in theory an economy such as the one studied here could function smoothly. We have also shown that this smooth functioning is unlikely and that bankruptcies probably will be common.

In theoretical models the problem of bankruptcies can be dealt with in various ways. One is to assume that, for each and every pair (t, i), each unit of $V(e_{ti})$ represents a contingent claim on purchasing power in period t which will pay an unknown fraction $\gamma(e_{ti})$ of a dollar if e_{ti} obtains and nothing otherwise. The actual value of $\gamma(e_{ti})$ if e_{ti} occurred could be taken to be the largest number in [0,1] which would allow the economy to achieve a temporary equilibrium in period t.

Since each and every consumer in our economy issues the same securities, it is natural that consumers should share the risk of bankruptcy in the way suggested above. In fact this method corresponds rather well to the way in which bankruptcy is handled in practice. The insolvent consumer declares bankruptcy, and his creditors decide on an equitable sharing of his assets. If an individual 'issued' only one kind of security, the share of each creditor would be proportional to his share of the consumer's outstanding debt.

Another way of handling bankruptcies is to establish a bank which at all times stands ready to lend to consumers the amount of purchasing power needed to keep everyone solvent and at the same time allow the economy to arrive at a temporary equilibrium in each and every period. In period t if the event e_{ti} occurs, this solution amounts (1) to changing condition (11.21) in the definition of a temporary equilibrium to

$$\sum_{k=1}^{r} V*^k(e_{(t+1)j}) = \boldsymbol{\delta}_j, \qquad j \in N[ti], \tag{11.21*}$$

where $\boldsymbol{\delta}_j$ is a negative parameter whose value is determined by the bank; and (2) to changing eq. (11.20) correspondingly to allow for the increase in those commodities which the bank gives in exchange for securities. Whether or not the bank always could find an appropriate sequence of $\boldsymbol{\delta}_j$s that would ensure the smooth functioning of the economy depends on consumers' preferences and price expectations, and on the bank's resources[13].

We have also shown that an economy such as the one studied here could achieve any prescribed Pareto-optimal allocation of resources but in practice is unlikely to realize one. *One important implication of this finding is that no solid microeconomic foundation presently exists for pur-*

[13] Different approaches to the problem of bankruptcy have been suggested by Arrow and Hahn (cf. their discussion of compensated equilibria in ref. [2, chapter 5, pp. 107–122]), and by Green [8].

suing either a vigorous antitrust policy or a staunch free-trade policy in an economy which operates under uncertainty and possesses markets for current goods and securities only. To justify such policies the notion of a Pareto-optimal allocation of resources must be replaced by some other welfare criterion.

One possible candidate is the concept of an admissible allocation as defined in the statement of theorem 11.4. According to theorem 11.4, if assumptions (1)–(5) hold, a temporary equilibrium is admissible. It can also be shown that an admissible allocation in period t, if the event e_{ti} occurs, can be sustained as a temporary equilibrium provided the initial distribution of purchasing power is appropriate. Hence for the economy studied here, the relationship between temporary equilibria and admissible allocations is the same as the relationship between competitive equilibria and Pareto-optimal allocations. This sounds, however, better than it is. The result obtained in theorem 11.4 and its converse are generally false if assumption (3) is not satisfied. If consumer price expectations are multivalued and vary over time, a given temporary equilibrium need not be admissible since another temporary equilibrium might exist at which no-one experiences lower expected utility and at least one person experiences higher expected utility (cf. ref. [14], p. 556 for further details). Also a given admissible allocation might be a monopolistically competitive equilibrium which could not be sustained as a temporary equilibrium! Thus the concept of an admissible allocation, while important from a theoretical point of view, cannot provide a basis for justifying pursuit of an anti-trust or a free-trade policy in a real-life economy.

References

[1] Arrow, K. J. The role of securities in the optimal allocation of risk-bearing. *Review of Economic Studies*, **31** (1964), 91–96.

[2] Arrow, K. J. and Hahn, F. H. *General Competitive Analysis*. Holden-Day, Inc: San Francisco (1971).

[3] Borch, K. Economic equilibrium under uncertainty. *International Economic Review* (1968), 334–347.

[4] Debreu, G. *Theory of Value*. Wiley: New York (1959).

[5] Diamond, P. A. The role of a stock market in a general equilibrium model with technological uncertainty. *American Economic Review*, **57** (1967), 759–776.

[6] Foley, D. K. Lindahl's solution and the core of an economy with public goods. *Econometrica*, **38** (January 1970), 66–92.

[7] Grandmont, J. M. On the short run equilibrium in a monetary economy. CEPREMAP Discussion Paper (February 1971).
[8] Green, J. R. Preexisting contracts and temporary general equilibrium. Harvard Discussion Paper (June 1972).
[9] Hicks, J. R. *Capital and Value.* Oxford University Press: Oxford (1939).
[10] Johansen, L. *Public Economics.* Rand-McNally: Chicago (1965).
[11] Kemeny, J. G. and Shell, J. L. *Mathematical Models in the Social Sciences.* Ginn: Boston (1962).
[12] Radner, R. Existence of equilibrium of plans, prices and price expectations in a sequence of markets. *Econometrica,* **40** (1972), 289–303.
[13] Sandmo, A. Equilibrium and efficiency in loan markets. *Economica,* **37** (February 1970), 23–38.
[14] Stigum, B. P. Competitive equilibria under uncertainty. *Quarterly Journal of Economics,* **83** (November 1969), 533–561.
[15] Stigum, B. P. Resource allocation under uncertainty. *International Economic Review,* **13** (October 1972), 431–459.
[16] Stigum, B. P. Competitive equilibria with infinitely many commodities (I). *Metroeconomica,* **24** (September 1973), 221–244.
[17] Stolper, W. and Samuelson, P. A. Protection and real wages. *Review of Economic Studies,* **9** (November 1941), 58–73.

APPENDIX

Proofs of Theorems 11.1–11.4

In this appendix we provide proofs of theorems 11.1–11.4. For brevity's sake details of the proofs are spelled out only to the extent deemed necessary.

Throughout we let

$$C \equiv \{\boldsymbol{q} = (\boldsymbol{q}(e_{01}), \boldsymbol{q}(e^{01})): 0 \leq \boldsymbol{q} < \infty\}$$

and assume that the topology in C is induced by the metric

$$\|\boldsymbol{q}\|* \equiv \sum_{t=0}^{\infty} 2^{-t}\left(\sum_{j=1}^{r^t} 2^{-j}[\|\boldsymbol{q}(e_{tj})\|/(1 + \|\boldsymbol{q}(e_{tj})\|)]\right),$$

where

$$\|\boldsymbol{q}(e_{tj})\| \equiv \sum_{i=1}^{m} |\boldsymbol{q}_i(e_{tj})|.$$

We also assume throughout that, for each $k = 1, \ldots, r$, $U_k^{01}(\cdot)$ is defined on all of C. Then assumption (5) implies that the $U_k^{01}(\cdot)$ are all continuous, increasing, strictly concave functions on C.

PROOF OF THEOREM 11.1. For each $k = 1, \ldots, r$, and $j = 1, \ldots, r_1$, let

$$f_{jk}(q(e_{01}), V(e_{1j})) \equiv \max_{(q(e_{1j}), q(e^{1j})) \in \Gamma_{jk}(\tilde{p}, V(e_{1j}))} U_k^{1j}(q(e_{01}), q(e_{1j}), q(e^{1j})), \quad \text{(A11.1)}$$

where

$$\Gamma_{jk}(\tilde{p}, V(e_{1j})) \equiv \{(q(e_{1j}), q(e^{1j})) \geq 0 : (\tilde{p}(e_{1j}), \tilde{p}(e^{1j}))\,[(q(e_{1j}), q(e^{1j})) - (\bar{q}^k(e_{1j}), \bar{q}^k(e^{1j}))] \leq V(e_{1j})\}. \quad \text{(A11.2)}$$

Moreover, let

$$F_k((q, V)(e_{01})) \equiv \sum_{j=1}^{r_1} Q_k(e_{1j}) f_{jk}(q(e_{01}), V(e_{1j})), \; k = 1, \ldots, r. \quad \text{(A11.3)}$$

Then $F_k((q, V)(e_{01}))$ measures the maximum expected utility which the kth consumer could achieve if he chose $(q, V)(e_{01})$ in the first period. It is easy to show that $F_k(\cdot)$ is well-defined, continuous, increasing, and strictly concave on

$$\tilde{A}_k \equiv \{0 \leq q(e_{01}) < \infty\} \times \prod_{j=1}^{r_1} \{-(\tilde{p}(e_{1j}), \tilde{p}(e^{1j}))(\bar{q}^k(e_{1j}), \bar{q}^k(e^{1j})) \leq V(e_{1j}) < \infty\}. \quad \text{(A11.4)}$$

It is equally easy to show that $((p, \beta)(e_{01}), (q*, V*)^1(e_{01}), \ldots, (q*, V*)^r(e_{01}))$ is a temporary equilibrium in the first period if and only if

$$(p, \beta)(e_{01}) > 0; \quad \text{(A11.5)}$$

$$\sum_{k=1}^{r} (q*, V*)^k(e_{01}) = \left(\sum_{k=1}^{r} \bar{q}^k(e_{01}), 0\right); \quad \text{(A11.6)}$$

and for each k, $k = 1, \ldots, r$, $(q*, V*)^k(e_{01})$ maximizes the value of $F_k(\cdot)$ subject to the constraints

$$(p, \beta)(e_{01})(q, V)(e_{01}) \leq p(e_{01})\bar{q}^k(e_{01});$$
$$0 \leq q(e_{01}); \; -(\tilde{p}(e_{1j}), \tilde{p}(e^{1j}))(\bar{q}^k(e_{1j}), \bar{q}^k(e^{1j})) \leq V(e_{1j}), j = 1, \ldots, r_1. \quad \text{(A11.7)}$$

With this characterization of temporary equilibria for the first period in mind we can establish theorem 11.1 in the following way: let $0 < \delta < 1$ and give δ units of each of the components of $\boldsymbol{q}(e_{01})$ and of each of the securities $V(e_{1j})$, $j = 1, \ldots, r_1$, to each consumer. Moreover, let

$$\bar{\boldsymbol{H}}_{kj} \equiv (\tilde{\boldsymbol{p}}(e_{1j}), \tilde{\boldsymbol{p}}(e^{1j}))(\bar{\boldsymbol{q}}^k(e_{1j}), \bar{\boldsymbol{q}}^k(e^{1j})),$$

and let consumer k's choice of $\boldsymbol{q}(e_{01})$ and $V(e_{1j})$, $j = 1, \ldots, r_1$, be restricted to the set

$$C_k^* \equiv \{(\boldsymbol{q}, V)(e_{01}) \colon 0 \leq \boldsymbol{q}(e_{01}) \leq \boldsymbol{H}, -\bar{\boldsymbol{H}}_{kj} \leq V(e_{1j}) \leq \boldsymbol{H}*, \; j = 1, \ldots, r_1\}, k = 1, \ldots, r, \tag{A11.8}$$

where $\boldsymbol{H}$ and $\boldsymbol{H}*$ are finite constants which satisfy the inequalities

$$\boldsymbol{H} > 2 \cdot \left(\sum_{k=1}^{r} \bar{\boldsymbol{q}}^k(e_{01}) \right), \tag{A11.9}$$

and

$$\boldsymbol{H}* > 2 \cdot \max_{1 \leq j \leq r_1} \left[\sum_{k=1}^{r} (\tilde{\boldsymbol{p}}(e_{1j}), \tilde{\boldsymbol{p}}(e^{1j}))(\bar{\boldsymbol{q}}^k(e_{1j}), \bar{\boldsymbol{q}}^k(e^{1j})) + 1 \right]. \tag{A11.10}$$

Then we can use standard methods to show that there exists a vector

$$((\boldsymbol{p}, \boldsymbol{\beta})(e_{01}), (\boldsymbol{q}*, V*)^1(e_{01}), \ldots, (\boldsymbol{q}*, V*)^r(e_{01}))$$

which satisfies the conditions

$$(\boldsymbol{p}, \boldsymbol{\beta})(e_{01}) > 0 \text{ and } \boldsymbol{p}(e_{01}) + \sum_{j=1}^{r_1} \boldsymbol{\beta}(e_{1j}) = 1; \tag{A11.11}$$

$$\sum_{k=1}^{r} \boldsymbol{q}*^k(e_{01}) = \sum_{k=1}^{r} \bar{\boldsymbol{q}}^k(e_{01}) + r(\delta, \ldots, \delta); \tag{A11.12}$$

$$\sum_{k=1}^{r} V*^k(e_{1j}) = r\delta, j = 1, \ldots, r_1; \tag{A11.13}$$

$$(\boldsymbol{q}*, V*)^k(e_{01}) \in \Gamma_k((\boldsymbol{p}, \boldsymbol{\beta})(e_{01}), \bar{\boldsymbol{q}}^k(e_{01}), \delta), k = 1, \ldots, r, \tag{A11.14}$$

and

$$\boldsymbol{F}_k((\boldsymbol{q}*, V*)^k(e_{01})) = \max_{(\boldsymbol{q}, V)(e_{01}) \in \Gamma_k((\boldsymbol{p}, \boldsymbol{\beta})(e_{01}), \bar{\boldsymbol{q}}^k(e_{01}), \delta)} \boldsymbol{F}_k((\boldsymbol{q}, V)(e_{01})), \tag{A11.15}$$

where

$$\Gamma_k((\boldsymbol{p}, \boldsymbol{\beta})(e_{01}), \bar{\boldsymbol{q}}^k(e_{01}), \delta) \equiv \{(\boldsymbol{q}, \boldsymbol{V})(e_{01}) \in C_k^*: (\boldsymbol{p}, \boldsymbol{\beta})(e_{01})[(\boldsymbol{q}, \boldsymbol{V})(e_{01}) - (\bar{\boldsymbol{q}}^k + \boldsymbol{\delta}, \boldsymbol{\delta})(e_{01})] \leq 0\}.$$[14]

Next we let $\{\delta^s\}$ be a sequence of numbers such that $\lim_{s\to\infty}\delta^s = 0$. For each δ^s we compute a vector

$$((\boldsymbol{p}, \boldsymbol{\beta})^s(e_{01}), (\boldsymbol{q}^s, \boldsymbol{V}^s)^1(e_{01}), \ldots, (\boldsymbol{q}^s, \boldsymbol{V}^s)^r(e_{01}))$$

that satisfies eqs. (A11.11)–(A11.15) with δ replaced by δ^s. Let

$$\{((\boldsymbol{p}, \boldsymbol{\beta})^{s_i}(e_{01}), (\boldsymbol{q}^{s_i}, \boldsymbol{V}^{s_i})^1(e_{01}), \ldots, (\boldsymbol{q}^{s_i}, \boldsymbol{V}^{s_i})^r(e_{01}))\}$$

be a convergent subsequence and let

$$((\boldsymbol{p}, \boldsymbol{\beta})^0(e_{01}), (\boldsymbol{q}^0, \boldsymbol{V}^0)^1(e_{01}), \ldots, (\boldsymbol{q}^0, \boldsymbol{V}^0)^r(e_{01})) = \lim_{s_i\to\infty} ((\boldsymbol{p}, \boldsymbol{\beta})^{s_i}(e_{01}), (\boldsymbol{q}^{s_i}, \boldsymbol{V}^{s_i})^1(e_{01}), \ldots, (\boldsymbol{q}^{s_i}, \boldsymbol{V}^{s_i})^r(e_{01})). \tag{A11.16}$$

Suppose $(\boldsymbol{p}, \boldsymbol{\beta})^0(e_{01}) > 0$. Then rather standard arguments will show that the left-hand side of eq. (A11.16) is a temporary equilibrium in the economy in which consumers are forced to choose their $(\boldsymbol{q}, \boldsymbol{V})(e_{01})$ vectors from $C*$. Moreover, since $(\boldsymbol{p}, \boldsymbol{\beta})^0(e_{01}) > 0$, the latter constraint can easily be seen to be non-binding. Therefore the left-hand side of eq. (A11.16) must be a temporary equilibrium in the original economy as well.

To show that $(\boldsymbol{p}, \boldsymbol{\beta})^0(e_{01}) > 0$ we proceed as follows: suppose that $\boldsymbol{\beta}^0(e_{11}) = 0$ and observe that at least one of the components of the limiting price vector must be positive. So suppose $\boldsymbol{\beta}^0(e_{12}) > 0$. Then note that there is at least one consumer, say the kth, for whom

$$(\tilde{\boldsymbol{p}}(e_{12}), \tilde{\boldsymbol{p}}(e^{12}))(\bar{\boldsymbol{q}}^k(e_{12}), \bar{\boldsymbol{q}}^k(e^{12})) > 0.$$

This consumer's demand functions (for $(\boldsymbol{q}, \boldsymbol{V})(e_{01})$) are continuous at $(\boldsymbol{p}, \boldsymbol{\beta})^0(e_{01})$. Therefore, since $\boldsymbol{F}_k(\cdot)$ is strictly increasing on $\tilde{A}_k$, for sufficiently large s_i he will demand more of $\boldsymbol{V}(e_{11})$ than is available. This contradicts the fact that the right-hand side of eq. (A11.16) for this s_i satisfies eqs. (A11.11)–(A11.15). Hence $\boldsymbol{\beta}^0(e_{11}) > 0$.

---

[14] Here

$$(\bar{\boldsymbol{q}}^k + \boldsymbol{\delta}, \boldsymbol{\delta})(e_{01}) \equiv (\bar{q}_1^k + \delta, \ldots, \bar{q}_m^k + \delta, \delta, \ldots, \delta)(e_{01}),$$

where $(\delta, \ldots, \delta)$ has r_1 components.

Arguments similar to those used for $\beta^0(e_{11})$ can be used to show that all the components of $(\boldsymbol{p}, \boldsymbol{\beta})^0(e_{01})$ must be positive. Since there is no need to give further details, the proof of the theorem is complete. *Q.E.D.*

PROOF OF THEOREM 11.2. To prove theorem 11.2 we begin by observing that there exists (*cf.* ref. [16, pp. 2–3]) a vector $(\boldsymbol{p}^*, \boldsymbol{q}^{*1}, \ldots, \boldsymbol{q}^{*r})$ which satisfies the following conditions

$$\boldsymbol{p}^* > 0; \tag{A11.17}$$

$$\boldsymbol{p}^*\left(\sum_{k=1}^{r} \bar{\boldsymbol{q}}^k\right) = 1; \tag{A11.18}$$

$$\boldsymbol{q}^{*k} \in C, \boldsymbol{p}^*\boldsymbol{q}^{*k} = \boldsymbol{p}^*\bar{\boldsymbol{q}}^k,\ k = 1, \ldots, r; \tag{A11.19}$$

$$U_k^{01}(\boldsymbol{q}^{*k}) = \max_{\boldsymbol{q} \in \{\boldsymbol{q} \in C:\ \boldsymbol{p}^*\boldsymbol{q} \leq \boldsymbol{p}^*\bar{\boldsymbol{q}}^k\}} U_k^{01}(\boldsymbol{q}); \tag{A11.20}$$

and

$$\sum_{k=1}^{r} (\boldsymbol{q}^{*k} - \bar{\boldsymbol{q}}^k) = 0. \tag{A11.21}$$

Then we use the vector $\boldsymbol{p}^*$ to construct a set of expected prices:

$$\boldsymbol{\beta}^e(e_{ti}) \equiv (\|\boldsymbol{p}^*(e_{ti})\| / \|\boldsymbol{p}^*(e_{(t-1)j(ti)})\|),\ t = 1, 2, \ldots;\ i = 1, \ldots, r_t, \tag{A11.22}$$

where $j_{(ti)}$ is the unique integer for which $e_{ti} \subset e_{(t-1)j(ti)}$, and

$$\boldsymbol{p}^e(e_{ti}) \equiv \left[\prod_{\substack{s=1 \\ e_{ti} \subset e_{sj_s}}}^{t} \boldsymbol{\beta}^e(e_{sj_s})\right]^{-1} \boldsymbol{p}^*(e_{ti}),\ t = 1, 2, \ldots;\ i = 1, \ldots, r_t. \tag{A11.23}$$

Next we define a family of security vectors: for each pair (t, i), $t = 1, 2, \ldots;\ i = 1, \ldots, r_t$, let $\varphi_s(e_{ti})$ be the set of all indices j such that $e_{sj} \subset e_{ti}$. Evidently, $\varphi_s(\cdot)$ is well-defined and non-empty for all $s \geq t$. Next, for each $\ell = t + 1, \ldots, s$, and $j \in \varphi_s(e_{ti})$ let ℓ_j be the unique integer such that $e_{sj} \subset e_{\ell\ell_j}$. Then for $t = 1, 2, \ldots;\ i = 1, \ldots r_t$ let

$$\boldsymbol{p}^{ti}(e_{sj}) \equiv \begin{cases} \boldsymbol{p}^e(e_{ti}) \text{ for } s = t, \text{ and} \\ \left[\prod_{\ell=t+1}^{s} \boldsymbol{\beta}^e(e_{\ell\ell_j})\right]\boldsymbol{p}^e(e_{sj}) \text{ for } s > t,\ j \in \varphi_s(e_{ti}). \end{cases} \tag{A11.24}$$

Finally, for each pair (t, i), $t = 1, 2, \dots, i = 1, \dots, r_t$, let

$$V^{*k}(e_{ti}) \equiv \sum_{s \geq t} \sum_{j \in \varphi_s(e_{ti})} \boldsymbol{p}^{ti}(e_{sj})(\boldsymbol{q}^{*k}(e_{sj}) - \bar{\boldsymbol{q}}^k(e_{sj})), \; k = 1, \dots, r. \tag{A11.25}$$

To prove theorem 11.2 we now let

$$(\boldsymbol{p}, \boldsymbol{\beta})(e_{01}) \equiv (\boldsymbol{p}^*, \boldsymbol{\beta}^e)(e_{01}), \tag{A11.26}$$

and

$$(\boldsymbol{p}, \boldsymbol{\beta})(e_{ti}) \equiv (\boldsymbol{p}, \boldsymbol{\beta})^e(e_{ti}), \; t = 1, 2, \dots; \; i = 1, \dots, r_t. \tag{A11.27}$$

Then we show that the family of vectors

$$\{((\boldsymbol{p}, \boldsymbol{\beta})(e_{ti}), (\boldsymbol{q}^*, V^*)^1(e_{ti}), \dots, (\boldsymbol{q}^*, V^*)^r(e_{ti})), \; t \geq 0, i = 1, \dots, r_t\}$$

is a feasible tree-structure of competitive equilibria: let (t, i) be chosen arbitrarily, let

$$A^{ti} \equiv \{[(\boldsymbol{q}^1(e_{ti}), \boldsymbol{q}^1(e^{ti})), \dots, (\boldsymbol{q}^r(e_{ti}), \boldsymbol{q}^r(e^{ti}))] \geq 0 \colon \sum_{k=1}^{r} [(\boldsymbol{q}^k(e_{ti}), \boldsymbol{q}^k(e^{ti})) - (\bar{\boldsymbol{q}}^k(e_{ti}), \bar{\boldsymbol{q}}^k(e^{ti}))] = 0\}, \tag{A11.28}$$

and let

$P^{ti} \equiv \{[(\boldsymbol{q}^1(e_{ti}), \boldsymbol{q}^1(e^{ti})), \dots, (\boldsymbol{q}^r(e_{ti}), \boldsymbol{q}^r(e^{ti}))] \in A^{ti}$: $\exists$ no $z \in A^{ti}$ such that $U_k^{ti}({}^{ti}\boldsymbol{q}^{*k}, z^k) \geq U_k^{ti}({}^{ti}\boldsymbol{q}^{*k}, (\boldsymbol{q}^k(e_{ti}), \boldsymbol{q}^k(e^{ti})))$ for all k with strict inequality holding for at least one $k\}$. (A11.29)

Observe that, under the assumption that each consumer k, $k = 1, \dots, r$, consumed the vector ${}^{ti}\boldsymbol{q}^{*k}$ during periods $s = 0, 1, \dots, t-1$, P^{ti} represents for period t and the event e_{ti} the set of all Pareto-optimal allocations of $\boldsymbol{q}(e_{ti})$ and contingent claims on $\boldsymbol{q}(e_{sj})$ for all $s > t$ and $j \in \varphi_s(e_{ti})$.

Next observe that

$$[(\boldsymbol{q}^{*1}(e_{ti}), \boldsymbol{q}^{*1}(e^{ti})), \dots, (\boldsymbol{q}^{*r}(e_{ti}), \boldsymbol{q}^{*r}(e^{ti}))] \in P^{ti},$$

and that, for all

$$k = 1, \dots, r(\boldsymbol{q}^{*k}(e_{ti}), \boldsymbol{q}^{*k}(e^{ti})) \in \Gamma_k^{ti},$$

where

$$\Gamma_k^{ti} \equiv \{(\boldsymbol{q}(e_{ti}), \boldsymbol{q}(e^{ti})) \geq 0 \colon (\boldsymbol{p}^{ti}(e_{ti}), \boldsymbol{p}^{ti}(e^{ti}))[(\boldsymbol{q}(e_{ti}), \boldsymbol{q}(e^{ti})) - (\bar{\boldsymbol{q}}^k(e_{ti}), \bar{\boldsymbol{q}}^k(e^{ti}))] \leq V^{*k}(e_{ti})\}. \tag{A11.30}$$

Note also that, for all $k = 1, \ldots, r$

$$U_k^{ti}({}^{ti}\boldsymbol{q}^{*k}, \boldsymbol{q}^{*k}(e_{ti}), \boldsymbol{q}^{*k}(e^{ti})) = \max_{(\boldsymbol{q}(e_{ti}), \boldsymbol{q}(e^{ti})) \in \Gamma_k^{ti}} U_k^{ti}({}^{ti}\boldsymbol{q}^{*k}, \boldsymbol{q}(e_{ti}), \boldsymbol{q}(e^{ti})). \quad \text{(A11.31)}$$

Finally note – it is easy to show – that for $k = 1, \ldots, r$

$$((\boldsymbol{q}*, V*)^k(e_{ti}), (\boldsymbol{q}*, V*)^k(e^{ti})) \in \Gamma_k^*, \quad \text{(A11.32)}$$

where

$$\Gamma_k^* \equiv \Gamma((\boldsymbol{p}, \boldsymbol{\beta})(e_{ti}), (\boldsymbol{p}, \boldsymbol{\beta})^e(e^{ti}), \bar{\boldsymbol{q}}^k(e_{ti}), \bar{\boldsymbol{q}}^k(e^{ti}), V*^k(e_{ti})).$$

From eq. (A11.32) it follows that, if for all $k = 1, \ldots, r$

$$((\tilde{\boldsymbol{q}}, \tilde{V})^k(e_{ti}), (\tilde{\boldsymbol{q}}, \tilde{V})^k(e^{ti})) \in \Gamma_k^*$$

and

$$U_k^{ti}({}^{ti}\boldsymbol{q}*^k, \tilde{\boldsymbol{q}}^k(e_{ti}), \tilde{\boldsymbol{q}}^k(e^{ti})) = \max_{((\boldsymbol{q},V)(e_{ti}),(\boldsymbol{q},V)(e^{ti})) \in \Gamma_k^*} U^{ti}({}^{ti}\boldsymbol{q}*^k, \boldsymbol{q}(e_{ti}), \boldsymbol{q}(e^{ti})), \quad \text{(A11.33)}$$

then

$$U_k^{ti}({}^{ti}\boldsymbol{q}*^k, \boldsymbol{q}*^k(e_{ti}), \boldsymbol{q}*^k(e^{ti})) \leq U^{ti}({}^{ti}\boldsymbol{q}*^k, \tilde{\boldsymbol{q}}^k(e_{ti}), \tilde{\boldsymbol{q}}^k(e^{ti})), k = 1, \ldots, r. \quad \text{(A11.34)}$$

On the other hand, it is easy to show that, for all $k = 1, \ldots, r$, $(\tilde{\boldsymbol{q}}^k(e_{ti}), \tilde{\boldsymbol{q}}^k(e^{ti})) \in \Gamma_k^{ti}$. This fact and eq. (A11.31) imply that equality must hold in eq. (A11.34) for all k. Consequently, since $U_k^{ti}({}^{ti}\boldsymbol{q}^{*k}, \cdot)$ is strictly concave

$$(\tilde{\boldsymbol{q}}^k(e_{ti}), \tilde{\boldsymbol{q}}^k(e^{ti})) = (\boldsymbol{q}^{*k}(e_{ti}), \boldsymbol{q}^{*k}(e^{ti})), k = 1, \ldots, r. \quad \text{(A11.35)}$$

To conclude the proof of the theorem we now observe that $(\boldsymbol{p}, \boldsymbol{\beta})(e_{ti})$ satisfies eq. (11.19) and that $(\boldsymbol{q}*^1(e_{ti}), \ldots, \boldsymbol{q}*^r(e_{ti}))$ satisfies eq. (11.20). We also note that eq. (A11.25) and the fact that $[(\boldsymbol{q}^{*1}(e_{ti}), \boldsymbol{q}^{*1}(e^{ti})), \ldots, (\boldsymbol{q}^{*r}(e_{ti}), \boldsymbol{q}^{*r}(e^{ti}))] \in P^{ti}$ imply that

$$\sum_{k=1}^{r} V^{*k}(e_{(t+1)j}) = 0, j \in N[ti]. \quad \text{(A11.36)}$$

But if that is true, then eqs. (A11.32), (A11.33) and (A11.35) imply that $((\boldsymbol{p}, \boldsymbol{\beta})(e_{ti}), (\boldsymbol{q}^*, V^*)^1(e_{ti}), \ldots, (\boldsymbol{q}^*, V^*)^r(e_{ti}))$ is in fact a temporary equilibrium relative to the distribution of purchasing power $(\boldsymbol{p}(e_{ti})\bar{\boldsymbol{q}}^1(e_{ti}) + V^{*1}(e_{ti}), \ldots, \boldsymbol{p}(e_{ti})\bar{\boldsymbol{q}}^r(e_{ti}) + V^{*r}(e_{ti}))$. This result and the fact that (t, i) was chosen arbitrarily establish the validity of the theorem. *Q.E.D.*

PROOF OF THEOREM 11.3. Suppose that $(\boldsymbol{q}*^1, \ldots, \boldsymbol{q}*^r) \in P$. By theorem 11.6 in ref. [16] there exists a vector $\boldsymbol{p}*$ such that $(\boldsymbol{p}*, \boldsymbol{q}*^1, \ldots, \boldsymbol{q}*^r)$ satisfies eqs. (A11.17)–(A11.21) with $\bar{\boldsymbol{q}}^k$ replaced by $\boldsymbol{q}*^k$ in eqs. (A11.19) and (A11.20). Consequently, if we define $(\boldsymbol{p}, \boldsymbol{\beta})^e(e^{01})$ as in eqs. (A11.22)–(A11.23), $\boldsymbol{p}^{ti}(e_{sj})$ as in eq. (A11.24), $\boldsymbol{V}*^k(e_{ti})$ as in eq. (A11.25) and $(\boldsymbol{p}, \boldsymbol{\beta})(e_{01})$ and $(\boldsymbol{p}, \boldsymbol{\beta})(e_{ti})$ as in eqs. (A11.26)–(A11.27), then we can show that the family of vectors

$$\{((\boldsymbol{p}, \boldsymbol{\beta})(e_{ti}), (\boldsymbol{q}*, \boldsymbol{V}*)^1(e_{ti}), \ldots, (\boldsymbol{q}*, \boldsymbol{V}*)^r(e_{ti})), t \geq 0, i = 1, \ldots, r_t\}$$

is a feasible tree-structure of competitive equilibria relative to the first-period distribution of purchasing power

$$((\boldsymbol{p}, \boldsymbol{\beta})(e_{01})(\boldsymbol{q}*, \boldsymbol{V}*)^1(e_{01}), \ldots, (\boldsymbol{p}, \boldsymbol{\beta})(e_{01})(\boldsymbol{q}*, \boldsymbol{V}*)^r(e_{01})).$$

Since there is no need to give further details, the proof of theorem 11.3 is complete. *Q.E.D.*

PROOF OF THEOREM 11.4. If we let

$$A* \equiv \{((\boldsymbol{q}, \boldsymbol{V})^1(e_{ti}), \ldots, (\boldsymbol{q}, \boldsymbol{V})^r(e_{ti})): \boldsymbol{q}^k \geq 0, \boldsymbol{V}^k \geq -(\tilde{\boldsymbol{p}}(e_{ti}), \tilde{\boldsymbol{p}}(e^{ti}))$$
$$(\bar{\boldsymbol{q}}^k(e_{ti}), \bar{\boldsymbol{q}}^k(e^{ti})), \sum_{k=1}^{r} ((\boldsymbol{q}, \boldsymbol{V})^k(e_{ti}) - (\bar{\boldsymbol{q}}^k, 0)(e_{ti})) = 0\}, \qquad \text{(A11.37)}$$

then the set of goods-securities allocations that are admissible in period t if the event e_{ti} occurs is

$*P^{ti} \equiv \{((\boldsymbol{q}, \boldsymbol{V})^1(e_{ti}), \ldots, (\boldsymbol{q}, \boldsymbol{V})^r(e_{ti})) \in A*$: $\exists$ no $z \in A*$ such that $\boldsymbol{F}_k(z^k) \geq \boldsymbol{F}_k((\boldsymbol{q}, \boldsymbol{V})^k(e_{ti}))$ for all k with inequality holding for at least one $k\}$. (A11.38)

From this definition of admissible strategies, it follows that theorem 11.4 can be proved by using standard methods. There is no need to spell out the details here. So we can consider the validity of theorem 11.4 established. *Q.E.D.*

COMMENTS

On sequences of temporary equilibrium

John O. Ledyard*

C11.1.1. Discussion

The papers by Green and Stigum, appearances notwithstanding, are complementary. They are also important contributions to the theory of temporary equilibrium; a theory of market systems in which the markets for trading commodities may be open each day. This model thus constitutes a generalization of the Arrow–Debreu model. In modeling such an economy, several issues must be faced immediately. One is the possibility that, at some date, the equilibrium price in certain markets may be different from what it was in the past. For example, the price paid today for delivery of goods in 1980 may be different from the price paid three days ago for the same delivery contract. Once this possibility is introduced into the model, it is necessary to introduce expectations (usually on prices) about the future possitilities for trades on the (currently) closed markets. Finally, once such expectations are introduced, speculation becomes not only possible but potentially profitable. Thus, individuals may contract to deliver commodities at some future date which they currently do not own (i.e., they sell short), under the expectation that they can buy up the appropriate amount on some day prior to the date delivery is to be made. Such behavior can lead to an inability to deliver if prices are not as expected, thus causing bankruptcy as a result of past behavior. Bankruptcy in turn may operate to insure that no temporary equilibrium exists, thus exposing a basic deficiency in the model.

Green and Stigum approach these problems in different ways. This can be more clearly seen if we initially consider a result of Arrow and Hahn (1, theorem 7, p. 121). They show that, under acceptable assumptions, a compensated temporary equilibrium exists even if bankruptcy

* I would like to thank James Gordan for helpful conversations and insights.

is allowed. In general, this compensated equilibrium will not be a market equilibrium; however, it is true that there is a redistribution of initial endowments such that the compensated equilibrium will be a market equilibrium after the redistribution. Although Stigum's paper does more, one of his results provides conditions on preferences and expectations such that the compensated equilibrium is a market equilibrium, *without a redistribution of endowments.* Green, on the other hand, accepts that a redistribution is likely to be necessary and provides a set of institutional rules which will accomplish this.

The contribution of each paper is now evident. From Stigum's work it is obvious that the sufficient conditions required to insure that a sequence of temporary equilibria exists without any consumer becoming bankrupt (what he calls a feasible tree structure of temporary equilibria), in economies with only markets for current goods and securities, are so strong that any possibility that reality is encompassed is effectively eliminated. Thus, Stigum's paper contains the motivation for Green's work. That is, if bankruptcy is as likely a possibility as Stigum's paper indicates, methods for dealing with it must be developed. Green, in his potentially seminal paper, provides us with one possibility.

Each paper is important; however, each has certain weaknesses. It is clear that each author ignored these in order to concentrate on what he considered to be the important issues. Thus my reason for discussing these points is not to be critical of the authors but to indicate the desirability of certain future work. It seems to me that the description of the behavior of equilibria in the presence of inactive markets is one of the main contributions of the theory of temporary equilibrium. Stigum's paper requires all commodity futures markets to be inactive: Green's requires none to be inactive. Each paper suffers a little from these extreme positions. Green's assumption implies that if there is no bankruptcy currently and if consumers expect with certainty that today's (relative) futures prices will be tomorrows' current prices, then expectations will be fulfilled and no bankruptcies will occur in the future. Another problem is that if some futures markets did not exist then the concept of the present value of wealth (defined at current prices) which he uses as a standard of bankruptcy is not defined. Personally I do not think this detracts from his major insights; however, it would be nice to see what happens in a world without complete futures markets. Stigum's assumption (that no commodity futures markets are active) is restrictive in a different sense. In particular it implys that all debt is paid off in 'dollars' (the unit of account). This immediately makes the price level (as opposed to relative prices) important. For example, normalization of first period prices is legitimate only if $V^k(e_{01})$, the dollar amount of securities maturing at time 0 held by consumer k, equals zero for all k. Otherwise, since commodity endowments are positive, the price level in period 1 could be set high enough such that all debt can be paid. (For additional comments on the role of the price level in a 'monetary' economy see Arrow and Hahn [1, pp. 347–369].) Thus assumption II of Stigum that $V(e_{01}) = 0$ is crucial and effectively eliminates the possibility of a past – precisely the phenomena Green introduces in his model. In Green's paper all debt is owed in commodities and, therefore, he does not have to face this problem. Clearly, more research is needed before the precise relationship between and impact of alternative debt forms (money, securities, or commodities) is understood.

Perhaps the most troubling aspect of Green's analysis is that he is only able to show the existence of an approximate equilibrium. This results from the fact that, because of the institutional arrangements to handle bankruptcy, demand correspondences can be non-convex. It is tempting to ask whether it is possible to revise the institutional rules for redistributing endowments in response to bankruptcies in a way which rescues the convexity of the consumers' demand correspondences. Since several of the disagreements between Green and Stigum must be resolved in such a revision, it is of interest to explore some possibilities.

The main disagreement arises over the definition of bankruptcy. Green defines a consumer to be bankrupt at the prices p if his net present value of wealth (both endowments and contracts) is negative when valued at today's current and future prices. Stigum defines a consumer to be bankrupt if he is unable to find someone who is willing to refinance his currently expiring contractual debts. Let us be a bit more precise. Let ${}^t w = (w_t, w_{t+1}, \ldots)$ be the consumer's current and future endowments at t. Let ${}^{t-1}e = ({}^{t-1}e_t, \ldots)$ be the consumer's current and future contracted commitments at date t. (${}^{t-1}e_{\tau k} > 0$ means he holds, at t, contracts for commodity k to be delivered to him at τ.) Given ${}^t w$, ${}^{t-1}e$ and prices ${}^t p$, the consumer must choose at t a vector of trades (current and futures contracts) ${}^t b$ such that

$$ {}^t p \cdot {}^t b \leq 0, \qquad \text{for all } t. \tag{C11.1} $$

(*Note:* if some markets are inactive then the appropriate entries in ${}^tb$ must equal zero.) Having signed these contracts the consumer consumes, at t, $x_t = w_t + {}^{t-1}e_t + {}^tb_t$ and has remaining contracts of ${}^te$ where ${}^te_\tau = {}^{t-1}e_\tau + {}^tb_\tau$ for $\tau \geq t$. He is constrained to choose (for survival reasons)

$$ {}^tx_t \geq 0 \qquad \text{for all } t. \tag{C11.2} $$

Green's definition of bankruptcy can be seen by rewriting (C11.1) as follows:

$$ {}^tp \cdot {}^tx \leq {}^tp \cdot {}^tw + {}^tp \cdot {}^{t-1}e \tag{C11.3} $$

where ${}^tx$ is his planned current and future consumption. A consumer is then bankrupt if there is no consumption *plan* which satisfies (C11.2) and (C11.3) simultaneously. This occurs if and only if the net present value of wealth, ${}^tp \cdot {}^tw + {}^tp \cdot {}^{t-1}e < 0$. This view is certainly consistent with the approach of Debreu [2], where ${}^0e = 0$ and ${}^1x_\tau = {}^\tau x_\tau$ for all $\tau \geq 1$. (*Note:* if some markets are inactive, it is not clear what 'expected' prices should be used to evaluate ${}^tw$. For our purposes, however, this is a side issue.)

Stigum's definition of bankruptcy can be seen by rewriting (C11.1) as follows:

$$ {}^tp_t \cdot {}^tx_t + {}^t\hat{p} \cdot {}^t\hat{b} \leq {}^tp_tw_t + {}^tp_t \cdot {}^{t-1}e_t \tag{C11.4} $$

where ${}^tp = ({}^tp_t, {}^t\hat{p})$ and ${}^tb = ({}^tb_t, {}^t\hat{b})$. A consumer is then bankrupt if there is no consumption vector ${}^tx_t$ satisfying (C11.2) and a trade ${}^t\hat{b}$ which can be completed *in equilibrium* such that (C11.4) holds. This view is consistent with that of Debreu [3] where ${}^tw < 0$ is possible in a limited way which ensures no bankruptcy of this type. However appealing this view might be as a representation of reality, I find it less compelling as a concept of bankruptcy in a *tatonnement* system since it is not independent of the existence of equilibrium. That is, it states an individual is bankrupt if there is no equilibrium such that he is not bankrupt. This seems particularly circular to me.

Another extreme view would be that a consumer is bankrupt if the value of currently maturing obligations ${}^tp_t \cdot w_t + {}^tp_t \cdot {}^{t-1}e_t < 0$. That is, he is given no opportunity to refinance his debt.

We thus have at least three possible views of bankruptcy from which to choose (Green's is an intermediate case). The only reason the choice must be made is that both Green's and Stigum's institutional rules force a consumer to declare bankruptcy if and only if he is bankrupt according to their criterion. A possible way out of this dilemma is to allow the consumer to choose the extent of his default as a decision variable. That is, we allow default plans just as we allow consumption plans. This would be more consistent with the idea of an informationally decentralized market system. Let me try to indicate how this might work, using Green's notations and concepts.

Instead of relying on Green's institutional default rule, $d(p, r) = \min\{\delta \geq 0 \mid p(w + re_+ + (1 - \delta)e_-) \geq 0\}$, we allow each consumer, i, to choose $d^i \in [0,1]$ given prices p and returns ratios r. Once each consumer has done so, new returns ratios are computed, as in Green, and prices are adjusted. Then new demands and default ratios are computed, etc. The only question is by what criterion does a consumer select a d^i? Clearly if there is no penalty connected with $d^i > 0$, and if preferences are monotonic then he would always maximize utility by choosing $d^i = 1$. That is, he would always desire to default on all commitments. On the other hand, if the penalty is severe (as, for example, in Green where $d^i > 0$ implies ${}^t x_\tau = 0$ for all τ), he might never desire to default unless forced to do so.

The first problem, the consumer always defaults, is inherent in any model where contracts are not strictly enforced. It is basically a problem of public goods (more precisely, bads) and involves an element of social trust. In this sense contracts are like money in that money won't be held (contracts won't be signed) unless there is some faith that it can be exchanged (that they will be carried out). Rather than sort this problem out, it is easier to assume that defaulted contracts carry some disutility (because of, say, social norms) to the defaulter. In particular we can let (as in Green) the utility function of a consumer be $u(x_1, y_1, x_2, y_2, \ldots)$ where x_t is consumption at t and $y_t = d_t^t p \cdot {}^{t-1}e_-$ is the dollar value of defaulted contracts at date t.

The second problem, that d might always be chosen to be 0, is not as easily solved. It is a problem because each consumer, given the price ${}^t p$, can always *plan* a trade ${}^t b$ which would refinance all his debt. However, there may be no price such that these plans add up (across consumers) to zero in which case there is no equilibrium. If t is the final decision period, the problem disappears since the consumer must choose d and ${}^t b$ such that $x_t \geq 0$. It is easily shown that if the present value of wealth is negative then d must be chosen to be greater than zero. For decision periods prior to the final period, one must introduce enough assumptions on expectations and utility to insure that the consumer does not expect to be able to refinance all of a large current debt. (Green does this through his assumptions on u and Ψ.)

An obvious objection to this model is that default does not require a declaration of bankruptcy (i.e. one need not even partially pay one's creditors except through the economy wide returns ratios). This is also a feature of Stigum's model in which default is paid for by everyone. Green on the other hand extracts the ultimate penalty even if the consumer defaults on only $\varepsilon\%$ of his contracts. Again a middle road might help reestablish convexity of Green's demand correspondence while forcing the defaulter to bear more of the burden of his actions. One possibility is to require that some percentage of defaulted contracts be covered by the consumer's own assets. Remembering that d is the percentage defaulted on, let $t(d)$ be the institutionally predetermined percentage of assets required to cover default. Then, a consumer's budget constraint would be: $px \leq (1 - d)pe_{-} + (1 - t(d))(pre_{+} + pw)$. In order for a consumer to be able to always attain a non-negative wealth position we would need $t(d) < d$. If $t(d)$ is continuous and convex in d then demand correspondences should be well behaved. That is, under assumptions similar to Green's, demands and default ratios should be upper-semi continuous, convex, non-empty correspondences of prices and returns ratios. If the formula for computing the returns ratio is then suitably adjusted, one should be able to establish the existence of temporary equilibrium. This remains to be shown.

I have concentrated on the aspects of each paper dealing with the question of existence. However, once existence is established, it is interesting to inquire about its optimality properties. Stigum's paper is one of the few to do this for economies with sequences of temporary equilibria. His results (embodied in theorems 11.3 and 11.4 and some counterexamples) deserve emphasizing. Briefly, he shows that, in general, a particular sequence of temporary equilibria is not Pareto-optimal *ex post*.

However, he does provide, without proof, very restrictive conditions such that *ex post* Pareto optimality does obtain. He also demonstrates that there exists a set of expectations and a redistribution of first-period purchasing power such that any *ex post* Pareto-optimal allocation can be supported by a sequence of temporary equilibria. That is, optima are equilibria (given a redistribution of endowments and specific expectations) but equilibria may not be optima except in fortuitous circumstances. Finally, he indicates (in theorem 11.4) that each temporary equilibrium plan is *ex ante* Pareto-optimal given the expectations of consumers. Thus, he has illuminated the complex relationship between sequences of temporary equilibria, *ex ante* Pareto-optimal plans, *ex post* Pareto-optimal allocations, and price expectations.

In summary, I consider both of these papers to be excellent contributions to our knowledge about the performance of market economies in which temporal sequences of markets exist. Stigum's work establishes that in general, 'a price mechanism confined mainly to current markets for current goods is likely to go astray'. (This is a view attributed to Keynes by Arrow and Hahn [1, p. 347].) Green's work initiates the important task of revising the usual rules of market behavior to allow sequences of temporary equilibria to proceed in an orderly fashion. Clearly this work has just begun.

References

[1] Arrow, K. J. and Hahn, F. H. *General Competitive Analysis*. Holden Day: San Francisco (1971).
[2] Debreu, G. *Theory of Value*. John Wiley and Sons, Inc.: New York (1959).
[3] Debreu, G. New concepts and techniques for equilibrium analysis. *International Economic Review*, **3** (September 1962), 257–273.

CHAPTER 12

BALANCED GROWTH UNDER UNCERTAINTY IN DECOMPOSABLE ECONOMIES

Harry Kesten and Bernt P. Stigum*

12.1. *Introduction and Statement of Results*

In this paper we study the asymptotic properties of d-dimensional positive (note we use the term positive to mean non-negative throughout this paper)[1] vector-valued random processes $\{\boldsymbol{x}_t = (x_{t,1}, \ldots, x_{t,d})\}_{t \geq 0}$, which for $t = 0, 1, \ldots$ satisfy a set of functional equations of the form

$$\begin{aligned} E\{x_{t+1,i} \mid \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t\} &= h_i(x_{t,1}, \ldots, x_{t,d_1}),\ 1 \leq i \leq d_1, \\ E\{x_{t+1,i} \mid \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t\} &= h_i(\boldsymbol{x}_t),\ d_1 + 1 \leq i \leq d. \end{aligned} \tag{12.1}$$

Here $E\{x_{t+1,i} \mid \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t\}$ denotes the conditional expectation of $x_{t+1,i}$ given the values taken on by $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_t$, and it is assumed that each $h_i(\cdot)$ is a positive, continuous, non-decreasing linearly homogeneous function of its arguments[2]. We shall show that under certain conditions these processes will eventually adopt a balanced growth pattern; i.e. for some $\lambda > 1$, $\lim_{t\to\infty} \{\boldsymbol{x}_t/\lambda^t\}$ exists and has a fixed direction with probability 1.

The need to study such processes arises in many different contexts in economics. Here are two simple examples.

---

* Partial support from the National Science Foundation for the second author is gratefully acknowledged.

[1] Throughout this paper we say that a constant c is *positive* if $c \geq 0$. Moreover, we say that a vector $\boldsymbol{x} = (x(1), \ldots, x(d))$ is *positive* if $\boldsymbol{x} \geq 0$; i.e. if $x(i) \geq 0$ for all $i = 1, \ldots, d$. Finally, we denote the set of all positive constants by R_+ and the set of all positive k-dimensional vectors by $(R_+)^k$, $k = 1, 2, \ldots$ with $(R_+)^1 = R_+$; i.e. $R_+ = \{c \geq 0\} = [0, \infty)$ and $(R_+)^k = \{\boldsymbol{x} = (x(1), \ldots, x(k)) \geq 0\}$.

[2] Throughout this paper a *positive function* on $(R_+)^k$ is a map from $(R_+)^k$ into R_+. More generally, a d-vector-valued *positive function* on $(R_+)^k$ is a map from $(R_+)^k$ into $(R_+)^d$.

12.1.1. A macro-model of an economy

Let $d = 2$. Moreover, let $x_{t,1}$ denote an economy's labor force in period t, $t = 0, 1, \ldots$, and assume that there exists a constant ρ such that, for all t,

$$E\{x_{t+1,1} \mid x_{0,1}, \ldots, x_{t,1}\} = \rho x_{t,1} \text{ w.pr. 1 } (\equiv \text{ with probability 1}). \quad (12.2)$$

Finally, let $x_{t,2}$, $t = 0, 1, \ldots$ denote the same economy's stock of capital in period t, and assume that there exists a positive, continuous, non-decreasing, linearly homogeneous function $h(\cdot)$ such that

$$E\{x_{t+1,2} \mid \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t\} = h(\boldsymbol{x}_t) \quad \text{w. pr. 1.} \quad (12.3)$$

Then $\boldsymbol{x}_t$, $t = 0, 1, \ldots$ can be seen to describe the behavior over time of an economy as it is reflected in the behavior of its stock of labor and capital.

A more specific interpretation of this model is obtained in the following special case. Replace eq. (12.2) by

$$x_{t+1,1} = \rho x_{t,1} \text{ w. pr. 1,} \quad (12.4)$$

and eq. (12.3) by

$$x_{t+1,2} = h(\boldsymbol{x}_t) \equiv x_{t,2} + sF(\boldsymbol{x}_t) \text{ w. pr. 1,} \quad (12.5)$$

where s denotes the consumers' average propensity to save and $F(\cdot)$ denotes the aggregate production function of the economy. The relations (12.4) and (12.5) represent a discrete analogue of Solow's classic model of economic growth [10, pp. 66–67] for an economy which behaves deterministically, that is, an economy which operates in a world in which the set of all possible states of the world consists of one and only one point. Eqs. (12.2) and (12.3) can therefore be viewed as a stochastic analogue of Solow's model.

Solow's result, when translated into discrete time, states that eqs. (12.4) and (12.5) imply for $\rho > 1$ that $\{\boldsymbol{x}_t/\rho^t\}$ converges to a fixed vector. Our theorem 12.1 generalizes this growth result from the deterministic to a stochastic context.

Another special case can be obtained as follows. Let A_t, $t = 0, 1, \ldots$ denote a sequence of identically and independently distributed random variables and suppose that the aggregate production function in period t can be represented by a continuous function $G(\boldsymbol{x}_t, A_t)$ which for each A_t is linearly homogeneous, and non-decreasing in $\boldsymbol{x}_t$. Moreover, suppose that there exists a constant s such that

$$x_{t+1,2} = x_{t,2} + sG(\boldsymbol{x}_t, A_t) \text{ w. pr. } 1. \tag{12.6}$$

Finally, suppose that the A_t's are independently distributed of the $\boldsymbol{x}_t$'s. Then the function $h(\cdot)$ in eq. (12.3) can be defined by the equation

$$h(\boldsymbol{x}_t) = x_{t,2} + sEG(\boldsymbol{x}_t, A_t), \tag{12.7}$$

where

$$EG(\boldsymbol{x}_t, A_t) = \int G(\boldsymbol{x}_t, y)\mathrm{d}F(y) \text{ and } F(y) = P\{A_t \leq y\};$$

i.e. $EG(\boldsymbol{x}_t, A_t)$ is the expectation of $G(\boldsymbol{x}_t, A_t)$ with respect to A_t.

The model obtained by substituting eq. (12.4) for eq. (12.2) and by retaining eq. (12.3) with $h(\cdot)$ defined as in eq. (12.7) represents a simplified version of a model of economic growth studied by Mirman (cf. [8], pp. 1–2, 3).

12.1.2. A dynamic model of a firm

Consider a closed-end investment company. Let $\boldsymbol{m} = (m_1, \ldots, m_{d_1})$ denote a vector each component of which represents a security in which the firm can invest; and let $x_{t,i}$, $i = 1, \ldots, d_1$ denote the t-period price of m_i. Then assume that the firm invests in equities only and that these equities either do not pay dividends or pay a fixed percentage of the current values of the shares. Assume also that the investment company does not pay its share holders dividends and that its manager in each period is paid a fixed percentage of the company's net worth. If these assumptions hold, then under reasonable conditions on the manager's risk preferences and price expectations the behavior over time of this firm as reflected in each period in the manager's choice of portfolio can be represented in the following way. Let $d = d_1 + 1$, let m_t denote the firm's portfolio at the beginning of period t, and let

$$x_{t,d} = \sum_{i=1}^{d_1} x_{t,i} m_{t,i}$$

denote the fund's net worth at the beginning of period t. Then there exists a continuous vector-valued function $\boldsymbol{m}(\cdot)$ that is homogeneous of degree zero and that for each $t \geq 0$ satisfies the equation

$$m_{t+1} = \boldsymbol{m}(\boldsymbol{x}_t).^3 \tag{12.8}$$

---

[3] We assume here both that the manager always expects prices to be positive and that prices in fact are positive with probability 1.

Suppose now that the $x_{t,i}$s, $1 \leq i \leq d_1$, are distributed independently of the m_ts and that there exist positive, continuous, linearly homogeneous functions $h_i(\cdot)$, $i = 1, \ldots, d_1$, such that for $1 \leq i \leq d_1$,

$$E\{x_{t+1,i} \mid \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t\} = h_i(x_{t,1}, \ldots, x_{t,d_1}) \text{ w. pr. 1.} \tag{12.9}$$

Then

$$\begin{aligned} E\{\boldsymbol{x}_{t+1,d} \mid \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t\} &= E\{\sum_{i=1}^{d_1} x_{t+1,i} m_{t+1,i} \mid \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t\} \\ &= \sum_{i=1}^{d_1} h_i(x_{t,1}, \ldots, x_{t,d_1}) m_i(\boldsymbol{x}_t) \equiv h_d(\boldsymbol{x}_t) \text{ w. pr. 1.,} \end{aligned} \tag{12.10}$$

where $h_d(\cdot)$ is a positive, continuous, linearly homogeneous function of its arguments.

From eqs. (12.9) and (12.10) it follows that, under certain conditions on the manager's risk preferences and price expectations and under certain conditions on the actual behavior of prices, eq. (12.9), we can study the growth of a closed-end investment company in terms of a model such as the one proposed in eq. (12.1).

The preceding examples pertain to economies and firms that operate in an uncertain world. For an interesting discussion of how the need to analyze solutions to deterministic analogues of eq. (12.1) arises in the study of economies that operate under certainty, the reader is referred to Fisher's fundamental paper on 'Decomposability, near decomposability and balanced price change under constant returns to scale' [4] (*cf.* in particular pp. 67–70). Other examples of deterministic analogues to eq. (12.1) can be found in H. Nikaido's discussion of linear and non-linear income propagation models (cf. ref. [9], pp. 98–100 and pp. 162–163)[4].

So much for examples, now on to our results. Note that we always take $d_1 \geq 1$. If $d_1 = d$, the situation reduces to the indecomposable one, which is treated in ref. [12]. This case could be subsumed under the present one, but it would make the write-up of several of our results cumbersome. We therefore have tacitly taken $d > d_1$ and the reader is referred to ref. [12] for details in the case $d = d_1$.

We begin by stating two definitions and two well-known lemmas. Throughout, $\boldsymbol{x}$ and $\boldsymbol{y}$ stand for the vectors $(x(1), \ldots, x(d))$, $(y(1), \ldots, y(d))$,

[4] For a systematic presentation of deterministic models of growth in decomposable economies the reader is referred to E. Burmeister and A. Rodney Dobell's book on *Mathematical Theories of Economic Growth* [2].

and $H(.) = (H_1(.), \ldots, H_d(.))$ is a map from $(R_+)^d = \{x = (x(1), \ldots, x(d)) \geq 0\}$ into $(R_+)^d$. Moreover, for $x, y \in (R_+)^d$, $N(x, y) = \{i: x(i) > y(i)\}$. If $x \in (R_+)^d$, we shall use x^1 for the d_1-vector $x^1 = (x(1), \ldots, x(d_1))$ and x^2 for the d-d_1-vector $x^2 = (x(d_1 + 1), \ldots, x(d))$. The vector x itself will often be written as $x = (x^1, x^2)$.

DEFINITION 12.1. The vector-valued function $H(\cdot)$ is called *indecomposable* if for any $x, y \in (R_+)^d$, with $x \geq y \geq 0$[5] and $\varnothing \neq N(x, y) \subset \{1, \ldots, d\}$ we have

$$H_i(x) \neq H_i(y) \text{ for some } i \notin N(x, y). \tag{12.11}$$

Moreover, $H(\cdot)$ is called *decomposable* if and only if it is not indecomposable.

Clearly a function $h(\cdot)$, with $h_i(\cdot)$ depending only on $x(1), \ldots, x(d_1)$ for $1 \leq i \leq d_1 < d$ as in eq. (12.1) is decomposable. This is the reason for the 'decomposable economies' in the title of this paper.

DEFINITION 12.2. We say that λ is an *eigenvalue* of $H(\cdot)$ if there exists an $x \in (R_+)^d$, $x \neq 0$ such that

$$H(x) = \lambda x. \tag{12.12}$$

A vector x satisfying these conditions for some λ is called an *eigenvector*.

For later purposes there are a few things to note about definition 12.2. First, the zero vector is not an eigenvector of $H(\cdot)$. Second, an eigenvector has to be a positive vector since $H(\cdot)$ is only defined on $(R_+)^d$. Lastly, the positivity of $H(\cdot)$ implies that any eigenvalue λ is positive (and *a fortiori* real).

LEMMA 12.1 (Samuelson and Solow). *Let*

$$H^1: (R_+)^{d_1} \to (R_+)^{d_1}$$

be a positive, continuous, non-decreasing function, which is homogeneous of degree one and indecomposable. Then there exists a unique eigenvalue λ_1 *with a corresponding eigenvector* V^1, *i.e.*

$$V^1 \in (R_+)^{d_1},\ V^1 \neq 0,\ H^1(V^1) = \lambda_1 V^1. \tag{12.13}$$

V^1 *is strictly positive and unique up to a positive multiplicative factor.*

---

[5] If $x = (x(1), \ldots, x(k))$ and $y = (y(1), \ldots, y(k))$, then $x \geq y$ means that $x(i) \geq y(i)$ for all $i = 1, \ldots, k$. Moreover, $x > y$ means that $x(i) > y(i)$ for all $i = 1, \ldots, k$.

PROOF. This lemma is an immediate consequence of theorems 10.1 and 10.4 in ref. [9, pp. 151 and 156] (see also Samuelson and Solow's proof, [11, pp. 415–416]).

If $\boldsymbol{H}^2(\cdot) = (H_{d_1+1}(\cdot), \ldots, H_d(\cdot))$ is a map from $(R_+)^d$ into $(R_+)^{d-d_1}$, which is positive, continuous and non-decreasing and such that the map $\boldsymbol{x}^2 \to \boldsymbol{H}^2(0, \boldsymbol{x}^2)$ is indecomposable (when viewed as a map from $(R_+)^{d-d_1}$ into $(R_+)^{d-d_1}$), then an analogue of lemma 12.1 shows that there exists a unique $\lambda_2 \geq 0$, $\bar{\boldsymbol{V}}^2 \in (R_+)^{d-d_1}$, $\bar{\boldsymbol{V}}^2 > 0$ such that

$$\boldsymbol{H}^2(0, \bar{\boldsymbol{V}}^2) = \lambda_2 \bar{\boldsymbol{V}}^2. \tag{12.14}$$

The following lemma shows that we can also find an eigenvector for

$$\boldsymbol{H} = (\boldsymbol{H}^1, \boldsymbol{H}^2): (R_+)^d \to (R_+)^d, \tag{12.15}$$

i.e. $H_i(\cdot) = H_i^1(\cdot)$ if $1 \leq i \leq d_1$, $H_i(\cdot) = H_i^2(\cdot)$ if $d_1 + 1 \leq i \leq d$. This lemma was proved by Fisher [4, pp. 79–81] under slightly stronger monotonicity assumptions on $\boldsymbol{H}(\cdot)$. We shall not, however, give a proof here since the gain in generality over Fisher's version is very slight.

LEMMA 12.2 (Fisher). *Let $\boldsymbol{H}^1: (R_+)^{d_1} \to (R_+)^{d_1}$ and $\boldsymbol{H}^2: (R_+)^d \to (R_+)^{d-d_1}$ be positive, continuous, non-decreasing and homogeneous of degree one. Assume that $\boldsymbol{H}^1(\cdot)$ and the map $\boldsymbol{x}^2 \to \boldsymbol{H}^2(0, \boldsymbol{x}^2)$ are indecomposable, that $\lambda_1, \lambda_2, \boldsymbol{V}_1, \bar{\boldsymbol{V}}^2$ satisfy eqs.* (12.13) *and* (12.14) *and that $\lambda_1 > \lambda_2$. Finally, let $\boldsymbol{H}$ be as in eq.* (12.15) *and assume that for any $\boldsymbol{x} \geq \boldsymbol{y}$ with $N(\boldsymbol{x}, \boldsymbol{y}) \neq \varnothing$, $[d_1 + 1, d]$ not contained in $N(\boldsymbol{x}, \boldsymbol{y})$, one has*

$$H_i^2(\boldsymbol{x}) > H_i^2(\boldsymbol{y}) \qquad \text{for some } i \in [d_1 + 1, d],\ i \notin N(\boldsymbol{x}, \boldsymbol{y}). \tag{12.16}$$

Then, there exists a unique vector $\boldsymbol{V}^2 = (V^2(d_1 + 1), \ldots, V^2(d)) \in (R_+)^{d-d_1}$ such that

$$\boldsymbol{V}^2 \neq 0 \text{ and } \boldsymbol{H}^2(\boldsymbol{V}^1, \boldsymbol{V}^2) = \lambda_1 \boldsymbol{V}^2. \tag{12.17}$$

This $\boldsymbol{V}^2$ is strictly positive.

If the conditions of lemma 12.2 hold, we shall write $\boldsymbol{W} = (\boldsymbol{V}^1, \boldsymbol{V}^2)$. By eqs. (12.13) and (12.17), $\boldsymbol{W} \in (R_+)^d$ is an eigenvector of $\boldsymbol{H}$ with eigenvalue λ_1, i.e.

$$\boldsymbol{H}(\boldsymbol{W}) = \lambda_1 \boldsymbol{W}. \tag{12.18}$$

Notice that $\lambda_1 > \lambda_2 \geq 0$ implies that λ_1 is strictly positive. Moreover, any $\boldsymbol{W}$ which satisfies $\boldsymbol{W} \in (R_+)^d$, $\boldsymbol{W}^1 \neq 0$, $\boldsymbol{H}(\boldsymbol{W}) = \lambda \boldsymbol{W}$ for some λ, also satisfies $\boldsymbol{H}^1(\boldsymbol{W}) = \boldsymbol{H}^1(\boldsymbol{W}^1) = \lambda \boldsymbol{W}^1$ so that $\boldsymbol{W}^1$ must be a strictly positive multiple of $\boldsymbol{V}^1$ and $\lambda = \lambda_1$. Thus, by lemma 12.2, $\boldsymbol{W}$ is the unique (up to a multiplicative constant) eigenvector of $\boldsymbol{H}$ with $\boldsymbol{W}^1 \neq 0$. Moreover, $\boldsymbol{W}$ is strictly positive, i.e.

$$\boldsymbol{W} > 0. \tag{12.19}$$

We introduce a last bit of notation in order to formulate our theorems:

$$|\boldsymbol{x}| = \sum_{i=1}^{d} |x(i)| \text{ for } \boldsymbol{x} \in R^d,^{6}$$

$$A = \{\boldsymbol{x} \in (R_+)^d : |\boldsymbol{x}| = 1\}, \tag{12.20}$$

and if $\boldsymbol{h}(\cdot)$ is the function in eq. (12.1), $h_{ij}(\boldsymbol{x}) = (\partial h_i(\boldsymbol{x})/\partial x_j)$.

THEOREM 12.1 ($\lambda_1 > \lambda_2$). *Assume that eq.* (12.1) *holds and that the functions* $\boldsymbol{H}(\cdot)$, $\boldsymbol{H}^1(\cdot)$, *and* $\boldsymbol{H}^2(\cdot)$ *defined by eq.* (12.15) *and*

$$H_i^1(\cdot) = h_i(\cdot),\ 1 \leq i \leq d_1$$

and

$$H_i^2(\cdot) = h_i(\cdot),\ d_1 + 1 \leq i \leq d \tag{12.21}$$

satisfy the conditions of lemma 12.2. *Let* $\boldsymbol{W}$ *be given by eqs.* (12.13), (12.17) *and* (12.18). *Suppose* $h_{ij}(\boldsymbol{W})$ *exists and suppose*

$$h_{ij}(\boldsymbol{W}) > 0 \text{ if } 1 \leq i, j \leq d_1;\ d_1 < i \leq d,\ 1 \leq j \leq d. \tag{12.22}$$

Moreover, suppose that

$$\lambda_1 > 1, 0 < \lambda_2 < \lambda_1,^{7} \tag{12.23}$$

and that

$$\boldsymbol{H}^1(\boldsymbol{x}^1) > 0 \textit{ and } \boldsymbol{H}^2(\boldsymbol{x}) > 0 \textit{ whenever } \boldsymbol{x}^1 \neq 0. \tag{12.24}$$

Finally, assume that there exists a neighborhood U *of* $\tilde{\boldsymbol{W}} = (\boldsymbol{W}/|\boldsymbol{W}|)$ *in* A *and constants* K *and* δ *such that* $0 < \delta < 1$ *and such that*

---

[6] Here $R^d = \{\boldsymbol{x} = (x(1), \ldots, x(d)) : -\infty < x(i) < \infty, i = 1, \ldots, d\}$.

[7] $\lambda_2 > 0$ is not much of an extra condition since $\boldsymbol{x}^2 \to \boldsymbol{H}(\boldsymbol{0}, \boldsymbol{x}^2)$ is already indecomposable. Thus $\lambda_2 = 0$ can occur only if $d = d_1 + 1$ and $\boldsymbol{H}(\boldsymbol{0}, \boldsymbol{x}^2) \equiv 0$.

$$E\{|x_{t+1,i} - h_i(x_t^1)|^2 | x_0, \ldots, x_t\} \le K |x_t^1|^{2(1-\delta)} \quad 1 \le i \le d_1; \tag{12.25}$$

$$E\{|x_{t+1,i} - h_i(x_t)|^2 | x_0, \ldots, x_t\} \le K |x_t|^{2(1-\delta)} \textit{ for } d_1 + 1 \le i \le d; \tag{12.26}$$

and

$$|H(x) - H(y)| \le K |x - y| \text{ for all } x, y \in U. \tag{12.27}$$

Then there exists a random variable g such that

$$\lim_{t\to\infty} \{x_t/\lambda_1^t\} = g\tilde{W} \textit{ w. pr. } 1.^8 \tag{12.28}$$

Moreover,

$$E\{g | x_0\} > 0 \text{ and } P\{g > 0 | x_0\} > 0 \tag{12.29}$$

whenever $x_0^1 \ne 0, |x_0| \ge M(|x_0^1|/|x_0|)$ *for a suitable finite function* $M(\cdot)$: $(0, 1] \to (0, \infty)$. *If* $P\{(|x_s^1|/|x_s|) \ge \varepsilon, |x_s| \ge M(\varepsilon)$ *for some* $s | x_0\} > 0$ *for some* $\varepsilon > 0$ *and all* x_0 *with* $x_0^1 \ne 0$, *then eq.* (12.29) *holds for all* $x_0^1 \ne 0$. *On the other hand,*

$$x_t^1 = 0 \textit{ for all } t \textit{ and } g = 0 \textit{ a.e. on } \{x_0^1 = 0\}.^9 \tag{12.30}$$

Finally,

$$E\{g | x_0\} < \infty \tag{12.31}$$

if the functions $h_i(\cdot)$, $1 \le i \le d_1$, *are concave.*

THEOREM 12.2 ($\lambda_1 < \lambda_2$). *Assume that eq.* (12.1) *holds and let* $H(\cdot)$, $H^1(\cdot)$ *and* $H^2(\cdot)$ *be given by eqs.* (12.15) *and* (12.21). *Assume that* $H^1(\cdot)$ *and* $H^2(\cdot)$ *are positive, continuous, non-decreasing and homogeneous of degree one. Moreover, let* $H^1(\cdot)$ *and the map* $x^2 \to H^2(0, x^2)$ *be indecomposable. If* $\lambda_1, \lambda_2, V^1$ *and* $\bar{V}^2$ *are as in eqs.* (12.13) *and* (12.14), *assume that*

$$0 < \lambda_1 < \lambda_2, \lambda_2 > 1; \tag{12.32}$$

$$H^2(x) > 0 \textit{ for } x \in (R_+)^d, x \ne 0; \tag{12.33}$$

and

$$h_{ij}(0, \bar{V}^2) \equiv h_{ij}(x) \textit{ at } x = (0, \bar{V}^2) \tag{12.34}$$

exists and is strictly positive for $d_1 + 1 \le i, j \le d$.

[8] Here as well as in eq. (12.35) it is understood that $g < \infty$ w. pr. 1.
[9] Here a.e. means 'almost everywhere'.

Finally, assume that there exists a neighborhood U of $(0, (\bar{V}^2/|\bar{V}^2|)$ *in A and positive constants K and* δ *satisfying eqs.* (12.25)–(12.27). *Then there exists a random variable g such that*

$$\lim_{t\to\infty} \{x_t/\lambda_2^t\} = g\cdot(0, \bar{V}^2) \text{ w. pr. 1.} \tag{12.35}$$

Moreover, for some constant $M < \infty$,

$$E\{g\,|\,x_0\} > 0 \text{ and } P\{g > 0\,|\,x_0\} > 0 \tag{12.36}$$

whenever $|x_0| \geq M$. *If for all* $x_0 \neq 0$, $P\{|x_s| \geq M \text{ for some } s > 0\,|\,x_0\} > 0$, *then eq.* (12.36) *holds for all* $x_0 \neq 0$.

Finally,

$$E\{g\,|\,x_0\} < \infty \tag{12.37}$$

if all the $h_i(\cdot)$*s are concave.*

At first glance neither the generality of theorems 12.1 and 12.2 nor their applicability in economics is obvious. So a few remarks concerning the assumptions made in these theorems are called for.

12.1.2.1. Remark 1. When $\max(\lambda_1, \lambda_2) > 1$ and $\lambda_1 \neq \lambda_2$, theorems 12.1 and 12.2 provide what are, from one point of view, the best sufficient conditions obtainable for the validity of eqs. (12.28), (12.29) and (12.35), (12.36) respectively. Specifically, there exist processes that satisfy all but one of these conditions whose asymptotic behavior cannot be described by eqs. (12.28), (12.29) or (12.35), (12.36). Here is an example.

Consider Mirman's model as presented on pp. 336–337 above. Make assumptions on $G(\cdot)$ that will ensure that the conditions of lemma 12.2 and conditions (12.22) and (12.27) are satisfied. Assume also that the distribution of A_t is absolutely continuous with respect to Lebesgue measure and concentrated on a compact interval, and that the corresponding density function is continuous and positive on this interval. Then $\boldsymbol{H}(\cdot)$ satisfies eq. (12.24) and $x_{t,1}$, $t = 0, 1, \ldots$ satisfies eq. (12.25). However, $x_{t,2}$, $t = 0, 1, \ldots$ does not satisfy eq. (12.26), and it follows trivially from theorem 3.3 in ref. [8, p. III.10] that $\{x_t/\lambda^t\}$ converges in distribution to a random vector whose distribution is not concentrated along a single ray from the origin in $(R_+)^2$. Thus, for this model, eqs. (12.28), (12.29) are false.

12.1.2.2. Remark 2. On the other hand, there exist processes that satisfy some but not all the conditions of theorems 12.1 and 12.2 whose asymptotic behavior can be described as in eqs. (12.28), (12.29) or (12.35), (12.36). Here is an example.

Let $\{\boldsymbol{x}_t = (x_{t,1}, x_{t,2})\}_{t \geq 0}$ be a decomposable Galton–Watson process with first-moment matrix

$$\begin{pmatrix} \lambda_1 & 0 \\ \boldsymbol{m} & \lambda_2 \end{pmatrix}$$

and assume that $\lambda_1 > 1$, $\lambda_2 > 0$, $\lambda_1 > \lambda_2$ and $\boldsymbol{m} > 0$. Then $\boldsymbol{H}(\cdot)$, $\boldsymbol{H}^1(\cdot)$, and $\boldsymbol{H}^2(\cdot)$ satisfy the conditions of lemma 12.2 and conditions (12.22), (12.24) and (12.27) as well. Moreover – cf. theorem 2.1 in ref. [7] – eqs. (12.28), (12.29) are valid if and only if $E\{x_{1,1} \log x_{1,1} \mid x_{0,1} = 1\} < \infty$. Thus, for such a process eqs. (12.28), (12.29) may hold even if eqs. (12.25) and (12.26) are not satisfied.

12.1.2.3. Remark 3. It is not intuitively obvious that any economy ever would satisfy conditions (12.25) and (12.26). So we next describe a simple economy that does. This hypothetical economy can be thought of as a free translation of Edward Bellamy's USA, year 2000.[10]

Consider an economy in which there are two primary inputs, labor $\boldsymbol{x}_1$ and capital $\boldsymbol{x}_2$. These factors can be combined to produce output ($\equiv$ net national product) according to a continuous, strictly quasi-concave, linearly homogeneous function $F(\cdot)$ that is increasing on $\{\boldsymbol{x}_1 > 0,\ \boldsymbol{x}_2 \geq 0\}$. We assume (1) that capital and output are both publicly owned, (2) that workers share equally in national output, each one's share being equal to a fraction of labor's average product[11], and (3) that the general surplus (i.e. net national product – wage allotments) is used *in toto* by government to augment the nation's capital stock.

Assume in addition that the share of national output credited on the public books to each worker is so ample that a worker is 'more likely not to spend it all'[12]. If a worker does not fully expend his credit, the balance is turned into general surplus. Under extraordinary circumstances a worker might be allowed to spend more than his allotment but never more than labor's average product. The excess above the usual allotment would be taken out of the general surplus.

[10] See ref. [1].
[11] Cf. ref. [1, p. 151].
[12] Cf. ref. [1, p. 148].

In more precise terms we are assuming (4) that the fraction of labor's average product consumed in each period by the ith worker can be represented by a random variable c_i with range (0,1], and (5) that, if $x_{t,1}$ and $x_{t,2}$ denote labor and capital in period t, then for all t:

$$x_{t+1,2} = x_{t,2} + \sum_{i=1}^{x_{t,1}} (1 - c_i)\{F(x_{t,1}, x_{t,2})/x_{t,1}\}. \tag{12.38}$$

We will also assume (6) that the distribution of c_i is independent of i and constant over time and that, for each pair (i, j) c_i and c_j are distributed independently of each other and of labor and capital. Finally we assume (7) that the growth of the labor force can be represented by a Galton–Watson process with mean $\lambda > 1$ and finite variance σ^2.

The preceding assumptions allow us to describe the development over time of our utopian economy in terms of a random process $\{\boldsymbol{x}_t = (x_{t,1}, x_{t,2})\}_{t\geq 0}$ with the following properties:

$$E\{x_{t+1,1} \mid \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t\} = \lambda x_{t,1} \text{ w. pr. 1,} \tag{12.39}$$

and

$$E\{x_{t+1,2} \mid \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t\} = x_{t,2} + sF(\boldsymbol{x}_t) \text{ w. pr. 1,} \tag{12.40}$$

where $s = E(1 - c_i)$. Moreover,

$$E\{(x_{t+1,1} - \lambda x_{t,1})^2 \mid \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t\} = \sigma^2 x_{t,1} \text{ w. pr. 1,} \tag{12.41}$$

and

$$E\{(x_{t+1,2} - x_{t,2} - sF(\boldsymbol{x}_t))^2 \mid \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t\} \leq K \mid x_t \mid \text{ w. pr. 1} \tag{12.42}$$

for a suitable finite constant K.

12.1.2.4. Remark 4. Theorems 12.1 and 12.2 can be generalized in several directions. Here are two ways suggested by theorem 12.3 in section 12.3 below.

Monotonicity. The only purpose of the condition that $\boldsymbol{H}(\cdot)$ be nondecreasing is to guarantee the existence of an eigenvector $\boldsymbol{W}$ of $\boldsymbol{H}(\cdot)$ and an exponential convergence of the iterates of $\{\boldsymbol{H}(\boldsymbol{x})/|\boldsymbol{H}(\boldsymbol{x})|\}$ to $\{\boldsymbol{W}/|\boldsymbol{W}|\}$. We could use theorem 12.3 to obtain eqs. (12.28) and (12.29) even when $\boldsymbol{H}(\cdot)$ is not monotonic, as long as there exists a vector $\boldsymbol{V}$ for which eq. (12.70) holds.

Condition (12.1). The preceding observation also suggests that eqs.

(12.28) and (12.29) may be valid even if condition (12.1) is not satisfied. Specifically, to ensure the validity of eqs. (12.28) and (12.29) it is sufficient to assume that there exist functions $\boldsymbol{H}(\cdot)$, $\boldsymbol{H}^1(\cdot)$ and $\boldsymbol{H}^2(\cdot)$ that satisfy eq. (12.21), the conditions of lemma 12.2 and eqs. (12.22)–(12.27). To bring this point home, consider the following extension of the example presented in section 12.1.2.3. In that example we insisted that the c_i's be independently distributed and have range in (0,1]. These assumptions can be relaxed in many ways without stopping the economy from eventually achieving a balanced growth path. Here is one possibility. Let $y_1, y_2, \ldots$ be identically and independently distributed, non-negative random variables and assume that

$$0 < E\{y_i\} \equiv \mu < 1, \tag{12.43}$$

and

$$E\{(y_i - \mu)^2\} \equiv \sigma_y^2 < \infty. \tag{12.44}$$

Assume also that the y_i's are distributed independently of $\boldsymbol{x}_t$, $t \geq 0$, where $\boldsymbol{x}_t$ is as defined in section 12.1.2.3. Finally, assume that in each period t the distribution of the c_i's, conditional upon the observed value of $x_{t,1}$, satisfies the following condition:

$$P\{(c_1, \ldots, c_{x_{t,1}}) \in A\} = P\{(y_1, \ldots, y_{x_{t,1}}) \in A \mid \sum_{i=1}^{x_{t,1}} y_i \leq x_{t,1}\}, \tag{12.45}$$

for all Borel subsets A of $(R_+)^{x_{t,1}}$.

When we make the above modification, but leave the model of section 12.1.2.3 otherwise unchanged, the relations (12.39) and (12.41) are still valid. Moreover, with $s = (1 - \mu)$ it is fairly easy to show that eq. (12.42) is still valid. However, eq. (12.40) is generally false unless the range of the y_is belong to (0,1]. Thus $\{x_{t,1}, x_{t,2})\}_{t \geq 0}$ need not satisfy eq. (12.1). Yet, with the help of theorem 12.3, the same arguments used to establish theorem 12.1 can be used verbatim to establish the validity of eqs. (12.28) and (12.29) for this economy.

12.1.2.5. Remark 5. Theorems 12.1 and 12.2 consider only decomposable economies for which $\lambda_1 \neq \lambda_2$. The case when $\lambda_1 = \lambda_2$ is more difficult. The behavior of $\boldsymbol{x}_t^1$ can still be characterized as in theorem 3 in ref. [12], but the behavior of $\boldsymbol{x}_t^2$ is not so easily determined. To show how 'badly behaved' $\boldsymbol{x}_t^2$ may be, we present in section 12.2 several results pertaining to the two-dimensional deterministic analogue of eq. (12.1) where $\lambda_1 = \lambda_2 > 0$. Specifically, in lemma 12.10 we show that, if

$$\liminf_{\varepsilon \to 0} \{[\boldsymbol{h}(\varepsilon,1) - \boldsymbol{h}(0,1)]/\varepsilon\} > 0,$$

then $\{x_{t,2}/\lambda_1^t\}$ grows at least as fast as t. In lemma 12.8 we give an example in which $\{x_{t,2}/\lambda_1^t\}$ grows faster than any power of t. In lemma 12.10 we show that, for all $\delta > 0$,

$$\limsup_{t \to \infty} \{x_{t,2}/(\lambda_1 + \delta)^t\} = 0.$$

Finally, in lemma 12.9 we show that if there exist constants K and γ such that $0 < \gamma < 1$ and such that $\boldsymbol{h}(\varepsilon,1) - \boldsymbol{h}(0,1) \leq K\varepsilon^\gamma$ for all $0 < \varepsilon \leq 1$, then

$$\limsup_{t \to \infty} \{x_{t,2}/t^{(1/\gamma)} \lambda_1^t\} < \infty.$$

For brevity's sake we have omitted the proofs of these results.

12.2. Auxiliary Lemmas

In this section we state and prove several lemmas that we need to establish theorems 12.1 and 12.2 as stated above. In reading it, note that $\boldsymbol{h}(\cdot) = (h_1(\cdot), \ldots, h_d(\cdot))$, where each $h_i(\cdot)$ is as defined in eq. (12.1), and that $\boldsymbol{h}^t(\cdot)$ denotes the tth iterate of $\boldsymbol{h}(\cdot)$. Note also that $\boldsymbol{x} = (x(1), \ldots, x(d)) \in (R_+)^d$ and that we frequently write $\boldsymbol{x} = (\boldsymbol{x}^1, \boldsymbol{x}^2)$, where

$$\boldsymbol{x}^1 = (x(1), \ldots, x(d_1)) \in (R_+)^{d_1},$$

and

$$\boldsymbol{x}^2 = (x(d_1 + 1), \ldots, x(d)) \in (R_+)^{d-d_1}.$$

In this notation

$$\boldsymbol{h}(\boldsymbol{x}) = (h_1(\boldsymbol{x}^1), \ldots, h_{d_1}(\boldsymbol{x}^1), h_{d_1+1}(\boldsymbol{x}), \ldots, h_d(\boldsymbol{x})).$$

Finally note that the numbers λ_1 and λ_2 and the vectors $\boldsymbol{V}^1$, $\bar{\boldsymbol{V}}^2$ and $\boldsymbol{W}$ are assumed to satisfy the equations

$$h_i(\boldsymbol{V}^1) = \lambda_1 V^1(i),\ 1 \leq i \leq d_1; \tag{12.46a}$$

$$h_i(0, \bar{\boldsymbol{V}}^2) = \lambda_2 \bar{V}^2(i),\ d_1 + 1 \leq i \leq d; \tag{12.46b}$$

and

$$h_i(\boldsymbol{W}) = \lambda_1 W(i),\ 1 \le i \le d. \tag{12.46c}$$

LEMMA 12.3. *Let $\boldsymbol{h}(\cdot)$ and $\boldsymbol{h}^t(\cdot)$ be as defined above and assume that the $h_i(\cdot)$'s are positive, continuous, non-decreasing and homogeneous of degree one. Moreover, assume that there exist a strictly positive constant λ_1 and a strictly positive vector $\boldsymbol{W}$ that satisfy eq.* (12.46c). *Finally, assume also that $\boldsymbol{h}(\boldsymbol{x}) > 0$ whenever $\boldsymbol{x}^1 \neq 0$ and that $h_{ij}(\boldsymbol{W}) = (\partial h_i(\boldsymbol{x})/\partial x_j)_{\boldsymbol{x}=\boldsymbol{W}}$ exists with $h_{ij}(\boldsymbol{W}) > 0$ for $1 \le i, j \le d_1$ and $d_1 + 1 \le i \le d$, $1 \le j \le d$. Then for each ε, $0 < \varepsilon < (|\boldsymbol{W}^1|/|\boldsymbol{W}|)$, there exist finite positive constants K_1 and α and a positive, continuous, non-decreasing, linearly homogeneous function $\gamma(\cdot)$ on $(R_+)^{d_1}$ such that $0 < \alpha < 1$ and such that for all $t \ge 1$ and all $\boldsymbol{x} \ge 0$, $\boldsymbol{x} \neq 0$ for which $(|\boldsymbol{x}^1|/|\boldsymbol{x}|) \ge \varepsilon$,*

$$|\{\boldsymbol{h}^t(\boldsymbol{x})/\lambda_1^t\} - \gamma(\boldsymbol{x}^1)\boldsymbol{W}| \le K_1\alpha^t|\boldsymbol{x}|. \tag{12.47}$$

K_1 depends on ε, but both $\gamma(\cdot)$ and α can be chosen to be independent of ε. Also $\gamma(\boldsymbol{x}^1) > 0$ if $|\boldsymbol{x}^1| > 0$.

PROOF. Let

$$\alpha(t, \boldsymbol{x}) = \min_{1 \le i \le d} \{h_i^t(\boldsymbol{x})/\lambda_1^t W(i)\},$$

and

$$\beta(t, \boldsymbol{x}) = \max_{1 \le i \le d} \{h_i^t(\boldsymbol{x})/\lambda_1^t W(i)\}, \tag{12.48}$$

where $h_i^t(\cdot)$ denotes the ith component of $\boldsymbol{h}^t(\cdot)$. Since

$$\alpha(t, \boldsymbol{x})\lambda_1^t \boldsymbol{W} \le \boldsymbol{h}^t(\boldsymbol{x}) \le \beta(t, \boldsymbol{x})\lambda_1^t \boldsymbol{W},$$

$$\alpha(t, \boldsymbol{x}) = \{ \min_{1 \le i \le d} \{h_i(\alpha(t, \boldsymbol{x})\lambda_1^t \boldsymbol{W})/\lambda_1^{t+1} W(i)\} \le \min_{1 \le i \le d} \{h_i(\boldsymbol{h}^t(\boldsymbol{x}))/\lambda_1^{t+1} W(i)\},$$

and

$$\beta(t, \boldsymbol{x}) = \max_{1 \le i \le d} \{h_i(\beta(t, \boldsymbol{x})\lambda_1^t \boldsymbol{W})/\lambda_1^{t+1} W(i)\} \ge \max_{1 \le i \le d} \{h_i(\boldsymbol{h}^t(\boldsymbol{x}))/\lambda_1^{t+1} W(i)\}. \tag{12.49}$$

From eq. (12.49) it follows that

$$\alpha(1, \boldsymbol{x}) \le \ldots \le \alpha(t, \boldsymbol{x}) \le \alpha(t+1, \boldsymbol{x}) \le \ldots \le \beta(t+1, \boldsymbol{x}) \le \beta(t, \boldsymbol{x}) \le \ldots \le \beta(1, \boldsymbol{x}). \tag{12.50}$$

Hence, for each $x \geq 0$ the limits

$$\alpha(x) = \lim_{t \to \infty} \alpha(t, x),$$

and

$$\beta(x) = \lim_{t \to \infty} \beta(t, x) \qquad (12.51)$$

exist and $\alpha(x) \leq \beta(x)$.

Next, let $\varepsilon > 0$ be fixed so that $\varepsilon < (|W^1|/|W|)$, and note that (by eq. (12.50)) on the set where $x \geq 0$ and $(|x^1|/|x|) \geq \varepsilon$,

$$\begin{aligned}[(\beta(t, x) - \alpha(t, x))/\alpha(t, x)] &\leq K_2(\beta(t, \tilde{x}) - \alpha(t, \tilde{x})) \\ &\leq K_2 \max_{\substack{1 \leq i \leq d \\ |y| = 1}} \{h_i(y)/\lambda_1 W(i)\} < \infty, \end{aligned} \qquad (12.52)$$

where $\tilde{x} = (x/|x|)$ and

$$K_2 = \{\min_{\{x \geq 0, |x| = 1, |x^1| \geq \varepsilon\}} \alpha(1, x)\}^{-1} < \infty. \qquad (12.53)$$

Note also that for some sufficiently small positive $\varepsilon_1 < 1$ and for all non-zero vectors $y = (y^1, y^2)$ such that $0 \leq y \leq \varepsilon_1 W$, and $(|y^1|/|y|) \geq \varepsilon$,

$$h(W + y) = h(W) + h'(W)y + o(y) \geq h(W) + \tfrac{1}{2}h'(W)y, \qquad (12.54)$$

and

$$h(W - y) = h(W) - h'(W)y + o(y) \leq h(W) - \tfrac{1}{2}h'(W)y, \qquad (12.55)$$

where $h'(W) \equiv \{h_{ij}(W)\}$.

Finally, note that by virtue of (12.52) there exists an $0 < \varepsilon_2 < 1$, depending on ε only, such that for all $t \geq 1$ and $x \geq 0$ with $(|x^1|/|x|) \geq \varepsilon$,

$$0 \leq \varepsilon_2[(\beta(t, x) - \alpha(t, x))/\alpha(t, x)]\, W(i)e_i \leq \varepsilon_1 W,\ i = 1, \ldots, d_1, \qquad (12.56)$$

where $e_i = (0, \ldots, 0, e_i(i) = 1, 0, \ldots, 0)$ is a d-dimensional vector.

Now fix $t \geq 1$ and $x \geq 0$ with $(|x^1|/|x|) \geq \varepsilon$ and note that at least one of the following assertions must be true:

$$\{h^t_{i_0}(x)/\lambda^t_1 W(i_0)\} \geq \alpha(t, x) + \tfrac{1}{2}(\beta(t, x) - \alpha(t, x)) \quad \text{for some } 1 \leq i_0 \leq d_1; \qquad (12.57)$$

$$\{h^t_i(x)/\lambda^t_1 W(i)\} \leq \beta(t, x) - \tfrac{1}{2}(\beta(t, x) - \alpha(t, x)) \quad \text{for all } 1 \leq i \leq d_1. \qquad (12.58)$$

Suppose that (12.57) holds. Then (12.56) and (12.54) imply that

$$\begin{aligned}\{\boldsymbol{h}^{t+1}(\boldsymbol{x})/\lambda_1^{t+1}\} &\geq (1/\lambda_1)\boldsymbol{h}(\alpha(t,\boldsymbol{x})\boldsymbol{W} + (\varepsilon_2/2)(\beta(t,\boldsymbol{x}) - \alpha(t,\boldsymbol{x}))W(i_0)\boldsymbol{e}_{i_0})\\ &\geq \alpha(t,\boldsymbol{x})\boldsymbol{W} + (\varepsilon_2/4\lambda_1)(\beta(t,\boldsymbol{x}) - \alpha(t,\boldsymbol{x}))\boldsymbol{h}'(\boldsymbol{W})W(i_0)\boldsymbol{e}_{i_0}\\ &\geq \alpha(t,\boldsymbol{x})\boldsymbol{W} + (\delta\varepsilon_2/4\lambda_1)(\beta(t,\boldsymbol{x}) - \alpha(t,\boldsymbol{x}))\boldsymbol{W} \qquad (12.59)\end{aligned}$$

for $\delta = \min_{1 \leq i \leq d,\, 1 \leq i_0 \leq d_1} \{(\boldsymbol{h}'(\boldsymbol{W})W(i_0)e_{i_0})_i/W(i)\} > 0$. From (12.59) and (12.50) it follows that

$$\beta(t+1,\boldsymbol{x}) - \alpha(t+1,\boldsymbol{x}) \leq (\beta(t,\boldsymbol{x}) - \alpha(t,\boldsymbol{x}))(1 - (\delta\varepsilon_2/4\lambda_1)). \quad (12.60)$$

Suppose next that (12.58) holds. Then (12.55) and (12.56) imply that for any $1 \leq i_0 \leq d_1$,

$$\begin{aligned}\{\boldsymbol{h}^{t+1}(\boldsymbol{x})/\lambda_1^{t+1}\} &\leq (1/\lambda_1)\boldsymbol{h}(\beta(t,\boldsymbol{x})\boldsymbol{W} - (\varepsilon_2/2)(\beta(t,\boldsymbol{x}) - \alpha(t,\boldsymbol{x}))W(i_0)\boldsymbol{e}_{i_0})\\ &\leq \beta(t,\boldsymbol{x})\boldsymbol{W} - (\varepsilon_2/4\lambda_1)(\beta(t,\boldsymbol{x}) - \alpha(t,\boldsymbol{x}))\boldsymbol{h}'(\boldsymbol{W})W(i_0)\boldsymbol{e}_{i_0}\\ &\leq \beta(t,\boldsymbol{x})\boldsymbol{W} - (\delta\varepsilon_2/4\lambda_1)(\beta(t,\boldsymbol{x}) - \alpha(t,\boldsymbol{x}))\boldsymbol{W}. \qquad (12.61)\end{aligned}$$

From (12.61) and (12.50) it again follows that (12.60) is valid.

The preceding two paragraphs show that (12.60) holds for all $\boldsymbol{x} \geq 0$ with $(|\boldsymbol{x}^1|/|\boldsymbol{x}|) \geq \varepsilon$, and for all $t \geq 1$ as well. Hence – by simple iteration – we can deduce from (12.60) that for all $t \geq 1$ and all $\boldsymbol{x} \geq 0$ with $(|\boldsymbol{x}^1|/|\boldsymbol{x}|) \geq \varepsilon$,

$$\begin{aligned}(\beta(t+1,\boldsymbol{x}) - \alpha(t+1,\boldsymbol{x})) &\leq (\beta(1,\boldsymbol{x}) - \alpha(1,\boldsymbol{x}))(1 - (\delta\varepsilon_2/4\lambda_1))^t\\ &\leq K_3(1 - (\delta\varepsilon_2/4\lambda_1))^t|\boldsymbol{x}|, \qquad (12.62)\end{aligned}$$

where

$$K_3 = \max_{\{x \geq 0, |x| = 1\}} \{\beta(1,\boldsymbol{x}) - \alpha(1,\boldsymbol{x})\}.$$

From (12.62) and (12.49) it follows that, if we let

$$\gamma(\boldsymbol{x}) \equiv \alpha(\boldsymbol{x}), \quad (12.63)$$

then on the set where $\boldsymbol{x} \geq 0$ and $(|\boldsymbol{x}^1|/|\boldsymbol{x}|) \geq \varepsilon$

$$\begin{aligned}|\{\boldsymbol{h}^t(\boldsymbol{x})/\lambda_1^t\} - \gamma(\boldsymbol{x})\boldsymbol{W}| &\leq |\boldsymbol{W}|(\beta(t,\boldsymbol{x}) - \alpha(t,\boldsymbol{x}))\\ &\leq (K_3|\boldsymbol{W}|)(1 - (\delta\varepsilon_2/4\lambda_1))^{t-1}|\boldsymbol{x}|,\ t \geq 1. \qquad (12.64)\end{aligned}$$

This shows the existence of the required α, K_1 and $\gamma(\cdot)$, but with α depending on ε and $\gamma(\cdot)$ possibly depending on $\boldsymbol{x}^2$.

Next we will show that $\gamma(\cdot)$ has all the required properties. Evidently, $\gamma(\cdot)$ is positive, non-decreasing and homogeneous of degree one. So it suffices to show that $\gamma(\cdot)$ is a function of $\boldsymbol{x}^1$ alone and that it is continuous. To do that observe first that our choice of ε above was arbitrary. Therefore (12.64) implies that for all $\boldsymbol{x} \geq 0$ with $\boldsymbol{x}^1 \neq 0$,

$$\gamma(\boldsymbol{x})W(i) = \lim_{t \to \infty} \{h_i^t(\boldsymbol{x})/\lambda_1^t\}, \; i = 1, \ldots, d. \tag{12.65}$$

Since $h_i^t(\cdot)$ for $i = 1, \ldots, d_1$ is a function of $\boldsymbol{x}^1$ only and since $\boldsymbol{W} > 0$, eq. (12.65) in turn implies that, for all $\boldsymbol{x} \geq 0$ with $\boldsymbol{x}^1 \neq 0$,

$$\gamma(\boldsymbol{x}) = \gamma(\boldsymbol{x}^1, 0). \tag{12.66}$$

Also observe that (12.64) implies that on the set $B = \{\boldsymbol{x} \geq 0, \, |\boldsymbol{x}| = |\boldsymbol{x}^1| = 1\}$ $\alpha(\boldsymbol{x}^1, 0) = \beta(\boldsymbol{x}^1, 0)$. Hence on B, $\gamma(\cdot)$ is the limit of both an increasing and a decreasing sequence of continuous functions. From this and from theorem 33 in ref. [5, p. 130] it follows that $\gamma(\cdot)$ is continuous on B. But if that is so, then the fact that $\gamma(\cdot)$ is homogeneous of degree one, and eq. (12.66), imply that $\gamma(\cdot)$ is continuous on the set $\{\boldsymbol{x} \geq 0, \, |\boldsymbol{x}^1| > 0\}$. By virtue of the homogeneity of $\gamma(\cdot)$ we can extend $\gamma(\cdot)$ continuously to all of $\{\boldsymbol{x} \geq 0\}$ by putting $\gamma(0) = 0$.

Finally, observe that $\gamma(\boldsymbol{x}) = \gamma(\boldsymbol{x}^1, 0) \geq \alpha(1, (\boldsymbol{x}^1, 0)) > 0$ if $\boldsymbol{x}^1 \neq 0$. Thus $\gamma(\cdot)$ has all the required properties as was to be shown.

To show that α can be chosen independently of ε proceed as follows. First, observe that by eq. (12.65)

$$\gamma(\boldsymbol{h}^t(\boldsymbol{x})) = \lambda_1^t \, \gamma(\boldsymbol{x}). \tag{12.67}$$

Next, let $0 < \varepsilon^0 < (|\boldsymbol{W}^1|/|\boldsymbol{W}|)$ be fixed and for this ε^0 let

$$\alpha_1 = (1 - (\delta\varepsilon_2(\varepsilon^0)/4\lambda_1)). \tag{12.68}$$

Finally, observe that, if $\boldsymbol{x} \geq 0$ and $(|\boldsymbol{x}^1|/|\boldsymbol{x}|) \geq \varepsilon > 0$, then (12.64), (12.65) and $\gamma(\boldsymbol{x}) > 0$ imply that there exists a smallest integer t_0 – depending on ε but not on $\boldsymbol{x}$ – such that for all $t \geq t_0$,

$$\left\{ \sum_{i=1}^{d_1} h_i^t(\boldsymbol{x}^1)/|\boldsymbol{h}^t(\boldsymbol{x})| \right\} \geq \varepsilon^0.$$

Hence, for all $s \geq t_0$ and $\boldsymbol{x} \geq 0$ with $(|\boldsymbol{x}^1|/|\boldsymbol{x}|) \geq \varepsilon$,

$$\begin{aligned} |\{\boldsymbol{h}^s(\boldsymbol{x})/\lambda_1^s\} - \gamma(\boldsymbol{x})\boldsymbol{W}| &= |\{\boldsymbol{h}^{s-t_0}(\boldsymbol{h}^{t_0}(\boldsymbol{x})/\lambda_1^{t_0})/\lambda_1^{s-t_0}\} - \gamma(\boldsymbol{h}^{t_0}(\boldsymbol{x})/\lambda_1^{t_0})\boldsymbol{W}| \\ &\leq K_4\alpha_1^{s-t_0}|(\boldsymbol{h}^{t_0}(\boldsymbol{x})/\lambda_1^{t_0})| \leq [K_4\boldsymbol{h}^{t_0}(\tilde{\boldsymbol{x}})/(\lambda_1\alpha_1)^{t_0}]\alpha_1^s|\boldsymbol{x}|, \end{aligned} \tag{12.69}$$

where $K_4 = (K_3 |\boldsymbol{W}|/\alpha_1)$. This concludes the proof of the lemma. *Q.E.D.*

We also need the following lemma to establish theorem 12.1 in section 12.1.

LEMMA 12.4. *If the conditions of lemma* 12.3 *are satisfied, then for each* ε, $0 < \varepsilon < (|\boldsymbol{W}^1|/|\boldsymbol{W}|)$, *there exist finite positive constants* $K_5(\varepsilon)$ *and* α *such that for all* $t \geq 1$ *and all* $\boldsymbol{x} \geq 0$ *with* $(|\boldsymbol{x}^1|/|\boldsymbol{x}|) \geq \varepsilon$,

$$|\{\boldsymbol{h}^t(\boldsymbol{x})/|\boldsymbol{h}^t(\boldsymbol{x})|\} - \{\boldsymbol{W}/|\boldsymbol{W}|\}| \leq K_5(\varepsilon)\alpha^t. \tag{12.70}$$

PROOF. Eq. (12.70) is an almost immediate consequence of (12.47), and the proof of this lemma can be taken verbatim from the proof of the last half of lemma 3 in ref. [12]. We therefore consider lemma 12.4 established. *Q.E.D.*

Next we will establish an analogue of lemma 12.3 for the case $\lambda_1 < \lambda_2$.

LEMMA 12.5. *Let* $\boldsymbol{h}(\cdot)$ *and* $\boldsymbol{h}^t(\cdot)$ *be as defined above and assume that the functions* $h_i(\cdot)$ *are positive, continuous, non-decreasing and homogeneous of degree one. Moreover, assume that there exist a pair of positive constants* (λ_1, λ_2) *and a pair of strictly positive vectors* $(\boldsymbol{V}^1, \bar{\boldsymbol{V}}^2)$ *that satisfy eqs.* (12.46*a*), (12.46*b*) *and* $\lambda_1 < \lambda_2$. *Assume also that* $h_i(\boldsymbol{x}) > 0$, $i = d_1 + 1, \ldots, d$, *whenever* $\boldsymbol{x} \geq 0$, $\boldsymbol{x} \neq 0$, *and that the matrix*

$$\tilde{H}(\bar{\boldsymbol{V}}^2) \equiv \{h_{ij}(0, \bar{\boldsymbol{V}}^2)\}_{d_1+1 \leq i,j \leq d} \equiv \{(\partial h_i(\boldsymbol{x})/\partial x_j)_{\boldsymbol{x}=(0,\bar{\boldsymbol{V}}^2)}\}_{d_1+1 \leq i,j \leq d}$$

exists and is strictly positive. Finally assume that there exist strictly positive constants K_6 *and* γ *such that, for all* $0 < \varepsilon \leq 1$,

$$h_{d_1+i}(\varepsilon \boldsymbol{V}^1, \bar{\boldsymbol{V}}^2) - h_{d_1+i}(0, \bar{\boldsymbol{V}}^2) \leq K_6\varepsilon^\gamma, \qquad i = 1, \ldots, d - d_1. \tag{12.71}$$

Then there exists a positive, continuous, non-decreasing, linearly homogeneous function $\gamma(\cdot)$ *on* $(R_+)^d$ *and two finite positive constants* K_7 *and* ψ *such that* $0 < \Psi < 1$ *and such that for all* $t \geq 1$ *and all* $\boldsymbol{x} \geq 0$,

$$|\{\boldsymbol{h}^t(\boldsymbol{x})/\lambda_2^t\} - \gamma(\boldsymbol{x})(0, \bar{\boldsymbol{V}}^2)| \leq K_7 \Psi^t|\boldsymbol{x}|. \tag{12.72}$$

Moreover, there exists a strictly positive constant η *such that* $\gamma(\boldsymbol{x}) \geq \eta$ *for all* $\boldsymbol{x} \geq 0$, $|\boldsymbol{x}| = 1$.

PROOF. Let

$$\alpha(t, \boldsymbol{x}) = \min_{1 \leq i \leq d-d_1} \{h^t_{d_1+i}(\boldsymbol{x})/\lambda_2^t \bar{V}^2(d_1 + i)\}, \tag{12.73}$$

and

$$\beta(t, x) = \max_{1 \le i \le d - d_1} \{h^t_{d_1+i}(x)/\lambda^t_2 \bar{V}^2(d_1 + i)\}. \tag{12.74}$$

Then for $t \ge 1$ one has as in (12.50)

$$\alpha(t, x) \ge \alpha(t, (\mathbf{0}, x^2)) \ge \alpha(1, (\mathbf{0}, x^2)), \tag{12.75}$$

where $\mathbf{0}$ denotes a d_1-dimensional vector with all components equal to zero. Moreover, for $t \ge 2$,

$$\begin{aligned} \alpha(t, x) &= \min_{1 \le i \le d - d_1} \{h^t_{d_1+i}(x)/\lambda^t_2 \bar{V}^2(d_1 + i)\} \\ &= \min_{1 \le i \le d - d_1} \{h^{t-1}_{d_1+i}(h(x)/\lambda_2)/\lambda^{t-1}_2 \bar{V}^2(d_1 + i)\} \\ &= \alpha(t - 1, (h(x)/\lambda_2)). \end{aligned} \tag{12.76}$$

Finally, note that there exists a positive constant c such that

$$c \le (h_{d_1+i}(x)/\lambda_2) \qquad \text{for all } 1 \le i \le d - d_1 \text{ and all } x \ge 0, |x| = 1. \tag{12.77}$$

Thus if we let $e = (1, \ldots, 1)$, a $(d - d_1)$-dimensional vector, then eqs. (12.75–12.77) imply that for all $t \ge 2$ and all $x \ge 0$, $|x| = 1$,

$$\alpha(t, x) = \alpha(t - 1, (h(x)/\lambda_2)) \ge \alpha(1, (0, ce)) > 0. \tag{12.78}$$

We also need to define several constants. First, let $H^2(\cdot) = (h_{d_1+i}(\cdot), \ldots, h_d(\cdot))$ and let $0 < \varepsilon_3 < 1$ be so small that, for all $0 \le y \le \varepsilon_3 \bar{V}^2$,

$$H^2(0, \bar{V}^2 + y) \ge H^2(0, \bar{V}^2) + \tfrac{1}{2}\tilde{H}(\bar{V}^2)y. \tag{12.79}$$

Next, let ε_4, $0 < \varepsilon_4 < 1$, be so small that for all $t \ge 2$ and all $x \ge 0$, $|x| = 1$,

$$0 \le \varepsilon_4[(\beta(t, x) - \alpha(t, x))/\alpha(t, x)]\bar{V}^2(d_1 + i)e_i \le \varepsilon_3 \bar{V}^2, 1 \le i \le d - d_1, \tag{12.80}$$

where $e_i = (0, \ldots, 0, e_i(d_1 + i) = 1, 0, \ldots, 0)$ is a $(d - d_1)$-dimensional vector.

Finally, let $\delta > 0$ be so small that

$$\tilde{H}(\bar{V}^2)\bar{V}^2(d_1 + i)e_i \ge \delta \bar{V}^2 > 0 \qquad \text{for all } 1 \le i \le d - d_1. \tag{12.81}$$

The existence of ε_3 is obvious (cf. also (12.54) above). The existence of δ follows from the fact that all the entries of $\tilde{H}(\bar{V}^2)$ are assumed to be strictly positive. Finally, the existence of ε_4 follows from (12.78) and the fact that $\beta(t,\cdot)$ is bounded on A uniformly in t. To establish the latter fact proceed as follows. As in (12.49) and (12.50) one sees that

$$\beta^*(t, x^1) = \max_{1 \le i \le d_1} \{h_i^t(x^1)\}/\{\lambda_1^t V^1(i)\}$$

is non-increasing in t. Hence, if

$$K_8 = \max_{|x^1|=1} \beta^*(1, x^1),$$

then for all $t \ge 1$ and all $x^1 \ge 0$,

$$h_i^t(x^1) \le K_8 \,|\, x^1 \,|\, \lambda_1^t V^1(i), \; 1 \le i \le d_1. \tag{12.82}$$

Hence, for some i_0, $1 \le i_0 \le d - d_1$, a finite constant $K_9 = [K_8/\alpha(1,(0, ce))] \ge [K_8/\inf_{t\ge 2} \alpha(t, x)] \ge [K_8/\inf_{t \ge 2} \beta(t, x)]$, a suitable $K_{10} < \infty$, and all $t \ge 2$, $x \in A$,

$$\begin{aligned}\beta(t+1, x)\bar{V}^2(d_1+i_0) &\le (1/\lambda_2)h_{d_1+i_0}(K_8\,|\,x^1\,|\,(\lambda_1/\lambda_2)^t V^1, \beta(t,x)\bar{V}^2)\\ &\le (1/\lambda_2)\beta(t,x)h_{d_1+i_0}(K_9(\lambda_1/\lambda_2)^t V^1, \bar{V}^2)\\ &\le \beta(t,x)(1+K_{10}\varphi^t)\bar{V}^2(d_1+i_0),\end{aligned} \tag{12.83}$$

where $\varphi = (\lambda_1/\lambda_2)^\gamma < 1$ (recall (12.71) and the fact that $\bar{V}^2(d_1 + i_0) > 0$).
Thus,

$$\beta(t+1, x) \le \beta(2, x)\prod_{s=2}^{t}(1 + K_{10}\varphi^s) \le \beta(2, x)\prod_{s=2}^{\infty}(1 + K_{10}\varphi^s) < \infty. \tag{12.84}$$

Since both $\beta(1,\cdot)$ and $\beta(2,\cdot)$ are bounded on A, (12.84) implies that $\beta(t, x)$ is bounded uniformly in t on A.

Next, let x be a given vector such that $x \ge 0$ and $|\,x\,| = 1$. Then observe that

$$\{h_{d_1+i_0}(x)/\lambda_2^t\} \ge \alpha(t,x)\bar{V}^2(d_1+i_0) + (\beta(t,x) - \alpha(t,x))\bar{V}^2(d_1+i_0) \tag{12.85}$$

for some $1 \le i_0 \le d - d_1$. Hence for $t \ge 2$ and $1 \le i \le d - d_1$,

$$\begin{aligned}\{h_{d_1+i}^{t+1}(x)/\lambda_2^{t+1}\} &\ge (1/\lambda_2)h_{d_1+i}(0, (h_{d_1+1}^t(x)/\lambda_2^t), \ldots, (h_d^t(x)/\lambda_2^t))\\ &\ge (1/\lambda_2)h_{d_1+i}(0, \alpha(t,x)\bar{V}^2 + (\beta(t,x) - \alpha(t,x))\bar{V}^2(d_1+i_0)e_{i_0})\\ &\ge \alpha(t,x)\bar{V}^2(d_1+i) + (\delta\varepsilon_4/2\lambda_2)(\beta(t,x)-\alpha(t,x))\bar{V}^2(d_1+i).\end{aligned} \tag{12.86}$$

(See (12.79) – (12.81) and compare with (12.59).)

From (12.86) we conclude that

$$\alpha(t+1, \boldsymbol{x}) \geq \alpha(t, \boldsymbol{x}) + (\delta\varepsilon_4/2\lambda_2)(\beta(t, \boldsymbol{x}) - \alpha(t, \boldsymbol{x})). \qquad (12.87)$$

Observe also that (12.83) and (12.84) imply that there exists a finite constant K_{11} such that

$$\beta(t+1, \boldsymbol{x}) \leq \beta(t, \boldsymbol{x}) + K_{11}\varphi^t \text{ all } \boldsymbol{x} \in A, \qquad (12.88)$$

But then it follows from (12.87) that for all $t \geq 2$ and $\boldsymbol{x} \in A$,

$$\begin{aligned}(\beta(t+1, \boldsymbol{x}) - \alpha(t+1, \boldsymbol{x})) &\leq K_{11}\varphi^t + (\beta(t, \boldsymbol{x}) - \alpha(t, \boldsymbol{x}))(1 - (\delta\varepsilon_4/2\lambda_2)) \\ &\leq K_{11}\{\varphi^t + (1 - (\delta\varepsilon_4/2\lambda_2))\varphi^{t-1}\} + (\beta(t-1, \boldsymbol{x}) \\ - \alpha(t-1, \boldsymbol{x}))(1 - (\delta\varepsilon_4/2\lambda_2))^2 &\leq \ldots \leq K_{11} \sum_{s=2}^{t} \varphi^s(1 - (\delta\varepsilon_4/2\lambda_2))^{t-s} \\ + (\beta(2, \boldsymbol{x}) - \alpha(2, \boldsymbol{x}))(1 - (\delta\varepsilon_4/2\lambda_2))^{t-1} &\leq K_{12}\Psi^t, \end{aligned} \qquad (12.89)$$

where $\Psi = \frac{1}{2} + \frac{1}{2}\max\{(1 - (\delta\varepsilon_4/2\lambda_2)), \phi, (\lambda_1/\lambda_2)\} < 1$, and where K_{12} is a large constant chosen independently of t and x.

From (12.87), (12.89), (12.82), (12.73), and (12.74), it follows easily that, if we let

$$\gamma(\boldsymbol{x}) = \lim_{t \to \infty} \alpha(t, \boldsymbol{x}), \; \tilde{\boldsymbol{x}} = (\boldsymbol{x}/|\boldsymbol{x}|), \qquad (12.90)$$

then $\gamma(\cdot)$ is well-defined, homogeneous of degree one, non-decreasing, and satisfies the relations

$$\begin{aligned}|\{\boldsymbol{h}^t(\boldsymbol{x})/\lambda_2^t\} - \gamma(\boldsymbol{x})(0, \bar{V}^2)| &= \sum_{i=1}^{d_1} \{h_i^t(\boldsymbol{x}^1)/\lambda_2^t\} + \sum_{i=1}^{d-d_1} |\{h_{d_1+i}^t(\boldsymbol{x})/\lambda_2^t\} \\ - \gamma(\boldsymbol{x})\bar{V}^2(d_1+i)| &\leq K_8|\boldsymbol{x}^1|\,|V^1|(\lambda_1/\lambda_2)^t + |\boldsymbol{x}|(\beta(t, \tilde{\boldsymbol{x}}) - \alpha(t, \tilde{\boldsymbol{x}}))|\bar{V}^2| \\ &\leq K_{13}\Psi^t|\boldsymbol{x}|, \qquad \text{for all } t \geq 1, \end{aligned} \qquad (12.91)$$

where K_{13} is a large finite constant chosen independently of t and $\boldsymbol{x}$.

The preceding paragraph establishes the validity of (12.72). To conclude the proof of the lemma it remains to show that $\gamma(\cdot)$ is continuous and strictly positive on A. To do that let $\boldsymbol{x}^0$ and $\varepsilon > 0$ be fixed and observe that on A, for suitable $K_{14} < \infty$,

$$\begin{aligned}|\gamma(\boldsymbol{x}) - \gamma(\boldsymbol{x}^0)| &\leq |\gamma(\boldsymbol{x}) - \alpha(t, \boldsymbol{x})| + |\alpha(t, \boldsymbol{x}) - \alpha(t, \boldsymbol{x}^0)| \\ &+ |\alpha(t, \boldsymbol{x}^0) - \gamma(\boldsymbol{x}^0)| \leq 2K_{14}\Psi^t + |\alpha(t, \boldsymbol{x}) - \alpha(t, \boldsymbol{x}^0)|. \end{aligned} \qquad (12.92)$$

Hence, if we pick t so large that $2K_{14}\Psi^t < (\varepsilon/2)$ and a neighborhood of $\boldsymbol{x}^0$ in A, say U, such that for all $\boldsymbol{x} \in U$, $|\alpha(t, \boldsymbol{x}) - \alpha(t, \boldsymbol{x}^0)| < \varepsilon/2$, then for all $x \in U \mid \gamma(x) - \gamma(x^0) \mid < \varepsilon$. So $\gamma(\cdot)$ is continuous on A. Since $\gamma(0) = 0$, it follows from the homogeneity of $\gamma(\cdot)$ that $\gamma(\cdot)$ is continuous on $(R_+)^d$. Next observe that (12.90) and (12.78) imply that $\gamma(\cdot)$ is strictly positive on A. This concludes the proof. *Q.E.D.*

We will establish an analogue of lemma 12.4 for the case $\lambda_1 < \lambda_2$.

LEMMA 12.6. *If the conditions of lemma 12.5 are satisfied, then there exists a finite constant K_{15} such that for all $t \geq 1$ and $x \geq 0$, $x \neq 0$,*

$$|\{h^t(x)/|h^t(x)|\} - \{(0, \bar{V}^2)/|\bar{V}^2|\}| \leq K_{15}\Psi^t, \tag{12.93}$$

where Ψ is as in lemma 12.5.

PROOF. Observe first that

$$|\{h^t(x)/|h^t(x)|\} - \{(0, \bar{V}^2)/|\bar{V}^2|\}| = \sum_{i=1}^{d_1} \{h_i^t(x^1)/|h^t(x)|\} + \sum_{i=1}^{d-d_1} |\{h_{d_1+i}^t(x)/|h^t(x)|\} - \{\bar{V}^2(d_1+i)/|\bar{V}^2|\}|. \tag{12.94}$$

Then note that by (12.91)

$$\gamma(\tilde{x})|\bar{V}^2| - K_{13}\Psi^t \leq \sum_{i=1}^{d} \{h_i^t(\tilde{x})/\lambda_2^t\} \leq \gamma(\tilde{x})|\bar{V}^2| + K_{13}\Psi^t, \tag{12.95}$$

and

$$\gamma(\tilde{x})\bar{V}^2(d_1+i) - K_{13}\Psi^t \leq \{h_{d_1+i}^t(\tilde{x})/\lambda_2^t\} \leq \gamma(\tilde{x})\bar{V}^2(d_1+i) + K_{13}\Psi^t, \quad i = 1, \ldots, d - d_1. \tag{12.96}$$

For some large integer t_0 the left-hand sides of (12.95) and (12.96) are positive for all $t \geq t_0$, uniformly in x. Thus for $t \geq t_0$, (12.82) and (12.95) imply that

$$\sum_{i=1}^{d_1} \{h_i^t(x^1)/|h^t(x)|\} \leq [K_{16}|x^1|\lambda_1^t/K_{17}|x|\lambda_2^t] \leq K_{18}(\lambda_1/\lambda_2)^t \leq K_{18}\Psi^t, \tag{12.97}$$

for suitable constants $0 < K_{16}, K_{17}, K_{18} < \infty$. For $t \geq t_0$, $K_{16} - K_{18}$ can be chosen to be independent of x and t. Moreover, for suitably large constants K_{19} and K_{20}, (12.95) and (12.96) imply that

$$\{h_{d_1+i}^t(x)/|h^t(x)|\} \leq \{(\gamma(\tilde{x})\bar{V}^2(d_1+i) + K_{13}\Psi^t)/\gamma(\tilde{x})|\bar{V}^2| (1 - (K_{13}\Psi^t/\gamma(\tilde{x})|\bar{V}^2|))\} \leq (\bar{V}^2(d_1+i)/|\bar{V}^2|) + K_{19}\Psi^t, \tag{12.98}$$

and

$$\{h^t_{d_1+i}(x)/|h^t(x)|\} \geq \{(\gamma(\tilde{x})\bar{V}^2(d_1+i) - K_{13}\Psi^t)/\gamma(\tilde{x})|\bar{V}^2| \\ (1+(K_{13}\Psi^t/\gamma(\tilde{x})|\bar{V}^2|))\} \geq (\bar{V}^2(d_1+i)/|\bar{V}^2|) - K_{20}\Psi^t. \quad (12.99)$$

It follows from (12.94), (12.97), (12.98) and (12.99) that (12.93) is satisfied for all $t \geq t_0$ and all $x \geq 0$, $x \neq 0$ with $K_{15} = K_{18} + d(K_{19} + K_{20})$. By taking K_{15} larger if necessary we can satisfy (12.93) for all $t \geq 1$. *Q.E.D.*

Lemmas 12.3 and 12.5 can be used to characterize the properties of solutions to non-linear difference equations of the form

$$x_{t+1} = h(x_t), \quad (12.100)$$

where $h(\cdot)$ is as defined in (12.1). Specifically, they give sufficient conditions when $\lambda_1 \neq \lambda_2$ that there exists a positive, continuous, non-decreasing linearly homogeneous function $\gamma(\cdot)$ and a positive vector V such that if $\lambda = \max(\lambda_1, \lambda_2)$,

$$\lim_{t\to\infty} \{x_t/\lambda^t\} = \gamma(x_0)V. \quad (12.101)$$

These lemmas also show that the convergence in eq. (12.101) occurs at an exponential rate.

We will conclude this section by stating several lemmas for the case $\lambda_1 = \lambda_2$. The proofs are omitted for brevity's sake.

LEMMA 12.7. *Let $\lambda > 0$ and suppose that*

$$x_{t+1,1} = \lambda x_{t,1}, \quad (12.102)$$

$$x_{t+1,2} = p x^{\alpha}_{t,1} x^{1-\alpha}_{t,2} + \lambda x_{t,2}. \qquad 0 < \alpha < 1. \quad (12.103)$$

Then

$$x_{t,1} = \lambda^t x_{0,1}, \text{ and, if } x_{0,2} \neq 0 \quad (12.104)$$

$$\lim_{t\to\infty} \{x_{t,2}/t^{(1/\alpha)}\lambda^t\} = [p x^{\alpha}_{0,1}\alpha/\lambda]^{(1/\alpha)}. \quad (12.105)$$

LEMMA 12.8. *Let $\lambda > 0$ and suppose that $x_{0,1} \neq 0$,*

$$x_{t+1,1} = \lambda x_{t,1}, \quad (12.106)$$

$$x_{t+1,2} = \sum_{k=1}^{\infty} p_k x_{t,1}^{\alpha_k} x_{t,2}^{(1-\alpha_k)} + \lambda x_{t,2}, \tag{12.107}$$

where

$$p_k > 0, 0 < \alpha_k < 1, k = 1, 2, \ldots, \sum_{k=1}^{\infty} p_k < \infty$$

and $\lim_{k \to \infty} \alpha_k = 0.$

Then (12.104) *is valid and if* $x_{0,2} \neq 0$, *then* $\{x_{t,2}/\lambda^t\}$ *grows faster than any power of* t.

LEMMA 12.9. *Let* $\lambda > 0$ *and let* $h(\cdot)$ *be a non-decreasing, continuous, linearly homogeneous function from* $(R_+)^2$ *to* R_+ *such that* $\lambda = h(0,1)$ *and such that* $h(\boldsymbol{x}) > 0$ *if* $\boldsymbol{x} \neq 0$. *Moreover, let* $\boldsymbol{x}_t = (x_{t,1}, x_{t,2})$ *and suppose that*

$$x_{t+1,1} = \lambda x_{t,1}, \tag{12.108}$$

$$x_{t+1,2} = h(\boldsymbol{x}_t). \tag{12.109}$$

Finally, suppose that there exist finite constants K and γ *such that* $0 < \gamma < 1$ *and such that for all* $0 < \varepsilon \leq 1$,

$$h(\varepsilon, 1) - h(0, 1) \leq K\varepsilon^{\gamma}. \tag{12.110}$$

Then $x_{t,2} = 0(t^{(1/\gamma)}\lambda^t)(t \to \infty)$ *whenever* $\boldsymbol{x}_0 > 0$.

LEMMA 12.10. *Let* $\lambda > 0$ *and let* $h(\cdot)$ *be a non-decreasing, continuous, linearly homogeneous function from* $(R_+)^2$ *to* R_+ *such that* $\lambda = h(0, 1)$ *and such that* $h(\boldsymbol{x}) > 0$ *if* $\boldsymbol{x} \neq 0$. *Also let* $\boldsymbol{x}_t = (x_{t,1}, x_{t,2})$ *and suppose that eqs.* (12.108) *and* (12.109) *are satisfied. Then for all* $\delta > 0$,

$$\limsup_{t \to \infty} \{x_{t,2}/(\lambda + \delta)^t\} = 0. \tag{12.111}$$

Moreover, if there exists a constant η *such that for all* $0 < \varepsilon \leq 1$,

$$\{[h(\varepsilon, 1) - h(0, 1)]/\varepsilon\} \geq \eta > 0, \tag{12.112}$$

then

$$\liminf_{t \to \infty} \{x_{t,2}/t\lambda^t\} > 0 \text{ whenever } x_0 > 0. \tag{12.113}$$

12.3. A Basic Convergence Theorem

In this section we prove a basic convergence theorem for sequences of positive random vectors. Throughout our discussion the probability space $(\Omega, \mathscr{F}, P)$ is kept fixed and a vector-valued function on Ω is said to be a random vector if and only if it is measurable with respect to $\mathscr{F}$. Moreover, we consider a given sequence of positive d-dimensional random vectors $\boldsymbol{x}_t$, $t = 0, 1, \ldots$, and write $P_t\{E\}$ for the conditional probability of the event E given the σ-field of events generated by $\boldsymbol{x}_s$, $0 \leq s \leq t$. Finally, $\boldsymbol{x}$ without a subscript denotes an 'ordinary' d-dimensional vector, $A \equiv \{\boldsymbol{x} \geq 0 : |\boldsymbol{x}| = 1\}$, and $\tilde{\boldsymbol{x}} \equiv (\boldsymbol{x}/|\boldsymbol{x}|)$ whenever $\boldsymbol{x} \neq 0$.

THEOREM 12.3 [13]. *Let T be a continuous transformation from A into A, B a closed subset of A, p a fixed point of T, and U a neighborhood of p such that the following relations hold for some constants K, $0 < \lambda < 1$, β, δ, $\tau > 0$, $\rho > 1$ and a function $R(\cdot)$:*

$$Tp = p,\ p \in U \subset B; \tag{12.114}$$

$$|T\boldsymbol{x} - T\boldsymbol{y}| \leq K|\boldsymbol{x} - \boldsymbol{y}| \qquad \text{for } \boldsymbol{x}, \boldsymbol{y} \in U; \tag{12.115}$$

$$|T^k\boldsymbol{x} - p| \leq K\lambda^k \qquad \text{for } k \geq 0 \text{ and } \boldsymbol{x} \in B; \tag{12.116}$$

$$P_t\{|\boldsymbol{x}_{t+1} - R(\boldsymbol{x}_t)T\tilde{\boldsymbol{x}}_t| \geq |\boldsymbol{x}_t|^{1-(\delta/2)}\} \leq K|\boldsymbol{x}_t|^{-\delta}; \tag{12.117}$$

$$|\{R(\boldsymbol{x})/|\boldsymbol{x}|\} - \rho| \leq K\{|\tilde{\boldsymbol{x}} - p|^{\beta} + |\boldsymbol{x}|^{-\beta}\} \text{ for } \boldsymbol{x} \geq 0,\ \boldsymbol{x} \neq 0; \tag{12.118}$$

and

$$R(\boldsymbol{x}) \geq \tau|\boldsymbol{x}| \qquad \text{for } \boldsymbol{x} \geq 0. \tag{12.119}$$

Then there is a positive constant $K_1 < \infty$ such that for any integer $t_0 \geq 0$

$$P_{t_0}\{\lim_{t\to\infty} \{\boldsymbol{x}_t/\rho^t\} = g \cdot p \qquad \text{for some } g > 0\} \geq 1 - K_1|x_{t_0}|^{-\delta} \tag{12.120}$$

whenever $\boldsymbol{x}_{t_0} \neq 0$ and $\tilde{\boldsymbol{x}}_{t_0} \in B$.

---

[13] This theorem is an extension of theorem 6.1 in ref. [6, pp. 91–92].

PROOF. The proof will be broken down into several steps. We show by an inductive argument that with high probability

$$|x_{t+1} - R(x_t)T\tilde{x}_t| \le |x_t|^{1-(\delta/2)} \text{ eventually.}$$

This will imply that $\tilde{x}_{t+1}$ is close to $T\tilde{x}_t$ which will allow us to show that

$$\lim_{t\to\infty} \tilde{x}_t = \lim_{t\to\infty} T^{t-t_1}\tilde{x}_{t_1} = p.$$

Throughout the proof C_0, C_1, ... will be suitable finite, positive constants, whereas θ_i will always denote some number with $|\theta_i| \le 1$. Finally, we will use the abbreviation $R_t \equiv R(x_t)$.

Step 1. Let $C_0 = [8K/(\rho - 1)]^{1/\beta}$, and

$$V_1 = \{x \in A: |x - p| \le K\lambda^{k_1}\},$$

where k_1 is chosen so large that

$$V_1 \subset U, \tag{12.121}$$

and

$$[R(x)/|x|] \ge [(3\rho + 1)/4] \qquad \text{for } \tilde{x} \in V_1, |x| \ge C_0. \tag{12.122}$$

(The existence of such a k_1 is assured by (12.118)). Also let

$$V_2 = \{x \in A: |x - p| \le K\lambda^{k_1+1}\}.$$

Then there exists a neighborhood

$$V_3 = \{x \in A: |x - p| \le K\lambda^{k_3}\} \subset V_2 \text{ (with } k_3 \ge k_1 + 1)$$

such that

$$T^j x \in V_2 \subset V_1 \text{ for all } j \ge 0 \text{ whenever } x \in V_3. \tag{12.123}$$

Proof. Since $V_2 \subset U \subset B$, we deduce from (12.116) that

$$T^j V_2 \subset V_2 \text{ for all } j \ge k_1 + 1. \tag{12.124}$$

Now choose k_3 so large that

$$T^j x \subset V_2 \text{ for } j = 0, 1, \dots, k_1 \text{ if } |x - p| \le K\lambda^{k_3}. \tag{12.125}$$

Such a k_3 exists by (12.115) and the fact that $T^j p = p$. Clearly, with this k_3, V_3 satisfies (12.123) since for any $x \in V_3$ and $j \ge 0$ either (12.124) or (12.125) applies.

Step 2. Assume that

$$T^{s-t_1}\tilde{\boldsymbol{x}}_{t_1} \in V_1 \text{ for } t_1 \leq s < t_2; \tag{12.126}$$

$$|\boldsymbol{x}_{t_1}| \geq C_1 \equiv [4/(\rho - 1)]^{(2/\delta)} + 2^{(2/\delta)} + C_0, \tag{12.127}$$

$$\tilde{\boldsymbol{x}}_s \in V_1 \text{ for } t_1 \leq s < t_2, \tag{12.128}$$

and

$$|\boldsymbol{x}_s - R_{s-1}T\tilde{\boldsymbol{x}}_{s-1}| \leq |\boldsymbol{x}_{s-1}|^{1-(\delta/2)} \text{ for } t_1 < s \leq t_2. \tag{12.129}$$

Then

$$|\boldsymbol{x}_{t_2}| \geq ((\rho + 1)/2)^{t_2 - t_1} |\boldsymbol{x}_{t_1}|, \tag{12.130}$$

and for $C_2 = K + 5$ and $C_3 = \sum_{n=0}^{\infty} ((\rho + 1)/2)^{-(\delta n/2)}$,

$$\begin{aligned} |\tilde{\boldsymbol{x}}_{t_2} - T^{t_2 - t_1}\tilde{\boldsymbol{x}}_{t_1}| &\leq |\boldsymbol{x}_{t_1}|^{-(\delta/2)} \sum_{j=t_1+1}^{t_2} C_2^{j-t_1}((\rho + 1)/2)^{-((t_2 - j)\delta/2)} \\ &\leq C_3 C_2^{t_2 - t_1} |\boldsymbol{x}_{t_1}|^{-(\delta/2)}. \end{aligned} \tag{12.131}$$

Proof. If

$$|\boldsymbol{x}_{s-1}| \geq C_1, \tilde{\boldsymbol{x}}_{s-1} \in V_1, \tag{12.132}$$

and

$$|\boldsymbol{x}_s - R_{s-1}\, T\tilde{\boldsymbol{x}}_{s-1}| \leq |\boldsymbol{x}_{s-1}|^{1-(\delta/2)}, \tag{12.133}$$

then it follows from (12.122) and the definition of C_1 in (12.127) that

$$\begin{aligned} |\boldsymbol{x}_s| \geq R_{s-1} - |\boldsymbol{x}_{s-1}|^{1-(\delta/2)} &\geq |\boldsymbol{x}_{s-1}|[((3\rho + 1)/4) - ((\rho - 1)/4)] \\ &= [(\rho + 1)/2]|\boldsymbol{x}_{s-1}|. \end{aligned} \tag{12.134}$$

Since (12.132) and (12.133) hold by assumption for $s = t_1 + 1$, it follows from (12.134) that $|\boldsymbol{x}_{t_1+1}| \geq [(\rho + 1)/2]|\boldsymbol{x}_{t_1}| \geq C_1$. But if that is so, then either $t_2 = t_1 + 1$ and (12.130) is established, or (12.128) and (12.129) imply that (12.132) and (12.133) hold for $s = t_1 + 2$. Evidently in the latter case (12.134) also holds for $s = t_1 + 2$. By induction we find that (12.134) holds for all $t_1 < s \leq t_2$, which implies

$$|\boldsymbol{x}_s| \geq [(\rho + 1)/2]^{s - t_1}|\boldsymbol{x}_{t_1}| \text{ for } t_1 \leq s \leq t_2. \tag{12.135}$$

For $s = t_2$ this is just (12.130).

To prove (12.131) proceed as follows. First, let

$$\alpha_s \equiv |\tilde{\boldsymbol{x}}_s - T^{s-t_1}\tilde{\boldsymbol{x}}_{t_1}|, t_1 \leq s \leq t_2.$$

Next, observe that (12.129) implies that

$$x_s = R_{s-1}T\tilde{x}_{s-1} + \theta_{s-1}|x_{s-1}|^{1-(\delta/2)}y_{s-1},\ t_1 < s \le t_2, \tag{12.136}$$

for some vector y_{s-1} with $|y_{s-1}| = 1$. Observe also that (12.136), (12.122) and (12.135) imply that

$$\begin{aligned} \| x_s| - R_{s-1}| &\le |x_{s-1}|^{1-(\delta/2)} \\ &\le [4/(3\rho + 1)]R_{s-1}[(\rho + 1)/2]^{-(\delta/2)(s-1-t_1)}|x_{t_1}|^{-(\delta/2)} \\ &\le R_{s-1}[(\rho + 1)/2]^{-(\delta/2)(s-1-t_1)}|x_{t_1}|^{-(\delta/2)} \le (R_{s-1}/2). \end{aligned}$$

From the last inequality and from (12.136) it follows that

$$\begin{aligned} \tilde{x}_s = \frac{x_s}{|x_s|} &= \frac{R_{s-1}T\tilde{x}_{s-1} + \theta_{s-1}|x_{s-1}|^{1-(\delta/2)}y_{s-1}}{R_{s-1}\{1 + \theta'_{s-1}[(\rho + 1)/2]^{-(\delta/2)(s-1-t_1)}|x_{t_1}|^{-(\delta/2)}\}} \\ &= \{T\tilde{x}_{s-1} + [\theta_{s-1}/R_{s-1}]|x_{s-1}|^{1-(\delta/2)}y_{s-1}\} \\ &\quad \times \{1 + 2\theta''_{s-1}[(\rho + 1)/2]^{-(\delta/2)(s-1-t_1)}|x_{t_1}|^{-\delta/2}\} \end{aligned} \tag{12.137}$$

for some $|\theta'_{s-1}|$, $|\theta''_{s-1}| \le 1$. Finally, observe that (12.115) and the relation $\tilde{x}_{s-1} \in V_1$ and $T^{s-1-t_1}\tilde{x}_{t_1} \in V_1$, $t_1 < s \le t_2$, imply that for $t_1 < s \le t_2$,

$$|T\tilde{x}_{s-1} - T^{s-t_1}\tilde{x}_{t_1}| = |T\tilde{x}_{s-1} - T(T^{s-1-t_1}\tilde{x}_{t_1})| \le K\alpha_{s-1}.$$

Hence, from (12.137), (12.122) and (12.135), we can deduce that

$$\begin{aligned} \alpha_s = |\tilde{x}_s - T^{s-t_1}x_{t_1}| &\le K\alpha_{s-1} + 3|x_{s-1}|^{-(\delta/2)}[4/(3\rho + 1)] \\ &\quad + 2[(\rho + 1)/2]^{-(\delta/2)(s-1-t_1)}|x_{t_1}|^{-(\delta/2)} \\ &\le C_2\{\alpha_{s-1} + |x_{t_1}|^{-(\delta/2)}[(\rho + 1)/2]^{-(\delta/2)(s-1-t_1)}\}. \end{aligned}$$

From the last inequality and from the fact that $\alpha_{t_1} = 0$ the validity of (12.131) follows by simple iteration.

Step 3. Let $t_1 \ge t_0$ be fixed for the moment and denote by E_s, $s \ge t_1$, the event

$$E_s = \{|x_j| \ge [(\rho + 1)/2]^{j-t_1}|x_{t_1}| \ge C_1,\ \tilde{x}_j \in V_1 \text{ for all } t_1 \le j \le s,$$

$$\text{and} \quad |x_j - R_{j-1}T\tilde{x}_{j-1}| \le |x_{j-1}|^{1-(\delta/2)} \text{ for all } t_1 < j \le s\}.$$

Then, for

$$C_4 = C_1 + \{K\lambda^{k_3}(1 - \lambda)C_3^{-1}C_2^{-k_3-2}\}^{-(2/\delta)},$$

and some $C_5 < \infty$,

$P_{t_1}\{E_s$ holds for all $s \geq t_1$ and $\tilde{x}t_{1+m(k_3+1)} \in V_3$ for all $m \geq 0\} \geq 1 - C_5 |x_{t_1}|^{-\delta}$

on the set where

$$|x_{t_1}| \geq C_4, \tilde{x}_{t_1} \in V_3.$$

Proof. The proof is obtained by an iterative argument. Let k_3 be as in step 1 and observe that, if $\tilde{x}_{t_1} \in V_3$ and $|x_{t_1}| \geq C_4$, then E_{t_1} occurs. Next, assume that $|x_{t_1}| \geq C_4$, $t_1 \leq t_3 \leq t_1 + k_3 + 1$ and that E_{t_3} occurs. Then obtain an estimate of the probability that E_{t_3+1} will occur as well in the following way. Note first that (12.117) implies that on E_{t_3},

$$\begin{aligned} P_{t_3}\{|x_{t_3+1} - R_{t_3}T\tilde{x}_{t_3}| \leq |x_{t_3}|^{1-(\delta/2)}\} &\geq 1 - K|x_{t_3}|^{-\delta} \\ &\geq 1 - K|x_{t_1}|^{-\delta}[(\rho + 1)/2]^{-\delta(t_3 - t_1)}. \end{aligned} \tag{12.138}$$

Next, note that, if E_{t_3} occurs and if in addition

$$|x_{t_3+1} - R_{t_3}T\tilde{x}_{t_3}| \leq |x_{t_3}|^{1-(\delta/2)}, \tag{12.139}$$

then the assumptions (12.127)–(12.129) hold with t_2 replaced by $t_3 + 1$. Moreover, (12.126) holds for $\tilde{x}_{t_1} \in V_3$ by (12.123). Thus it follows from step 2 that (12.130) and (12.131) hold for $t_2 = t_3 + 1$; i.e.

$$|x_{t_3+1}| \geq [(\rho + 1)/2]^{t_3+1-t_1}|x_{t_1}| \tag{12.140}$$

and

$$\begin{aligned} |\tilde{x}_{t_3+1} - T^{t_3+1-t_1}\tilde{x}_{t_1}| &\leq C_3C_2^{t_3+1-t_1}|x_{t_1}|^{-(\delta/2)} \\ &\leq C_3C_2^{k_3+2}|x_{t_1}|^{-(\delta/2)}. \end{aligned} \tag{12.141}$$

Finally, note that, if $\tilde{x}_{t_1} \in V_3$, then (cf. step 1) $T^{t_3+1-t_1}\tilde{x}_{t_1} \in V_2$; i.e.

$$|T^{t_3+1-t_1}\tilde{x}_{t_1} - p| \leq K\lambda^{k_1+1}. \tag{12.142}$$

From (12.141) and (12.142) it follows that

$$\begin{aligned} |\tilde{x}_{t_3+1} - p| &\leq K\lambda^{k_1+1} + C_3C_2^{k_3+2}|x_{t_1}|^{-(\delta/2)} \\ &\leq K\lambda^{k_1+1} + C_3C_2^{k_3+2}C_4^{-(\delta/2)} \leq K\lambda^{k_1}. \end{aligned}$$

Hence, if (12.139) occurs, and $\tilde{x}_{t_1} \in V_3$, then $\tilde{x}_{t_3+1} \in V_1$. This last observation together with (12.138) and (12.139) imply that on the set where $|x_{t_1}| \geq C_4$, $\tilde{x}_{t_1} \in V_3$ and E_{t_3} occurs, $t_3 \leq t_1 + k_3 + 1$

$$P_{t_3}\{E_{t_3+1} \text{ occurs}\} \geq 1 - K|x_{t_1}|^{-\delta}[(\rho + 1)/2]^{-\delta(t_3 - t_1)}.$$

By repeating the preceding arguments $k_3 + 1$ times we find that, on the set where $|x_{t_1}| \geq C_4$ and $\tilde{x}_{t_1} \in V_3$,

$$P_{t_1}\{E_{t_1+k_3+1} \text{ occurs}\} \geq \prod_{j=0}^{k_3} \{1 - K|x_{t_1}|^{-\delta}[(\rho + 1)/2]^{-\delta j}\}.$$

We also find on $E_{t_1+k_3+1}$ (cf. (12.141) for $t_3 = k_3 + t_1$) that

$$\begin{aligned}|\tilde{x}_{t_1+k_3+1} - p| &\leq |T^{k_3+1}\tilde{x}_{t_1} - p| + C_3 C_2{}^{k_3+1} C_4{}^{-(\delta/2)} \\ &\leq K\lambda^{k_3+1} + C_3 C_2{}^{k_3+1} C_4{}^{-(\delta/2)} \leq K\lambda^{k_3},\end{aligned}$$

which implies that $\tilde{x}_{t_1+k_3+1} \in V_3$.

From the last result it follows that we can start over again with t_1 replaced by $t_1 + k_3 + 1$ and $|x_{t_1}|$ by $|x_{t_1+k_3+1}| \geq |x_{t_1}|[(\rho + 1)/2]^{k_3+1}$. Then we obtain on the set where $|x_{t_1}| \geq C_4$, $\tilde{x}_{t_1} \in V_3$, and $E_{t_1+k_3+1}$ occurs

$$\begin{aligned}&P_{t_1+k_3+1}\{E_{t_1+2(k_3+1)} \text{ occurs}\} \\ &\qquad \geq \prod_{j=0}^{k_3} \{1 - K|x_{t_1}|^{-\delta}[(\rho + 1)/2]^{-\delta(k_3+1+j)}\}.\end{aligned}$$

A repetition of the argument for t_1 replaced by $t_1 + m(k_3 + 1)$ and

$$|x_{t_1}| \text{ by } |x_{t_1+m(k_3+1)}| \geq [(\rho + 1)/2]^{m(k_3+1)}|x_{t_1}|,\ m \geq 1$$

finally yields: on the set where $|x_{t_1}| \geq C_4$ and $\tilde{x}_{t_1} \in V_3$,

$$\begin{aligned}&P_{t_1}\{E_s \text{ holds for all } s \geq t_1 \text{ and } \tilde{x}_{t_1+m(k_3+1)} \in V_3 \text{ for } m \geq 0\} \\ &\qquad \geq \prod_{m=0}^{\infty} \prod_{j=0}^{k_3} \{1 - K|x_{t_1}|^{-\delta}[(\rho + 1)/2]^{-\delta(m(k_3+1)+j)}\} \\ &\qquad = \prod_{s=0}^{\infty} \{1 - K|x_{t_1}|^{-\delta}[(\rho + 1)/2]^{-\delta s}\} \geq 1 - C_5|x_{t_1}|^{-\delta}\end{aligned}$$

for suitable $C_5 < \infty$.

Step 4. Any sample path for which

$$\tilde{x}_{t_1+m(k_3+1)} \in V_3,\ m \geq 0,\ |x_{t_1}| \geq C_4, \tag{12.143}$$

$$\tilde{x}_s \in V_1,\ s \geq t_1, \tag{12.144}$$

$$|x_s| \geq [(\rho + 1)/2]^{s-t_1}|x_{t_1}|,\ s \geq t_1, \tag{12.145}$$

and

$$|x_s - R_{s-1}T\tilde{x}_{s-1}| \leq |x_{s-1}|^{1-(\delta/2)},\ s > t_1, \tag{12.146}$$

hold, satisfies

$$\lim_{t\to\infty} \{x_t/\rho^t\} = g \cdot p \text{ for some } g > 0. \tag{12.147}$$

Proof. Since $\tilde{x}_{t_1+m(k_3+1)} \in V_3 \subset V_1$, it follows from step 1 that

$$T^j\tilde{x}_{t_1+m(k_3+1)} \in V_1,\ j \geq 0. \tag{12.148}$$

Also (cf. (12.116)),

$$|T^j\tilde{x}_{t_1+m(k_3+1)} - p| \leq K\lambda^j. \tag{12.149}$$

By virtue of (12.144)–(12.146) and (12.148) we may now apply step 2 with t_1 replaced by $t_1 + m(k_3 + 1)$ and t_2 by $s \geq t_1 + m(k_3 + 1)$. (12.131), (12.145) and (12.149) yield

$$\begin{aligned}
|\tilde{x}_s - p| &\leq |\tilde{x}_s - T^{s-t_1-m(k_3+1)}\tilde{x}_{t_1+m(k_3+1)}| \\
&\quad + |T^{s-t_1-m(k_3+1)}\tilde{x}_{t_1+m(k_3+1)} - p| \\
&\leq C_3C_2{}^{s-t_1-m(k_3+1)}|x_{t_1+m(k_3+1)}|^{-(\delta/2)} + K\lambda^{s-t_1-m(k_3+1)} \\
&\leq C_3C_2{}^{s-t_1-m(k_3+1)}[(\rho+1)/2]^{-(\delta/2)m(k_3+1)}C_4^{-(\delta/2)} + K\lambda^{s-t_1-m(k_3+1)}.
\end{aligned} \tag{12.150}$$

If we let

$$m = \left[\frac{s-t_1}{k_3+1} \cdot \frac{\log(C_2/\lambda)}{(\delta/2)\log[(\rho+1)/2] + \log(C_2/\lambda)}\right]$$

and

$$\lambda_1 = \exp\left\{\frac{(\delta/2)\log[(\rho+1)/2]\log\lambda}{(\delta/2)\log[(\rho+1)/2] + \log(C_2/\lambda)}\right\} < 1,$$

then the last member of (12.150) is bounded by

$$C_6\lambda_1^{s-t_1} \tag{12.151}$$

for suitable $C_6 < \infty$. Thus (12.143)–(12.146) imply that

$$|\tilde{x}_s - p| \leq C_6\lambda_1^{s-t_1},\ s \geq t_1, \tag{12.152}$$

and hence that $\lim_{t\to\infty} \tilde{x}_t = p$. Moreover, (12.118), (12.152), (12.143) and (12.145) imply that for $s \geq t_1$,

$$\begin{aligned}
|\{R_s/|x_s|\} - \rho| &= |\{R(x_s)/|x_s|\} - \rho| \\
&\leq K\{C_6^\beta\lambda_1^{\beta(s-t_1)} + [(\rho+1)/2]^{-\beta(s-t_1)}C_4{}^{-\beta}\} \\
&\leq C_7\lambda_2{}^{(s-t_1)},
\end{aligned} \tag{12.153}$$

where $\lambda_2 = \max(\lambda_1^\beta, [(\rho + 1)/2]^{-\beta}) < 1$, and $C_7 = K\{C_6^\beta + C_4^{-\beta}\}$. By substituting (12.153) in (12.146) we obtain

$$\begin{aligned}|\boldsymbol{x}_s| &= R_{s-1} + \theta_{s-1}|\boldsymbol{x}_{s-1}|^{1-(\delta/2)} \\ &= |\boldsymbol{x}_{s-1}|\{\rho + \theta'_{s-1}C_7\lambda_2^{s-1-t_1} + \theta_{s-1}[(\rho+1)/2]^{-(\delta/2)(s-1-t_1)}C_4^{-(\delta/2)}\}\end{aligned}$$

and therefore for any $s \geq t_2 > t_1$

$$\frac{|\boldsymbol{x}_s|}{\rho^{s-t_2}|\boldsymbol{x}_{t_2}|} = \prod_{j=0}^{s-t_2-1}\{1 + (\theta'_{t_2+j}/\rho)C_7\lambda_2^{t_2-t_1+j} + (\theta_{t_2+j}/\rho)[(\rho+1)/2]^{-(\delta/2)(t_2-t_1+j)}C_4^{-(\delta/2)}\}$$

as well as

$$\lim_{t\to\infty}\{|\boldsymbol{x}_t|/\rho^t\} = (|\boldsymbol{x}_{t_2}|/\rho^{t_2})\prod_{j=0}^{\infty}\{1 + (\theta'_{t_2+j}/\rho)C_7\lambda_2^{t_2-t_1+j} + (\theta_{t_2+j}/\rho)\cdot[(\rho+1)/2]^{-(\sigma/2)(t_2-t_1+j)}C_4^{-(\delta/2)}\}. \quad (12.154)$$

From (12.152) and (12.154) the validity of (12.147) follows with

$$\begin{aligned}g &= (|\boldsymbol{x}_{t_2}|/\rho^{t_2})\prod_{j=0}^{\infty}\{1 + (\theta'_{t_2+j}/\rho)C_7\lambda_2^{t_2-t_1+j} \\ &\quad + (\theta_{t_2+j}/\rho)[(\rho+1)/2]^{-(\delta/2)(t_2-t_1+j)}C_4^{-(\delta/2)}\} \\ &\geq [(\rho+1)/2]^{t_2-t_1}(C_4/\rho^{t_2})\prod_{j=0}^{\infty}\{1 - (C_7/\rho)\lambda_2^{t_2-t_1+j} \\ &\quad - (C_4^{-(\delta/2)}/\rho)[(\rho+1)/2]^{-(\delta/2)(t_2-t_1+j)}\} \\ &> 0 \text{ for sufficiently large } t_2 - t_1.\end{aligned}$$

Steps 3 and 4 show that on the set where $|\boldsymbol{x}_{t_1}| \geq C_4$ and $\tilde{\boldsymbol{x}}_{t_1} \in V_3$

$$P_{t_1}\{\lim_{t\to\infty}\{\boldsymbol{x}_t/\rho^t\} = g\cdot p \text{ for some } g > 0\} \geq P_{t_1}\{(12.143)\text{–}(12.146) \text{ hold}\} \geq 1 - C_5|\boldsymbol{x}_{t_1}|^{-\delta}. \quad (12.155)$$

We now have to show that we can start with any $\boldsymbol{x}_{t_0} \neq 0$, $\tilde{\boldsymbol{x}}_{t_0} \in B$. This is done in step 5.

Step 5. The relation (12.120) holds whenever $\boldsymbol{x}_{t_0} \neq 0$ and $\tilde{\boldsymbol{x}}_{t_0} \in B$.

Proof. Let

$$C_8^{(\delta/2)} \geq 1 + (\tau/2)^{-1-(\delta/2)k_3} \quad (12.156)$$

and assume that

$$|x_{t_0}| \geq C_8,\ \tilde{x}_{t_0} \in B, \tag{12.157}$$

as well as

$$|x_{t+1} - R_t T\tilde{x}_t| \leq |x_t|^{1-(\delta/2)},\ t_0 \leq t \leq t_0 + k_3, \tag{12.158}$$

where k_3 is as in step 1. Then, by (12.119),

$$|x_{t_0+1}| \geq |x_{t_0}|(\tau - |x_{t_0}|^{-(\delta/2)}) \geq |x_{t_0}|(\tau - C_8^{-(\delta/2)}) \geq |x_{t_0}|(\tau/2),$$

and by induction on t one has, for $t_0 \leq t \leq t_0 + k_3$,

$$|x_t| \geq |x_{t_0}|(\tau/2)^{t-t_0}. \tag{12.159}$$

Returning to (12.158), we obtain

$|x_{t+1}| = R_t + \theta_t |x_t|^{1-(\delta/2)}$ and for some vector y_t with $|y_t| = 1$

$$\begin{aligned}|\tilde{x}_{t+1} - T\tilde{x}_t| &= \left|\frac{R_t T\tilde{x}_t + \theta_t' |x_t|^{1-(\delta/2)} y_t}{R_t + \theta_t |x_t|^{1-(\delta/2)}} - T\tilde{x}_t\right| \\ &= |\{T\tilde{x}_t + \theta_t'[|x_t|/R_t]|x_t|^{-(\delta/2)} y_t\}\{1 + 2\theta_t''[|x_t|/R_t]|x_t|^{-(\delta/2)}\} - T\tilde{x}_t| \\ &\leq 5[|x_t|^{-(\delta/2)}/\tau] \leq (5/\tau)(\tau/2)^{-(\delta/2)(t-t_0)} C_8^{-(\delta/2)},\ t_0 \leq t \leq t_0 + k_3\end{aligned} \tag{12.160}$$

(cf. our estimates in step 2). Since T is uniformly continuous on the compact set A, there exists an $\varepsilon_1 > 0$ such that

$$|Tx - Ty| \leq \tfrac{1}{2}K\lambda^{k_3}(1 - \lambda) \text{ for } x, y \in A \text{ with } |x - y| \leq \varepsilon_1.$$

Consequently, if

$$|\tilde{x}_{t_0+k_3} - T^{k_3}\tilde{x}_{t_0}| \leq \varepsilon_1, \tag{12.161}$$

and if C_8 is so large that

$$(5/\tau)(\tau/2)^{-(\delta/2)k_3} C_8^{-(\delta/2)} \leq \tfrac{1}{2}K\lambda^{k_3}(1 - \lambda),$$

then

$$\begin{aligned}|\tilde{x}_{t_0+k_3+1} - p| &\leq |\tilde{x}_{t_0+k_3+1} - T\tilde{x}_{t_0+k_3}| + |T\tilde{x}_{t_0+k_3} - T(T^{k_3}\tilde{x}_{t_0})| \\ &\quad + |T^{k_3+1}\tilde{x}_{t_0} - p| \leq K\lambda^{k_3}(1 - \lambda) + K\lambda^{k_3+1} = K\lambda^{k_3}.\end{aligned}$$

Thus, if (12.157) and (12.158) hold with sufficiently large C_8, then

$$\tilde{x}_{t_0+k_3+1} \in V_3, \tag{12.162}$$

provided (12.161) holds as well. But we can find an $\varepsilon_2 > 0$ such that $|Tx - Ty| \leq (\varepsilon_1/2)$ for $x, y \in A$ with $|x - y| \leq \varepsilon_2$. If now

$$|\tilde{x}_{t_0+k_3-1} - T^{k_3-1}\tilde{x}_{t_0}| \leq \varepsilon_2, \tag{12.163}$$

and $(5\tau)(\tau/2)^{-(\delta/2)(k_3-1)}C_8^{-(\delta/2)} \leq (\varepsilon_1/2)$, then

$$|\tilde{x}_{t_0+k_3} - T^{k_3}\tilde{x}_{t_0}| \leq |\tilde{x}_{t_0+k_3} - T\tilde{x}_{t_0+k_3-1}| + |T\tilde{x}_{t_0+k_3-1} - T(T^{k_3-1}\tilde{x}_{t_0})| \leq \varepsilon_1,$$

i.e. (12.161). Thus (12.162) holds if (12.157), (12.158) and (12.163) hold with a sufficiently large C_8. After k_3 such steps we find that (12.162) will hold as soon as (12.157) and (12.158) hold with a sufficiently large C_8. However, we can easily estimate the probability of the occurrence of (12.158) on the set where (12.157) holds. Indeed, if (12.157) holds as well as

$$|x_{j+1} - R_jT\tilde{x}_j| \leq |x_j|^{1-(\delta/2)} \text{ for } t_0 \leq j < t, \tag{12.164}$$

then (12.159) is still valid, and by (12.117),

$$P_t\{(12.164) \text{ holds for } j = t\} \geq 1 - K|x_t|^{-\delta} \geq 1 - K(\tau/2)^{-(t-t_0)\delta}|x_{t_0}|^{-\delta}.$$

Thus, by induction,

$$P_{t_0}\{(12.158) \text{ holds}\} \geq \prod_{t=t_0}^{t_0+k_3}\{1 - K(\tau/2)^{-(t-t_0)\delta}|x_{t_0}|^{-\delta}\} \geq 1 - C_9|x_{t_0}|^{-\delta}$$

on the set where (12.157) hold, for suitable $C_9 < \infty$. Combined with the above remarks and (12.159) this proves that there exist $C_8, C_9 < \infty$ such that

$$P_{t_0}\{\tilde{x}_{t_0+k_3+1} \in V_3, |x_{t_0+k_3+1}| \geq (\tau/2)^{k_3+1}|x_{t_0}|\} \geq 1 - C_9|x_{t_0}|^{-\delta} \tag{12.165}$$

on the set where (12.157) holds. If C_8 is chosen so large that also $(\tau/2)^{k_3+1}C_8 \geq C_4$, then (12.165) can be combined with (12.155) with $t_1 = t_0 + k_3 + 1$ to obtain on the set where (12.157) holds that[14]

[14] $E_{t_0}\{X; F\}$ denotes the integral of X over the set F with respect to the measure P_{t_0}. The first inequality in (12.166) is based on the property of conditional expectations

$$P_{t_0}\{F_1 \cap F_2\} = E_{t_0}\{P_t\{F_2\}; F_1\}, t \geq t_0,$$

where F_1, F_2 are events with $F_1 \in \mathscr{F}_t$. Here, $\mathscr{F}_t$ denotes the σ-field of events generated by x_s, $0 \leq s \leq t$.

$$
\begin{aligned}
P_{t_0}\{\lim_{t\to\infty} (x_t/\rho^t) = g\cdot p \text{ for some } g > 0\} \\
\geq E_{t_0}\{P_{t_0+k_3+1}\{\lim_{t\to\infty} (x_t/\rho^t) = g\cdot p \text{ for some } g > 0\}; \\
\tilde{x}_{t_0+k_3+1} \in V_3, |x_{t_0+k_3+1}| \geq (\tau/2)^{k_3+1}|x_{t_0}|\} \\
\geq E_{t_0}\{1 - C_5|x_{t_0+k_3+1}|^{-\delta}; \tilde{x}_{t_0+k_3+1} \in V_3, |x_{t_0+k_3+1}| \geq (\tau/2)^{k_3+1}|x_{t_0}|\} \\
\geq 1 - C_5(\tau/2)^{-\delta(k_3+1)}|x_{t_0}|^{-\delta}P_{t_0}\{\tilde{x}_{t_0+k_3+1} \in V_3, |x_{t_0+k_3+1}| \\
\geq (\tau/2)^{k_3+1}|x_{t_0}|\} \\
\geq \{1 - C_5(\tau/2)^{-\delta(k_3+1)}|x_{t_0}|^{-\delta}\}\{1 - C_9|x_{t_0}|^{-\delta}\} \geq 1 - K_1|x_{t_0}|^{-\delta}
\end{aligned} \tag{12.166}
$$

for suitable $K_1 < \infty$. This proves (12.120) whenever $\tilde{x}_{t_0} \in B$, $|x_{t_0}| \geq C_8$. But clearly (12.120) remains valid for $\tilde{x}_{t_0} \in B$ and $0 < |x_{t_0}| \leq C_8$ if we take $K_1 C_8^{-\delta} \geq 1$. This completes the proof of the theorem. *Q.E.D.*

12.4. Proofs of Theorems 12.1 and 12.2

In this section we will use the results obtained in sections 12.2 and 12.3 to prove theorems 12.1 and 12.2. As in section 12.3, $C, C_1, C_2, \ldots$ denote finite, strictly positive constants. Note that a C_i in this section does not correspond to any C_j in section 12.3.

12.4.1. Proof of theorem 12.1

It is quite clear that on the Ω-set where $\limsup_{t\to\infty} |x_t| < \infty$ the random variable g defined by

$$gW = \lim_{t\to\infty} \{x_t/\lambda_1^t\} \tag{12.167}$$

is well defined and equal to zero. Thus to prove the theorem it suffices to show that g is well-defined on the set where $\limsup_{t\to\infty} |x_t| = \infty$. This we do in two steps.

Step 1. In this step we will establish the following auxiliary result. Suppose that the conditions of theorem 12.1 are satisfied and let

$$y_t \equiv \max(|x_t^1|, (\lambda_1^t/t + 1)), \; t = 0, 1, \ldots .$$

Then there exists a finite positive constant C such that

$$P\{\limsup_{t\to\infty} [|x_t^2|/y_t] > C\} = 0. \tag{12.168}$$

Proof. First note that

$$\lim_{t\to\infty} \{y_{t+1}/y_t\} = \lambda_1 \text{ w. pr. } 1. \tag{12.169}$$

This follows from the following considerations. It follows from theorem 3 in the appendix of ref. [12] (cf. p. A. 1) that w. pr. 1 either $|\boldsymbol{x}_t^1|$ remains bounded or $\lim_{t\to\infty} \{|\boldsymbol{x}_t^1|/\lambda_1^t\}$ exists and is positive. Thus if we ignore an exceptional null Ω-set, then for each sample path ω we can find an integer $t^*(\omega)$ so large that either $y_t = (t+1)^{-1}\lambda_1^t$ for all $t \geq t^*(\omega)$ or $y_t = |\boldsymbol{x}_t^1|$ for all $t \geq t*(\omega)$. But if that is true, then (12.169) is necessarily valid.

Next let $\boldsymbol{V} = (\boldsymbol{0}, \bar{\boldsymbol{V}}^2)$ and

$$\gamma_t = \max_{d_1+1\leq j\leq d} \{x_{t,j}/V(j)\}. \tag{12.170}$$

Moreover, let $\varepsilon > 0$ be so small that

$$\lambda_2 + 4\varepsilon < \lambda_1, \text{ and } \lambda_1 - \varepsilon > 1. \tag{12.171}$$

Finally, let C_1 be a finite positive constant so large that

$$\max_{d_1+1\leq j\leq d} \{h_j(\boldsymbol{u}, [\boldsymbol{x}^2/|\boldsymbol{x}^2|])/V(j)\} \leq (\lambda_2 + \varepsilon) \max_{d_1+1\leq j\leq d} \{x(j)/|\boldsymbol{x}^2|\, V(j)\} \tag{12.172}$$

whenever $|\boldsymbol{u}| \leq (1/C_1)$. Such a C_1 exists by the uniform continuity of $h_j(\boldsymbol{x}^1, \boldsymbol{x}^2)$ on $|\boldsymbol{x}^1| \leq 1$, $|\boldsymbol{x}^2| \leq 1$ and by the inequalities

$$h_j(\boldsymbol{0}, \boldsymbol{x}^2) \leq \max_{d_1+1\leq i\leq d} [x(i)/V(i)] \cdot h_j(\boldsymbol{0}, \bar{\boldsymbol{V}}^2) = \lambda_2 \max_{d_1+1\leq i\leq d} [x(i)/V(i)] \cdot V(j),$$

which are valid for $d_1 + 1 \leq j \leq d$. We now show that there exist finite positive constants C_2 and C_3 such that

$$P\{|\boldsymbol{x}_{t+1}^2| > 2C_1C_2y_t \,|\, \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t, |\boldsymbol{x}_t^2| \leq C_1y_t\} \leq C_3y_t^{-2\delta}, \tag{12.173}$$

$$P\{\gamma_{t+1} > (\lambda_2 + 2\varepsilon)\gamma_t \,|\, \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t, |\boldsymbol{x}_t^2| > C_1y_t\} \leq C_3y_t^{-2\delta}. \tag{12.174}$$

To establish (12.173) proceed as follows. Let

$$C_2 = \{\max_{\substack{|\boldsymbol{u}|\leq 1/C_1\\ |\boldsymbol{v}|\leq 1}} |\boldsymbol{h}^2(\boldsymbol{u}, \boldsymbol{v})|\} \tag{12.175}$$

and note that on the set where $|x_t^2| \le C_1 y_t$

$$|E\{x_{t+1}^2 \mid x_0, \ldots, x_t\}| = |h^2(x_t^1, x_t^2)| \le C_1 y_t |h^2([x_t^1/C_1 y_t], [x_t^2/C_1 y_t])| \le C_1 C_2 y_t. \tag{12.176}$$

Hence, by (12.26) and Chebychev's inequality

$$\begin{aligned} P\{|x_{t+1}^2| > 2C_1C_2y_t \mid x_0, \ldots, x_t, |x_t^2| \le C_1 y_t\} \\ \le P\{|x_{t+1}^2 - h^2(x_t^1, x_t^2)| \ge C_1C_2y_t \mid x_0, \ldots, x_t, |x_t^2| \le C_1 y_t\} \\ \le \frac{K|x_t|^{2(1-\delta)}(d-d_1)^3}{(C_1C_2y_t)^2} \le C_3 y_t^{-2\delta} \end{aligned} \tag{12.177}$$

for some suitable constant C_3.

To establish (12.174) note that (12.172) implies that

$$\begin{aligned} &\{\omega: \gamma_{t+1} > (\lambda_2 + 2\varepsilon)\gamma_t, |x_t^2| > C_1 y_t\} \\ &\subset \bigcup_{j=d_1+1}^{d} \{\omega: (x_{t+1,j}/V(j)) > (\lambda_2 + 2\varepsilon)\gamma_t, |x_t^2| > C_1 y_t\} \\ &\subset \bigcup_{j=d_1+1}^{d} \{\omega: (x_{t+1,j}/V(j)) > (|x_t^2|\, h_j\left(\frac{x_t^1}{|x_t^2|}, \frac{x_t^2}{|x_t^2|}\right)/V(j)) \\ &\quad + \varepsilon\gamma_t, |x_t^2| > C_1 y_t\} \\ &\subset \bigcup_{j=d_1+1}^{d} \{\omega: |x_{t+1,j} - h_j(x_t^1, x_t^2)| > \varepsilon\gamma_t V(j), |x_t^2| > C_1 y_t\}. \end{aligned} \tag{12.178}$$

Hence, by (12.26), (12.170) and Chebychev's inequality

$$\begin{aligned} P\{\gamma_{t+1} > (\lambda_t + 2\varepsilon)\gamma_t \mid x_0, \ldots, x_t, |x_t^2| > C_1 y_t\} \\ \le \frac{K(d-d_1)|x_t|^{2(1-\delta)}}{\{\min_{d_1+1 \le j \le d} V(j)^2\}\varepsilon^2\gamma_t^2} \le C_3 y_t^{-2\delta} \end{aligned} \tag{12.179}$$

for some suitable finite constant C_3.

So much for (12.173) and (12.174). Next, let

$$A_t = \{\omega: |x_{t+1}^2| > 2C_1C_2y_t, |x_t^2| \le C_1 y_t\}, t = 0, 1, \ldots$$

and let

$$B_t = \{\omega: \gamma_{t+1} > (\lambda_2 + 2\varepsilon)\gamma_t, |x_t^2| > C_1 y_t\}, t = 0, 1, \ldots$$

and observe that

$$\sum_{t=0}^{\infty} y_t^{-2\delta} \le \sum_{t=0}^{\infty} (t+1)^{2\delta}(\lambda_1^{-2\delta})^t < \infty. \tag{12.180}$$

From (12.180), (12.173), (12.174) and from corollary 2, p. 324 in ref. [3] it follows that w. pr. 1 only finitely many of the events A_t and B_t occur. Thus there exists a random integer t_0 such that for all $t \geq t_0$,

$$|x_t^2| \leq C_1 y_t \Rightarrow |x_{t+1}^2| \leq 2C_1C_2 y_t, \tag{12.181}$$

$$|x_t^2| > C_1 y_t \Rightarrow \gamma_{t+1} \leq (\lambda_2 + 2\varepsilon)\gamma_t, \tag{12.182}$$

and

$$y_{t+1} > (\lambda_1 - \varepsilon) y_t \geq (\lambda_2 + 3\varepsilon) y_t \quad \text{(recall (12.169))}. \tag{12.183}$$

In the remainder of this step we ignore the exceptional Ω-set on which no such finite t_0 exists. Moreover, we let $t_1(\omega)$ be the first time after and including t_0 that $|x_{t_1}^2| \leq C_1 y_{t_1}$ and let $t_i(\omega)$, $i > 1$, be the first time after $t_{i-1}(\omega)$ that $|x_{t_i}^2| \leq C_1 y_{t_1}$. The existence and almost sure finiteness of $t_i(\omega)$ is an immediate consequence of (12.170) and (12.181). (See also estimates (12.184) – (12.186).)

It is now easy to prove (12.168) with $C = 2C_1(C_2 + 1)M(1 + |V|)$, where

$$M = (1 + \max_{d_1+1 \leq i \leq d} (1/V(i))).$$

Indeed, if $t_{i+1} = t_i + 1$, then

$$\max_{t_i < t \leq t_{i+1}} \{|x_t^2|/y_t\} = \{|x_{t_i+1}^2|/y_{t_i+1}\} \leq C_1. \tag{12.184}$$

If on the other hand $t_{i+1} > t_i + 1$, then by (12.170) and (12.181)

$$\{\gamma_{t_i+1}/y_{t_i+1}\} \leq M\{|x_{t_i+1}^2|/y_{t_i+1}\} \leq 2C_1C_2M, \tag{12.185}$$

and by (12.182), (12.183) and (12.185) for all $t_i + 1 \leq t < t_{i+1}$

$$\begin{aligned}\{|x_{t+1}^2|/y_{t+1}\} &\leq \{\gamma_{t+1}|V|/y_{t+1}\} \leq \{(\lambda_2 + 2\varepsilon)\gamma_t|V|/(\lambda_2 + 3\varepsilon)y_t\} \\ &\leq \{\gamma_t|V|/y_t\} \leq \dots \leq \{\gamma_{t_i+1}|V|/y_{t_i+1}\} \\ &\leq 2C_1C_2M|V|. \end{aligned} \tag{12.186}$$

Thus, in any case, for all $i \geq 1$,

$$\sup_{t_i < t \leq t_{i+1}} \{|x_t^2|/y_t\} \leq C,$$

and hence

$$\sup_{t \geq t_1} \{|x_t^2|/y_t\} \leq C.$$

Step 2. The preceding step shows that on the set where $\limsup_{t\to\infty} |x_t^1| < \infty$,

$$\lim_{t\to\infty} \{|x_t^2|/\lambda_1^t\} = 0 \text{ a.e. (and of course also } \limsup_{t\to\infty} \{|x_t^1|/\lambda_1^t\} = 0). \tag{12.187}$$

Hence g as defined in (12.167) is well-defined and equals zero a.e. on the set where $\lim_{t\to\infty} \sup |x_t^1| < \infty$. In this step we will show that g is well-defined and positive a.e. on the set where $\limsup_{t\to\infty} |x_t^1| = \infty$.

Let $C_4 = 2C$ and let

$$S = \{\omega: \limsup_{t\to\infty} |x_t^1| = \infty\}. \tag{12.188}$$

As pointed out above it was shown in theorem 3 in the appendix to ref. [12] that a.e. on S $\lim_{t\to\infty} \{x_t^1/\lambda_1^t\} = g_1 V^1$ for some $g_1 > 0$. Thus, a.e. on S $\{|x_t^1|/\lambda_1^t\}$ is bounded away from zero and $y_t = |x_t^1|$ eventually by step 1. From this it follows that a.e. on S we can find an integer $t^{**}(\omega)$ such that for all $t \geq t^{**}(\omega)$,

$$\frac{|x_t^1(\omega)|}{|x_t^1(\omega)| + |x_t^2(\omega)|} = \left\{\frac{1}{1 + [|x_t^2|/|x_t^1|]}\right\} \geq \frac{1}{1 + C}. \tag{12.189}$$

Hence, for $D = (1/(1 + C)) > 0$,

$$\liminf_{t\to\infty} \{|x_t^1|/|x_t|\} \geq D \text{ a.e. on } S. \tag{12.190}$$

Next, let $A = \{x \geq 0: \ |x| = 1\}$ $\quad Tx = \{h(x)/|h(x)|\}$, $x \in A$, and $R(x) = |h(x)|$. By lemmas 12.1 and 12.2, $\tilde{W} = (W/|W|)$ is strictly positive. Hence we can find a $0 < \varepsilon < D$ with $|\tilde{W}^1| > \varepsilon$. Let $B \equiv \{x \in A: |x^1| \geq \varepsilon\}$. Without loss of generality we may assume that the neighborhood U specified in theorem 12.1 is contained in B. It then follows from lemmas 12.2 and 12.4 that T satisfies (12.114) with $p = \tilde{W} = \{W/|W|\}$ and (12.116) with $\lambda = \alpha$. Moreover, by (12.24) and by the fact that the map $x^2 \to H^2(0, x^2)$ is indecomposable and $\lambda_2 > 0$, there exists a $\tau > 0$ such that

$$|h(x)| \geq \tau > 0 \text{ for all } x \in A. \tag{12.191}$$

(12.27) together with (12.191) implies that T satisfies (12.115) and that $R(x)$ satisfies (12.119). In addition, since $h(\cdot)$ is homogeneous of degree one, (12.118) also follows from (12.27) if we let

$$\rho = |h(\tilde{W})| = |\lambda_1 \tilde{W}| = \lambda_1 \text{ and } \beta = 1.$$

Indeed,

$$\{R(\boldsymbol{x})/|\boldsymbol{x}|\} = \{|\boldsymbol{h}(\boldsymbol{x})|/|\boldsymbol{x}|\} = |\boldsymbol{h}(\tilde{\boldsymbol{x}})|$$

and

$$\left||\boldsymbol{h}(\tilde{\boldsymbol{x}})| - |\boldsymbol{h}(\tilde{\boldsymbol{W}})|\right| \leq |\boldsymbol{h}(\tilde{\boldsymbol{x}}) - \boldsymbol{h}(\tilde{\boldsymbol{W}})| \leq K|\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{W}}|.$$

Finally, (12.117) is an immediate consequence of (12.25), (12.26) and Chebychev's inequality. Thus all conditions of theorem 12.3 hold for T and the family $\{\boldsymbol{x}_t\}_{t \geq 0}$.

Consequently, for any integer $t_0 \geq 0$,

$$P_{t_0}\{\lim_{t\to\infty} \{\boldsymbol{x}_t/\lambda_1^t\} = g\tilde{\boldsymbol{W}} \text{ for some } g > 0\} \geq 1 - K_1 |\boldsymbol{x}_{t_0}|^{-\delta} \quad (12.192)$$

almost everywhere on the set $\{\boldsymbol{x}_{t_0} \neq 0, \tilde{\boldsymbol{x}}_{t_0} \in B\}$. Now, let $\eta > 0$, $L_\eta = (K_1/\eta)^{(1/\delta)}$ and $k = k(\eta, \omega) = \inf\{n: |\boldsymbol{x}_n(\omega)| \geq L_\eta, \tilde{\boldsymbol{x}}_n(\omega) \in B\}$. By (12.190), the definition of S and the choice of ε,

$$k < \infty \text{ a.e. on } S. \quad (12.193)$$

Finally, by (12.192) and the fact that

$$\tilde{\boldsymbol{x}}_s \in B \text{ and } 1 - K_1|\boldsymbol{x}_s|^{-\delta} \geq 1 - K_1 L_\eta^{-\delta} = 1 - \eta \text{ on } \{k(\eta, \omega) = s\},$$

we have

$$\begin{aligned}
&P\{\lim_{t\to\infty} \{\boldsymbol{x}_t/\lambda_1^t\} = g\tilde{\boldsymbol{W}} \text{ for some } g > 0 \text{ and } k(\eta, \omega) < \infty\} \\
&\quad = \sum_{s=0}^{\infty} P\{k = s \text{ and } \lim_{t\to\infty} \{\boldsymbol{x}_t/\lambda_1^t\} = g\tilde{\boldsymbol{W}} \text{ for some } g > 0\} \\
&\quad = \sum_{s=0}^{\infty} P\{k = s\} P_s\{\lim_{t\to\infty} \{\boldsymbol{x}_t/\lambda_1^t\} = g\tilde{\boldsymbol{W}} \text{ for some } g > 0 \mid k = s\} \\
&\quad \geq (1 - \eta) \sum_{s=0}^{\infty} P\{k = s\} = (1 - \eta) P\{k(\eta, \omega) < \infty\}. \qquad (12.194)
\end{aligned}$$

Thus, for *any* $\eta > 0$ (cf. (12.143)),

$$\begin{aligned}
&P\{\{\boldsymbol{x}_t\} \in S \text{ but not } \lim_{t\to\infty} \{\boldsymbol{x}_t/\lambda_1^t\} = g\tilde{\boldsymbol{W}} \text{ for any } g > 0\} \\
&\quad \leq P\{\{\boldsymbol{x}_t\} \in S \text{ and } k(\eta, \omega) = \infty\} \\
&\quad + P\{k(\eta, \omega) < \infty \text{ but not } \lim_{t\to\infty} \{\boldsymbol{x}_t/\lambda_1^t\} = g\tilde{\boldsymbol{W}} \text{ for any } g > 0\} \\
&\quad \leq \eta P\{k(\eta, \omega) < \omega\} \leq \eta.
\end{aligned}$$

That is, $\lim_{t\to\infty} \{x_t/\lambda_1^t\} = g\tilde{W}$ for some $g > 0$ a.e. on S. Thus (12.28) has been proved in general. Essentially the same argument shows

$$P\{g > 0 \mid x_0\} \geq (1 - \eta)P\{k(\eta, \omega) < \infty \mid x_0\} > 0$$

as soon as

$$P\{k(\eta, \omega) < \infty \mid x_0\} > 0 \text{ for some } 0 < \eta < 1.$$

The above proof is applicable for any $0 < \varepsilon < \min(D, |\tilde{W}^1|)$. Of course K_1 and L_η depend on ε and also B is defined in terms of ε. To bring out this dependence we shall write $K_1(\varepsilon)$, $L_\eta(\varepsilon)$, and $B(\varepsilon)$ for the remainder of the proof of (12.29). Then, for any x_0 with $x_0^1 \neq 0$ we have

$$\tilde{x}_0 \in B(|x_0^1|/|x_0|),$$

and

$$k(\tfrac{1}{2}, \omega) = 0 \text{ on } \{|x_0| \geq L_{(1/2)}(|x_0^1|/|x_0|) = [2K_1(|x_0^1|/|x_0|)]^{(1/\delta)}\},$$

i.e.

$$P\{g > 0 \mid x_0\} \geq \tfrac{1}{2} \text{ on } \{x_0^1 \neq 0, |x_0| \geq [2K_1(|x_0^1|/|x_0|)]^{(1/\delta)}\}.$$

Clearly $E\{g \mid x_0\} > 0$ as soon as $P\{g > 0 \mid x_0\} > 0$. Hence $P\{g > 0 \mid x_0\} > 0$ and $E\{g \mid x_0\} > 0$ whenever $x_0^1 \neq 0$, $|x_0|$ sufficiently large. On the other hand, on $x_0^1 = 0$, $E_0\{x_1^1\} = H^1(0) = 0$. So $x_1^1 = 0$ and by induction $x_t^1 = 0$ a.e. on $\{x_0^1 = 0\}$. This completes the proof of (12.29) and (12.30) as well as the comment between them, with $M(y) = [2K_1(y)]^{1/\delta}$.

Lastly, when $H^1(\cdot) = (h_1(\cdot), \ldots, h_{d_1}(\cdot))$ is concave, the following inequality holds:

$$\begin{aligned} E\{(x_t^1/\lambda_1^t) \mid x_0\} &= E\{E\{(x_t^1/\lambda_1^t) \mid x_0, \ldots, x_{t-1}\} \mid x_0\} \\ &= (1/\lambda_1)E\{H^1((x_{t-1}^1/\lambda_1^{t-1})) \mid x_0\} \\ &\leq (1/\lambda_1)H^1(E\{(x_{t-1}^1/\lambda_1^{t-1}) \mid x_0\}) \\ &\leq \ldots \leq (1/\lambda_1)^t(H^1)^t(x_0). \end{aligned} \tag{12.195}$$

Since by lemma 12.3,

$$\limsup_{t\to\infty} (1/\lambda_1^t)|(H^1)^t(x_0)| = \lim_{t\to\infty} (1/\lambda_1^t)|(H^1)^t(x_0)| = \gamma(x_0)|W^1| < \infty,$$

we find from Fatou's lemma that

$$\begin{aligned} E\{g \mid x_0\} &= (1/|\tilde{W}^1|)E\{\lim_{t\to\infty}(|x_t^1|/\lambda_1^t) \mid x_0\} \\ &\le (1/|\tilde{W}^1|)\limsup_{t\to\infty} E\{(|x_1^t|/\lambda_1^t) \mid x_0\} \\ &= \gamma(x_0)\,|W| < \infty, \end{aligned}$$

which proves (12.31). *Q.E.D.*

12.4.2. Proof of theorem 12.2

We closely follow step 2 in the proof of theorem 12.1. Again we take $A = \{x \ge 0\colon |x| = 1\}$, $Tx = \{h(x)/|h(x)|\}$, and $R(x) = |h(x)|$. For p and B we now take the vector $(0, (\bar{V}^2/|\bar{V}^2|))$ and A respectively. Finally U is the neighborhood of p specified in the assumptions of theorem 12.2. (12.114) is now satisfied by the definition of $\bar{V}^2$ (see (12.14)). Essentially, as in step 2 of the proof of theorem 12.1, (12.115), (12.118) and (12.119) follow from (12.33) and (12.27) if we now take

$$\rho = |h(0, (\bar{V}^2/|\bar{V}^2|))| = \lambda_2,\ \beta = 1,$$

(12.117) is a consequence of (12.25), (12.26) and Chebychev's inequality as before. Lastly, (12.116) with $\lambda = \Psi$ as in (12.72) is implied by lemma 12.6 since (12.27) implies the validity of (12.71) with $K_6 = (2K + \lambda_2)|V^1|$ and $\gamma = 1$. An application of theorem 12.3 now gives for any $\eta > 0$ an L_η such that

$$P_{t_0}\{\lim_{t\to\infty}(x_t/\lambda_2^t) = g(0, \bar{V}^2) \text{ for some } g > 0\} \ge 1 - \eta \text{ on the set } \{|x_{t_0}| \ge L_\eta\}.$$

As in step 2 of theorem 12.1 we have

$$P\{\lim_{t\to\infty}(x_t/\lambda_2^t) = g(0, \bar{V}^2) \text{ for some } g > 0\} \ge (1-\eta)P\{|x_s| \ge L_\eta \text{ for some } s\},$$

and

$$\lim_{t\to\infty}\{x_t/\lambda_2^t\} = g(0, \bar{V}^2) \text{ for some } g > 0 \tag{12.196}$$

a.e. on the set

$$S = \bigcap_{\eta>0}\{|x_s| \ge L_\eta \text{ for some } s\} = \{\limsup_{s\to\infty}|x_s| = \infty\}. \tag{12.197}$$

Clearly, $\lim_{t\to\infty} \{x_t/\lambda_2^t\} = 0$ on the complement of S so that by (12.196) g can be defined a.e. by

$$\lim_{t\to\infty} \{x_t/\lambda_2^t\} = g\cdot(\mathbf{0}, \bar{V}^2).$$

Moreover, $g > 0$ a.e. on S and $g = 0$ off S. This proves (12.35) and the remainder of the theorem with $M = L_{\frac{1}{2}}$ is proved in the same way as (12.29) and (12.31). *Q.E.D.*

References

[1] Bellamy, E. In *Looking Backward 2000–1778,* ed. John L. Thomas. Harvard University Press: Cambridge, Mass. (1967).

[2] Burmeister, E. and Dobell, A. R. *Mathematical Theories of Economic Growth.* The Macmillan Company: New York (1970).

[3] Doob, J. L. *Stochastic Processes.* Wiley: New York (1953).

[4] Fisher, F. M. Decomposability, near decomposability, and balanced price change under constant returns to scale. *Econometrica,* **31** (1963), 67–89.

[5] Graves, L. M. *Theory of Functions of Real Variables.* McGraw-Hill: New York (1956).

[6] Kesten, H. Quadratic transformations: a model for population growth I and II. *Advances in Applied Probability,* **2** (1970), 1–82.

[7] Kesten, H. and Stigum, B. P. Limit theorems for decomposable multi-dimensional Galton–Watson processes. *Journal of Mathematical Analysis and Applications,* **17** (1967), 307–338.

[8] Mirman, L. J. The steady state behavior of a class of one sector growth models with uncertain technology. *Journal of Economic Theory,* **6** (1973), 346–354.

[9] Nikaido, H. Convex structures and economic theory. Academic Press: New York (1968).

[10] Solow, R. M. A contribution to the theory of economic growth. *Quarterly Journal of Economics,* **70** (1956), 65–94.

[11] Solow, R. M. and Samuelson, P. A. Balanced growth under constant returns to scale. *Econometrica,* **21** (1953), 412–424.

[12] Stigum, B. P. Balanced growth under uncertainty. *Journal of Economic Theory,* **5** (1972), 42–68.

COMMENTS

Balanced growth under uncertainty in decomposable economies

Roy Radner

In this paper with H. Kesten and in a previous paper, 'Balanced growth under uncertainty', Professor Stigum has given us an interesting analysis of a class of stochastic growth processes in which the distribution of the state vector is asymptotically concentrated on a ray through the origin. Roughly speaking, the mathematical expectation of the (vector valued) process is assumed to be governed by equations like those studied by Solow and Samuelson in their paper 'Balanced growth under constant returns to scale'. Furthermore, the standard deviation of each coordinate of the process is assumed to grow more slowly than the 'length' of the state vector. The expected value of the process grows asymptotically like λ^t, where λ is greater than 1. I may perhaps paraphrase the argument heuristically as follows. For those realizations of the process $\boldsymbol{x}(t)$ that grow more slowly than λ^t, the 'normalized' process, $\boldsymbol{x}(t)/\lambda^t$, approaches 0; whereas for those realizations that grow as fast as λ^t the conditional standard deviation of $x_i(t)/\lambda^t$, given the past, approaches 0, and so the normalized process converges to a particular ray through the origin (this heuristic argument does not, of course, do justice to the complexity of the problem).

Without detracting from the interest of the paper, I would like here to amplify Professor Stigum's own remarks concerning classes of stochastic economies that do *not* satisfy his assumptions (12.25) and (12.26) about the variance of the process. First, it may clarify his remark 1 on p. 347 if we consider a special case of the example on p. 340 in which $G(\boldsymbol{x}, A) = F(\boldsymbol{x})A$. It is easily verified that if F is homogeneous of degree one, then the conditional variance of the second coordinate of $\boldsymbol{x}_{t+1}$, given the past, is equal to

$$s^2 F\left(\frac{\boldsymbol{x}_t}{|\boldsymbol{x}_t|}\right)^2 (\text{Var } A)\, |\boldsymbol{x}_t|^2,$$

thus assumption (12.26) is not satisfied.

I think that this situation can be expected to be quite common. Thus, in a von Neumann model in which the input and output coefficients are random variables, assumptions (12.25) and (12.26) will typically not be satisfied.

Other concepts of balanced growth are of interest if the distribution of the state vector is *not* asymptotically concentrated along a single ray through the origin. For example, we may say that the process $\{x_t\}$ exhibits weakly balanced growth, if there exist stationary processes $\{\Lambda_t\}$ and $\{\bar{x}_t\}$, with values, respectively, in R and R^n such that $\bar{x}_t = x_t/\Lambda^{(t)}$, where $\Lambda^{(t)} = \Lambda_1\Lambda_2 \ldots \Lambda_t$. Λ, a random variable that has the same distribution as all the Λ_ts, is called the steady growth factor.

The following example is due to P. Jeanjean[1].

We have n commodities, and n industries, each one producing one commodity with the production function

$$y_j = r_j\beta_j \prod_{i=1}^{n} x_{ij}^{\alpha_{ij}},$$

where x_{ij} is the input of the ith commodity in the jth process, β_j and α_{ij} are constants with

$$\alpha_{ij} \geq 0, \qquad \sum_{i=1}^{n} \alpha_{ij} = 1, \qquad \beta_j > 0,$$

and the r_js are non-negative, bounded random variables not necessarily independent, such that for every j, $Er_j = 1$ and $r_j \geq \varepsilon_j > 0$.

When the total stock of good x is available, it is allocated among the different industries according to a matrix

$$K = (k_{ij}), \qquad k_{ij} \geq 0, \qquad \sum_{j=1}^{n} k_{ij} \leq 1$$

in such a way that $x_{ij} = k_{ij}x_i$. The matrix K is thus the control variable.

It is useful to reformulate the problem in logarithmic terms. Denote $A = (\alpha_{ij})' =$ transpose of the matrix (α_{ij}),

$$y = \begin{bmatrix} \log y_1 \\ \cdot \\ \cdot \\ \cdot \\ \log y_n \end{bmatrix}, R = \begin{bmatrix} \log r_1 \\ \cdot \\ \cdot \\ \cdot \\ \log r_n \end{bmatrix}, \beta = \begin{bmatrix} \log \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \log \beta_n \end{bmatrix}, \eta = \begin{bmatrix} \eta_1(K) \\ \cdot \\ \cdot \\ \cdot \\ \eta_n(K) \end{bmatrix},$$

[1] P. Jeanjean. Optimal growth with stochastic technology in a closed economy. Unpublished dissertation, University of California, Berkeley (1972).

where

$$\eta_j(K) = \sum_{i=1}^{n} \alpha_{ij} \log k_{ij}.$$

For feasibility, $x_i(t + 1) = y_i(t)$.

Denote $\zeta = \beta + \eta(K)$. The law of evolution of the system is then $Y(t + 1) = R + \zeta + AY(t)$ which admits the explicit solution

$$Y(t + 1) = A^t Y(0) + \sum_{k=1}^{t} A^k \zeta + \sum_{k=1}^{t} A^k R(k), \qquad t \geq 1,$$

where $R(t)$ are the successive values of the random vector R, assumed independent from each other, and $Y(0)$ is the initial endowment.

Let us assume:

(1) the random variables $R_k = \log r_k$ have second order moment; and

(2) the matrix A is fully regular[2].

It follows from the second assumption that there exists a non-negative matrix $\bar{A}$, whose rows are all identical – denote them by a – such that

$$A^t \xrightarrow[t \to \infty]{} \bar{A} = \begin{bmatrix} a \\ \cdot \\ \cdot \\ \cdot \\ a \end{bmatrix}.$$

It is possible to show that there exists a random vector Ω, with $E\Omega = 0$, and a non-random vector $\bar{Y}$, such that

$$Y(t) - t\bar{A}\left[\zeta + \frac{1}{t}\sum_{k=1}^{t} r(k)\right] \xrightarrow[t \to \infty]{\text{a.s.}} \bar{Y} + \Omega.$$

Asymptotically, the economy exhibits weakly balanced growth at a steady rate $a(\zeta + R)$ (the same for each commodity).

The quantity

$$f(t) = a\left[\zeta + \frac{1}{t}\sum_{k=1}^{t} R(k)\right]$$

[2] A fully regular matrix is a matrix with non-negative coefficients, each of whose columns sums to 1, and for which 1 is the only eigenvalue of modulus 1 and is a simple root of the characteristic equation.

may be considered as the average rate of growth. By an ergodic theorem,

$$f(t) \xrightarrow{\text{a.s.}} \rho = a(\zeta + ER) = E\rho(t)$$

ρ is the asymptotic long-run rate of growth.

If we call the *associated deterministic process* the process obtained by taking the expected output as deterministic output, its asymptotic rate of growth is $a\zeta$. Hence, the asymptotic long-run rate of growth of the economy is well defined and not random. It is strictly smaller than the asymptotic rate of growth of the associated deterministic process.

On the application spectrum for the Kesten–Stigum theory

Michael Balch

This paper by Professors Kesten and Stigum is a continuation of the analysis begun by the latter author (1972), in which the standard deterministic growth model of neoclassical theory is embedded in a class of stochastic processes that are Solow–Samuelson 'in expectation'. Thus, while the instantaneous evolution of state variables is random, to be sure, it is assumed that the corresponding vector of mathematical expectations satisfies a growth equation of the Solow–Samuelson type[1]. The object of their study is to restrict this class so that the principal structural inference of the neoclassical model – namely, that the system evolves along a distinguished ray of balanced growth – carries over in an appropriate asymptotic way for the more general stochastic setting. This is neatly achieved by requiring (conditions (12.25), (12.26)) that the standard deviation of each component of the (vector-valued) process be bounded in terms of current state size; very roughly put, the 'first moment' influence of past realizations must dominate the 'second moment' influence of current uncertainties.

[1] This provides the structural 'skeleton' for the Kesten–Stigum analysis. But see their remark 4 on relaxing condition (12.1), which provides far greater generality than the strict 'equation' form of (12.1) would suggest (some might find it vaguely disturbing that the evolution of the conditional expectation of a random macro-process can be known with such precision).

In the comment preceding this one Professor Radner questions whether the Kesten–Stigum sub-class is sufficiently robust – as a matter of what we conceive to be real world essence – to support our continued confidence in the sharpness of the neoclassical result; otherwise, as Professor Radner has suggested, we may have to settle for a 'fattened' version of balanced growth that more faithfully characterizes the stochastic situation. Radner's concern arises in connection with the Mirman (1973) model for macrostochastic growth which the present authors employ to indicate the scope and 'minimally sufficient' character of their theory. They note that the Mirman model neither belongs to their process class nor enjoys the asymptotic behavior that would otherwise follow. Radner sees this as a typical sort of limitation on the application spectrum for the Kesten–Stigum theory.

It seems to me, however, that the case for the Kesten–Stigum theory is stronger than might appear from a simple comparison with the Mirman model, since, in the latter, microrandomness is assumed to 'augment' with macroscale according to an incidence mode that does not take macrostochastic account of diversifications in the productive sector. More particularly, the Mirman model does not incorporate production contingencies of the 'firm-specific' type, i.e. of the type that may be expected to impact *separately* and in randomly incident fashion upon the individual productive units of a decentralized economy. Rather, analysis is based upon the polar assumption that, at any given epoch, all contingencies have universal incidence and 'lockstep' impact across firms[2]. In the case of a simple corn economy, for example, the Mirman mode does not allow for real production[3] contingencies that

[2] In fairness, it should be noted that this microfoundational characterization is not explicit in the Mirman construction; rather it follows inferentially from the nature of the law that is taken to govern macrostochastic evolution. The point is that random *aggregate* product is the summation of random *individual* products, and the stochastic nature of the former depends upon the incidence mode that is assumed for the latter.

[3] It is worth emphasizing that for the family of macrogrowth processes under present discussion, the formal concern is with modeling randomness that associates to the production function alone – that is, to those 'natural' determinants of real product that remain beyond the conditionalizing influence of the input decision – rather than to the (derivative) market conditions that will obtain once aggregate product has been realized. Thus, while *ex post* output prices do, of course, jointly contribute to the determination of microrewards, and while such market realizations do impact uniformly across all microunits in a given sector, these market variables do not feature explicitly in the 'real' equations at hand and so do not formally concern us here (though perhaps should, as Professor Kurz argues in chapter 13).

do not visit the (locationally separated) farms of that economy in universal fashion; as, for example, with selectively incident drought conditions, hailstorms, floods and the like. The effect of the Mirman mode is that scale-normalized variance for the total process is model-theoretically divorced from the (growing) scale of economic activity (in particular, does not shrink with scale), and it is just here that the Kesten–Stigum condition (12.26) is failed.

Even so, it is interesting to note that the Mirman model does manage to sit on the 'boundary' of the Kesten–Stigum class and, indeed, achieves class membership when the incidence mode for basic randomness is modified in the sense described above. To see this we may consider a single-good economy[4] in which capital growth is somehow modeled to take place on the *extensive* margin, i.e. through the entry of new firms rather than by capital augmentation of old ones. This could follow under an appropriate set of assumptions on the locational and (real) transactions cost structure of the economy so that – at least for some relevant interval of capital/labor ratios – an optimal capital size for individual firms is determinate, finite and, for the purpose of this present cartoon, independent of capital/labor ratio. Thus, for simplicity, suppose that at any epoch t the economy consists of a number of identical firms each of which utilizes the services of a fixed quantum of capital – say normalized at 1 per firm – so that there are $x_{t,2}$ firms at epoch t. Labor services are perfectly divisible, and the product of the (unit-quantized) ith firm is given, say, by $A^i f(x_{t,1}/x_{t,2})$, where the random variables A^i represent microrandomness of the decentralized and firm-specific type discussed above, and where the full employment argument of f is justified by the usual assumptions $f(0) = 0, f' > 0$ and $f'' < 0$. We may suppose that the A^i $(i = 1, \ldots, x_{t,2}, \ldots)$ are identically and independently distributed[5] and, for further simplicity, that these are distributed independently of $\boldsymbol{x}_t$. Then random macroproduct at epoch $t + 1$ is given by

$$F(x_{t,1}, x_{t,2}; \{A^i\}_1^\infty) = \sum_{i=1}^{x_{t,2}} A^i f(x_{t,1}/x_{t,2}).$$

---

[4] Like the one described in Kesten–Stigum by (12.3), (12.4) and (12.7), except as modified in obvious ways below.

[5] The corresponding assumption for the Mirman mode is that (unit-quantized) microproduct is given by $Af(x_{t,1}/x_{t,2})$, where the single random variable A has simultaneous and universal incidence across all firms.

Expected macroproduct is clearly homogeneous of degree 1 in $\boldsymbol{x}_t$. Moreover, in view of the independence of the A^i's, we have

$$V(F(\boldsymbol{x}_t; \cdot) \mid \boldsymbol{x}_0, \ldots, \boldsymbol{x}_t) = V(A) f^2(x_{t,1}/x_{t,2}) x_{t,2}.$$

Since $f(\cdot)$ is bounded in the 'relevant' capital/labor interval posited above (and since $|x_{t,2}| \leq |x_t|$) it follows that the Kesten–Stigum condition (12.26) is satisfied (with $\delta = \frac{1}{2}$) and that their theorem 12.1 obtains.

CHAPTER 13

THE KESTEN–STIGUM MODEL AND THE TREATMENT OF UNCERTAINTY IN EQUILIBRIUM THEORY

Mordecai Kurz*

13.1. Introduction

The treatment of uncertainty in social theory has always been a subject for diverse viewpoints and the present time is no exception. The view of uncertainty adopted in the papers by Kesten–Stigum and Radner which were presented in this meeting raises some very sharp questions. I shall devote the first section below to reviewing the Kesten–Stigum paper and then turn to the broader issues which this theory raises.

13.2. Review of the Kesten-Stigum Theory

We note first that the theory at hand has been developed in two papers. The first is an earlier paper by Stigum [8] entitled 'Balanced growth under uncertainty' and the second is the one by Kesten and Stigum entitled 'Balanced growth under uncertainty in decomposable economies'. These papers represent an extension of the earlier Samuelson–Solow [7] theory of balanced growth to the case in which the environment is random.

The model describes the random evolution of the aggregate stock vector $x(t)$ in the economy, where the random variable $x(t)$ must satisfy

$$E\{x(t+1) \mid x(0), x(1), \ldots, x(t)\} = H(x(t)),$$

* In preparing these remarks I benefited from conversations and correspondence with K. J. Arrow to whom I am indebted.

where $\boldsymbol{H}(\boldsymbol{x}(t))$ is a function which incorporates the technology, resources and rules of intertemporal allocations. It is to be understood that in the uncertainty case the economy has a vector $\boldsymbol{x}(t)$ of stocks which are used to produce output; part of this output is consumed and part reinvested. Thus the relation $\boldsymbol{x}(t+1) = \boldsymbol{H}(\boldsymbol{x}(t))$ means that the state of technology and resources as well as rules of production and consumption decisions are part of the function $\boldsymbol{H}(\boldsymbol{x}(t))$. Since in the uncertainty version of this model the conditional expectation of $\boldsymbol{x}(t+1)$ is $\boldsymbol{H}(\boldsymbol{x}(t))$ it is clear that the only uncertainty allowed here is that which arises in the exogenous environmental conditions provided they do not influence such structural elements which are part of $\boldsymbol{H}(\cdot)$. In particular, such random processes as improvements in the state of technology, or the increase in the availability of natural resources, are not allowed. Turning now to the papers themselves, it may appear that the results under the condition of indecomposability are different from those under the condition of decomposability. In fact this is not the case, and results are essentially the same provided the symbols and the concepts are properly interpreted. Let us review this issue now and indicate below in what way the results are different for the two cases.

(a) First we note that the model of 'decomposability' as proposed here is not as general as described by the authors and the concept of 'triangular' would probably be more appropriate. This follows from the fact that $\boldsymbol{H}(\cdot)$ has a 'two sector' structure of the form $\boldsymbol{H} = (\boldsymbol{H}^1, \boldsymbol{H}^2)$ where $H_i^1 = H_i$ if $1 \leq i \leq d_1$ and $H_i^2 = H_i$ if $d_1 + 1 < i < d$ and the two functions are defined over $\boldsymbol{x} = (\boldsymbol{x}^1, \boldsymbol{x}^2)$ as follows:

$$H_i = H_i(\boldsymbol{x}^1) \text{ for } 1 \leq i \leq d_1,\ \boldsymbol{x}^1 \text{ is } d_1\text{-dimensional,}$$

$$H_i = H_i(\boldsymbol{x}^1, \boldsymbol{x}^2) \text{ for } d_1 + 1 \leq i \leq d,\ \ \boldsymbol{x}^2 \text{ is } (d - d_1)\text{-dimensional,}$$

The sector 1 is an indecomposable sector composed of goods (industries) $1 \leq i \leq d_1$ and sector 2 is an indecomposable sector composed of goods (industries) $d_1 + 1 \leq i \leq d$.

(b) The basic theory proved by the authors of the two papers is the same for both the indecomposable and the decomposable cases. This theory can be very generally stated as follows: Given certain regularity conditions on $\boldsymbol{H}(\cdot)$ and the very strong 'bounded variances' conditions on the stochastic process $\boldsymbol{x}(t)$, there exist a positive real number λ and a non-negative, non-zero vector $\boldsymbol{V}$ and a random variable λ such that

$$\text{(i)}\quad \lim_{t\to\infty} \frac{\boldsymbol{x}(t)}{\lambda^t} = g\boldsymbol{V},$$

$$\text{(ii)}\quad E\{g \mid \boldsymbol{x}(0)\} < \infty.$$

This means that asymptotically the uncertainty regarding the growth rate vanishes while the uncertainty regarding the 'level' of the growth path is present but not essential. Thus we have again the basic conclusion similar to those obtained in other cases that for any given economy the uncertainty regarding the growth path will vanish with the passage of time. These results extend the Samuelson–Solow results to the case of uncertainty and represent a nice analytical achievement by the authors.

(c) The difference between the indecomposable and the decomposable cases are to be found in the values that λ and $\boldsymbol{V}$ take. In the indecomposable case, λ and $\boldsymbol{V}$ are simply the eigenvalue and vector of $\boldsymbol{H}(\cdot)$; thus $\boldsymbol{H}(\boldsymbol{V}) = \lambda\boldsymbol{V}$ and with the regularity conditions on $\boldsymbol{H}$ we have $\lambda > 0$ and $\boldsymbol{V} \gg 0$.

In the indecomposable case we let λ_1, λ_2, $\boldsymbol{V}^1$, $\bar{\boldsymbol{V}}^2$ and $\boldsymbol{V}^2$ be defined as follows:

$$\begin{aligned}
&\text{(i)}\quad \boldsymbol{H}^1(\boldsymbol{V}^1) = \lambda_1\boldsymbol{V}^1,\\
&\text{(ii)}\quad \boldsymbol{H}^2(\boldsymbol{0}, \bar{\boldsymbol{V}}^2) = \lambda_2\bar{\boldsymbol{V}}^2,\\
&\text{(iii)}\quad \boldsymbol{H}^2(\boldsymbol{V}^1, \boldsymbol{V}^2) = \lambda_1\boldsymbol{V}^2,
\end{aligned}$$

where the existence of all these scalars and vectors is proved. It is clear that λ_1 and $\boldsymbol{W} = (\boldsymbol{V}^1, \boldsymbol{V}^2)$ are the eigenvalue and vector of $\boldsymbol{H}(\cdot)$; thus $\boldsymbol{H}(\boldsymbol{W}) = \lambda_1\boldsymbol{W}$. However, λ_1 is not necessarily the balanced growth rate and $\boldsymbol{W}$ is not necessarily the balanced proportions. The results depend upon the relations between λ_1 and λ_2. The Kesten–Stigum theory applies only to the cases where $\max[\lambda_1, \lambda_2] > 1$ and $\lambda_1 \neq \lambda_2$.

Now, if $\lambda_1 > \lambda_2$ then the results of the decomposable case are identically the same as those of the indecomposable case: λ_1 is the asymptotic growth rate and $\boldsymbol{V} = \boldsymbol{W} = (\boldsymbol{V}^1, \boldsymbol{V}^2)$ is the vector of asymptotic proportions.

If $\lambda_1 < \lambda_2$, then the interpretation is different: the asymptotic growth rate of the economy becomes λ_2 and the first sector becomes asymptotically insignificant. The theory states that the second sector is asymptotically self-sustaining with the vector of relative stock proportions of $V = (0, \bar{V}^2)$.

(d) Finally, we note that the conditions in Professor Stigum's paper are essentially the same conditions used by Professors Kesten and Stigum in their paper:

(i) the critical bounded variances conditions and the regularity conditions on $H(\cdot)$ are essentially the same;

(ii) Professor Stigum's condition requiring the conditional probability of $x(t + 1)$ to be smaller (larger) than $H(x(t))$ in case λ is smaller (larger) than 1 is replaced in the decomposable case by the conditions on the function M and the initial condition $x(0)$; the function of all these conditions in the proofs is the same; and

(iii) the concavity condition on $H(\cdot)$ assumed by Professor Stigum is slightly altered; for the proof of convergence it is assumed, in the second paper, that $|H(x) - H(y)| \leq K|x - y|$ while for the proof that $0 < E\{g \mid x(0)\} < \infty$ Professors Kesten and Stigum assume concavity as well.

Leaving additional technical matters aside, I note that the uncertainty underlying the Kesten–Stigum universe and the one operating in the Radner [6] model presented in this meeting are the same. It is the random process of the environment which generates the random consequences of economic decisions. This fact is true in spite of the fact that the Kesten–Stigum economy is supposed to describe a decentralized economy while the Radner theory deals with central planning processes.

It may be argued that for the theory of planning the only relevant uncertainty is the one regarding the environment. This is probably true for a completely centralized economy where all decisions, including investment and consumption decisions for each individual, are made in the center. Such economies are of little interest, mostly because they do not exist. If we restrict our attention to decentralized economies with price guided allocation mechanism, then it is not clear at all that the important uncertainty arises from the random nature of the environment. In fact, I suspect that in competitive economies the random nature of the environment which gives rise to 'exogenous uncertainty' is probably small compared to the all-important endogenous uncertainty which we shall discuss now.

13.3. Exogenous vs Endogenous Uncertainties

The traditional theory of equilibrium under uncertainty was developed by Arrow and Debreu. In this theory uncertainty exists with respect to 'the state of the world'. This is usually viewed as the availability of resources and the possibilities of production and consumption. In this theory a set of contingency markets are to be established in which, for each set of prices for contingency contracts, individuals will buy and sell future commodities and services contingent upon the state of the world. This is a very good description of the insurance world: the consumer pays a price today for a delivery of a new home by the insurance company if his house burns down. The traditional Arrow–Debreu model was criticized by many writers, including Radner [5], who proposed an extension of the traditional model to allow different individuals to have different information available to them. Radner's analysis leads to the conclusion that, because of such differences in information, many contingency markets may not function. This is a generalization of the phenomenon of 'moral hazard' (see ref. [1]). There are other reasons, such as transaction cost, which may prevent the achievement of a complete system of contingency markets. However, all the attempts at modifying the Arrow–Debreu view of uncertainty are only designed to allow the individual decision maker to cope with a rather incredible task of specifying his utility of consumption in every state of the world and establishing his probability distribution over all states of the world. Note that, if consumer preferences are random, then the configuration of individual preferences is part of 'the state of the world'. This means that each participant must establish his probability distribution of all participant's preferences; there must exist markets in which delivery is contingent upon the actual preferences of each individual participant, and these must be revealed after the actual selection takes place.

Radner's solution to the above is simply to assume that individuals do not and cannot know certain things, and thus they form their supply and demand correspondences under conditions of ignorance. In such circumstances it is safe to assume that no contract will be signed between agents 1 and 2 where delivery will be contingent upon the utility function of agent 3. But this obviously means that the utility function of individual 3 will remain an endogenous random variable which gives rise to risk that cannot be insured against. Also, in many instances, the cost of establishing the precise state of the world, once it has occurred, may be very high. If the marginal gain from the functioning of this contingency market is less than the cost of producing the information, the market will not function.

Thus, all in all, diversity of information, the phenomenon of 'moral hazard', transaction cost, and the cost of establishing 'the state of the world', all contribute to the failure of contingency markets to function. This in turn gives rise to risks and uncertainties which, in many cases, the individual agent can neither insure against nor purchase more.

The above are examples of what I would regard as 'endogenous uncertainty' in the sense that this uncertainty is either created within the system or is a reflection of the internal functioning of the economy. These are not the only cases of endogenous uncertainty: we turn now to discuss other forms of such uncertainty and their consequences.

13.4. Endogenous Uncertainty and its Consequences

Uncertainty is related in an essential way to the sequential nature of economic activity. For this reason individual decision makers can discover from their own experience the value of knowledge and information. The Radner consumer who is assumed to have a fixed information structure will learn two fundamental facts: first, that he may produce or purchase information, and second, that what may be regarded by all people as inconceivable may indeed happen.

The above suggests that when information may be acquired or purchased, the amount of individual uncertainty becomes an endogenous element. This means that in his private decision making the consumer may elect to go to school or to engage the informational services of specialists in order to reduce his risk. Firms may select the amount of technological uncertainty by spending resources on research, and the risk regarding the availability of resources may depend on the resources spent on exploration. This endogenous uncertainty is amplified by the fundamental non-convexity in this economic activity: on the supply side, produced information has the character of public goods. On the demand side there is a problem in defining the demand for information when one does not know the information one buys (see for example, ref. [1, chapter 12]).

On the other extreme, the Radner agent who operates under conditions of ignorance may discover that his basic probability space was 'wrong' in the sense that certain events which he did not conceive of did happen; perhaps the most spectacular events are precisely those which very few people had conceived of. This is inconsistent with the behavior of Radner's agent since, according to Radner, the basic probability space is known to each consumer and he has a well defined utility which depends upon each set of states of the world. This may lead an individual to assign subjective probability to all those unknown events that may happen but whose nature is unknown. If this is so, the typical individual will have a probability measure of less than 1 on the space of all events which he may conceive of. Individual reinforcement to this view arises from the occurrence of events which 'nobody conceived of'.

We finally arrive at the most fundamental endogenous uncertainty which is associated with any economic activity: the uncertainty regarding the capacity, managerial skills and qualities of other agents.

Owners of capital know that most of their capital is managed by other individuals, and primary among the sources of the return to their capital are the managerial skills and success of those who use the capital. Workers who sell their services to an employer will have a basic uncertainty regarding the nature, quality and duration of their employment which will depend upon the abilities of the employer. Further, in the provision of any human service, from the repair of your car to a medical diagnosis, there is a fundamental uncertainty due to the fundamental random nature of human decisions. And finally, in signing any contract there is always the uncertainty regarding the desire and ability of agents to deliver.

The essence of the above examples is that individual capacity for optimal action entails qualities which are basic random variables. The random variations in labor and managerial performance is a fundamental endogenous uncertainty which is probably the most important uncertainty in any social system. Note that some of the above uncertainties do give rise to some kind of 'contingency' markets: unemployment insurance and insurance of the functioning of private durable goods are examples[1]. On the other hand there is no insurance against a fall in price of a commodity caused by excessive capacity in an industry. Here a lag in market adjustment – common in almost all industries – may lead to excessive rate of entry or exit. This is a typical random variable resulting from the nature of production, the organization of the markets and the ability of management. It is obvious that the rate of entry and exit is a very important uninsurable random variable to any profit maximizing firm.

The fundamental consequence of the existence of endogenous uncertainty is the same as the consequences of any source of uncertainty for which there are no contingency markets functioning. These consequences are simple: individuals must bear their own uncertainty and they cannot capitalize their future endowments, assets or obligations. On the other hand, the markets will reopen in the next period and there are commodities, services or contracts that will be traded during the next market date and for which prices will be available in the next market date but for which no futures markets and prices are available today. The above leads to the simple observation that, in spite of the fact that many futures markets are not functioning today and many futures prices are not available today, the consumer forms expectations about prices tomorrow and those expectations become an important basis for decision making today! It is then clear that from the viewpoint of the individual consumer functioning in *today's* market, *prices tomorrow are random variables.* Apart from the consumer's ability to forecast prices tomorrow, it is perfectly reasonable for him to make contracts *which are contingent on market prices tomorrow*! Thus a bakery can make contracts for future delivery of bread contingent on the prices of flour. In this case the bakery can also purchase contracts for the delivery of wheat at a later date and thus make a firm sale of future bread. However, a travel agent might make a touring contract with his clients contingent upon the airline and hotel prices remaining unaltered. Most contracts which are made contingent upon future market prices are usually made contingent upon 'prices not changing' and then certain rules for renegotiation in case changes do take place. Thus all contracts with escalation clauses related to inflation are in fact examples of such contracts. Also all options, put and call contracts can be interpreted as

---

[1] See below for further discussion.

contracts to buy or sell something in the *future* contingent upon the market price prevailing at that time. Thus consider my purchasing a six-month option to buy a certain asset at a fixed price p_0. This contract is *equivalent* to a contract to deliver the same asset at the future date, delivery being contingent upon $p > p_0$. If $p < p_0$ then the buyer does not want the delivery to take place. In general it appears that contracts for future delivery contingent upon future prices are much more common than is generally assumed.

13.5. Uncertainty and Expectations: a Search for a New Equilibrium Concept

One simple fact must be stated at this point: given the fact that the Arrow–Debreu contingency markets deal with a relatively small amount of uncertainty, the bulk of social uncertainty is left to be dealt with in a different manner. There are first of all those risks which must be born by the individual consumer since neither contingency markets nor their social substitutes can function. This leads to a fundamental endogenous uncertainty. Then there are all the social substitutes for contingency markets which are markets in which uncertainty is defined with respect to events, i.e. set of states. Thus there may not exist insurance against the detailed configuration of other agents' conduct, but there exists insurance against malpractice suits; there may not exist contingency contracts which may specify the possible events that may happen to a firm and its employees, but there exists an insurance against unemployment. These are examples of the diverse methods of transferring risk through markets which are not contingency markets in the Arrow–Debreu sense; risk is defined in these markets in terms of events.

The most important consequence of the failure of contingency markets to function is the translation of a great deal of this form of untraded risk into uncertainty about prices which will prevail in future markets. This is clearly not included in the notion of uncertainty regarding the 'state of the world'. However, from the point of view of the individual agent operating in a world of incomplete markets, this is a real endogenous risk and gives rise to markets in which this uncertainty is traded. This trading follows from the fact that the presence of this uncertainty creates expectations and diversity of expectations, endowments and tastes create an extensive market for trade in this uncertainty. This is one way of looking at securities markets: in this context an investing individual who purchases a position in a mutual fund is purchasing essentially the same service of reducing his risk as the insurance policy which insures his home.

The extensive market for financial intermediaries contains important segments in which the risk of future variations in prices is traded. Moreover, given the difficulties which we have enumerated with regard to the 'state of the world' model of risk *it is probably the simplest and most efficient procedure to allow individuals to trade their expectations of future market prices.* We are thus led to the inevitable rejection of the Arrow–Debreu model of equilibrium under uncertainty. A proper theory of equilibrium under uncertainty will have to be developed in the context of an economy:

(1) with a sequence of markets;

(2) where individuals are allowed to make contracts contingent upon prices tomorrow; and

(3) where individual price expectations are formed endogenously and influence the allocation today.

It is important to note that in principle what is emphasized here is that we should extend the notion of the 'state of the world' to include future prices. Since Hicks, it has been well understood that a theory of plans and expectations must be integrated into the theory of general equilibrium. Recent work in the theory of temporary equilibrium has incorporated some of the ideas presented here (see for example, refs. [2] and [3]). In all of these contributions it has been assumed that there exists a *fixed* expectational function which assigns to any vector $\boldsymbol{p}$ of today's prices a probability distribution of tomorrow's prices. With this device it is then possible to define equilibrium relative to the fixed expectation function. This is most unsatisfactory, since there are reasons why expectations are formed, and some of these reasons were discussed above; to assume a fixed expectation function does not resolve any issue at all. Since it is clearly understood that the structure of expectations today can have dramatic effects on the allocation today, the assumption of an arbitrary individual expectation function is almost equivalent to the statement that any allocation may be an equilibrium. The fundamental issue is the endogenous formation of expectations. This is an open question.

References

[1] Arrow, K. J. *Essays in the Theory of Risk-Bearing.* Markham Publishing Company: Chicago (1971).
[2] Grandmont, J. M. *On the short-run equilibrium in a monetary economy.* CEPREMAP Discussion Paper, Paris (February 1971).
[3] Green, J. *Temporary general equilibrium in a sequential trading model with spto and futures transactions.* CORE Discussion Paper No. 7127, Heverlee, Belgium (August 1971).
[4] Radner, R. Competitive equilibrium under uncertainty. *Econometrica,* **36** (1968), 31–58.
[5] Radner, R. Problems in the theory of markets under uncertainty. *American Economic Review,* **61** (1970), 454–460.
[6] Radner, R. *Optimal stationary consumption with stochastic production and resources.* Technical Report No. 14, Center for Research in Management Science, University of California, Berkeley, California (February 1972).
[7] Samuelson, P. A. and Solow, R. M. A complete capital model involving heterogeneous capital goods. *Quarterly Journal of Economics,* **70** (1956),
[8] Stigum, B. Balanced growth under uncertainty. *Journal of Economic Theory,* **5** (1972), 42–68.

BIBLIOGRAPHY

Afriat, S. N. The validity of the expected utility hypothesis. In *Recent Advances in Game Theory*. Princeton Univ. Press: Princeton, N.J. (1962), 73–82.

Akerlof, G. A. The market for lemons: Quantitative uncertainty and the market mechanism. *Quart. Jour. Econ.*, **84** (1970), 488–500.

Alchian, A. A. Uncertainty, evolution and economic theory. *Jour. Pol. Econ.*, **58** (1950), 211–221.

Allais, M. Le comportement de l'homme rationnel devant le risque: critique des postulats et axioms de l'école américaine. *Econometrica*, **21** (1953), 503–546.

Anscombe, F. J. and Aumann, R. J. A definition of subjective probability. *Annals Math. Stat.*, **34** (1963), 199–205.

Archibald, G. C. Uncertainty and the utility function. *Econ. Jour.*, **58** (1948), 1–10.

Archibald, G. C. Utility, risk, and linearity. *Jour. Pol. Econ.*, **67** (1959), 437–450.

Arditti, F. D. Risk and required rate of return. *Jour. Finance*, **22** (1967), 19–36.

Arditti, F. D. Risk and required rate of return: reply. *Jour. Finance*, **24** (1969), 720.

Armstrong, W. E. Uncertainty and the utility function. *Econ. Jour.*, **58** (1948), 1–10.

Arrow, K. J. Alternative approaches to the theory of choice in risk-taking situations. *Econometrica*, **19** (1951), 404–437.

Arrow, K. J. Utilities, attitudes, choices: A review note. *Econometrica*, **26** (1958), 1–23.

Arrow, K. J. Functions of a theory of behavior under uncertainty. *Metroeconomica*, **11** (1959), 12–20.

Arrow, K. J. Rational choice functions and orderings. *Economica*, N.S., **26** (1959), 121–127.

Arrow, K. J. Uncertainty and the welfare economics of medical care. *Am. Econ. Rev.*, **53** (1963), 941–973.

Arrow, K. J. Comment on the portfolio approach to the demand for money and other assets. *Rev. Econ. Stat.*, **45** (1963), 24–27.

Arrow, K. J. The role of securities in the optimal allocation of riskbearing. *Rev. Econ. Stud.*, **31** (1964), 91–96.

Arrow, K. J. Exposition of the theory of choice under uncertainty. *Synthese*, **16** (1966), 253–269.

Arrow, K. J. *Aspects of the Theory of Risk-Bearing* (Yrjo Jahnsson Lectures). Yrjo Jahnssonin Saatio: Helsinki (1965).

Arrow, K. J. The economics of moral hazard: Further. comment. *Am. Econ. Rev.*, **58** (1968), 537–539.

Arrow, K. J. *Essays in the Theory of Risk Bearing*. Markham Pub. Co.: Chicago, Illinois (1971).

Arrow, K. J., Hurwicz, L. and Uzawa, H. (eds.) *Studies in Linear and Non-linear Programming*. Stanford Univ. Press: Stanford, Calif. (1958).

Arrow, K. J., Karlin, S. and Scarf, H. *Studies in the Mathematical Theory of Inventory and Production.* Stanford Univ. Press: Stanford (1958).

Arrow, K. J. and Lind, R. C. Uncertainty and the evaluation of public investment decisions. *Am. Econ. Rev.*, **60** (1970), 364–378.

Arrow, K. J. and Lind, R. C. Uncertainty and the evaluation of public investment decisions: Reply. *Am. Econ. Rev.*, **62** (1972), 171–172.

Audley, R. J. A stochastic model for individual choice behavior. *Psychological Rev.*, **67** (1960), 1–15.

Aumann, R. J. *Subjectively mixed strategies.* Technical Report No. 45, Institute for Mathematical Studies in the Social Sciences, Stanford Univ. (1971).

Bailey, M. Formal criteria for investment decisions. *Jour. Pol. Econ.*, **67** (1959), 476–488.

Bailey, M. and Jensen, M. C. Risk and the discount rate for public investment. In *Studies in the Theory of Capital Markets* (ed. Jensen, M. C.). Praeger Publishers: New York (1972).

Baron, D. Price uncertainty, utility, and industry equilibrium in pure competition. *Intl. Econ. Rev.*, **11** (1970), 463–480.

Baron, D. Demand uncertainty in imperfect competition. *Intl. Econ. Rev.*, **12** (1971), 196–208.

Baumol, W. J. The Neumann–Morgenstern utility index: An ordinalist view. *Jour. Pol. Econ.*, **59** (1951), 61–66.

Baumol, W. J. An expected gain-confidence limit criterion for portfolio selection. *Management Science*, **10** (1963), 174–183.

Baumol, W. J. and Quandt, R. E. Rules of thumb and optimally imperfect decisions. *Am. Econ. Rev.*, **54** (1964), 23–46.

Bayes, T. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions*, **53** (1963), 376-398.

Becker, G. M. Objective measures of subjective probability and utility. *Psychological Review*, **69** (1962), 136–148.

Becker, G. M., Degroot, M. H. and Marschak, J. Probabilities of choices among very similar objects: An experiment to decide between two models. *Behavioral Science*, **8** (1963), 306-311.

Becker, G. M., Degroot, M. H. and Marschak, J. Stochastic models of choice behavior. *Behavioral Science*, **8** (1963), 41–55.

Becker, G. M. and McClintock, C. G. Value: Behavioral decision theory, *Annual Review of Psychology*, **18** (1967), 239–286.

Becker, G. S. Irrational behavior and economic theory. *Jour. Pol. Econ.*, **70** (1962), 1–13.

Becker, S. W. and Siegel, S. Utility of grades: Level of aspiration in a decision theory context. *Jour. Experimental Psychology*, **55** (1958), 81–85.

Beja, A. The structure of the cost of capital under uncertainty. *Rev. Econ. Stud.*, **38** (1971), 359–369.

Ben-Zion, U. and Balch, M. On the analysis of corporate financial theory under uncertainty: A comment. *Quart. Jour. Econ.*, **87** (1973), 290–295.

Bernhardt, I. 'Fixed costs' and the competitive firm under price uncertainty: Comment. *Am. Econ. Rev.*, **62** (1972), 193.

Bernoulli, D. Exposition of a new theory on the measurement of risk. *Econometrica*, **22** (1954), 23–36.

Bhattacharya, R. N. and Majumdar, M. Random exchange economies. *Jour. Econ. Theory*, **6** (1973), 37–67.

Bierwag, G. O. and Grove, M. A. Indifference curve in asset analysis. *Econ. Jour.*, **76** (1966), 337–343.

Billingsley, P. *Convergence of Probability Measures*. John Wiley: New York (1968).

Blackwell, D. and Girshick, M. A. *Theory of Games and Statistical Decisions*. John Wiley: New York (1954).

Block, M. K. and Heineke, J. M. A comment on uncertainty and household decisions. *Rev. Econ. Stud.*, **39** (1972), 523–525.

Block, M. K. and Heineke, J. M. The allocation of effort under uncertainty: The case of risk-averse behavior. *Jour. Pol. Econ.*, **81** (1973), 376–385.

Blume, M. E. Portfolio theory: A step toward its practical application. *Jour. Bus.*, **43** (1970), 152–173.

Bohman, G. A mutual reinsurance scheme. *The ASTIN Bulletin*, **6** (1971).

Bolker, E. D. A simultaneous axiomatization of utility and subjective probability. *Philosophy of Science*, **34** (1967), 333–340.

Borch, K. H. Reciprocal reinsurance treaties. *The ASTIN Bulletin*, **1** (1960).

Borch, K. H. Reciprocal reinsurance treaties seen as a two person cooperative games. *Skandinavaisk Aktuarietidshift* (1960).

Borch, K. H. Equilibrium in a reinsurance market. *Econometrica*, **30** (1962), 424–444.

Borch, K. H. A note on utility and attitudes to risk. *Management Science*, **9** (1963), 697–700.

Borch, K. H. Recent developments in economic theory and their application to insurance. *The ASTIN Bulletin*, **2** (1963), 322–341.

Borch, K. H. The economics of uncertainty. In *Essays in Mathematical Economics in Honor of Oscar Morgenstern* (ed. Shubik, M.). Princeton Univ. Press: Princeton, N. J. (1967), 197–210.

Borch, K. H. Decision rules on the probability of ruin. *Oxford Economic Papers*, **20** (1968), 1–10.

Borch, K. H. *The Economics of Uncertainty*. Princeton University Press: Princeton, N. J. (1968).

Borch, K. H. The theory of control processes applied to insurance companies. *Jour. Risk Insurance*, **34** (1967), 581–591.

Borch, K. H. The theory of risk. *Jour. Royal Stat. Society*, Ser. B, **29** (1967), 432–452.

Borch, K. H. Indifference curves and uncertainty. *Swedish Jour. Econ.*, **70** (1968), 19–24.

Borch, K. H. The optimal portfolio of assets in an insurance company: transactions of the 18th International Congress of Actuaries, Munich (1968), 21–31.

Borch, K. H. A note on uncertainty and indifference curves. *Rev. Econ. Stud.*, **36** (1969), 1–4.

Borch, K. H. Specification of objectives in decision problems. *Theory and Decision*, **1** (1971), 5–21.

Boussard, J. M. The introduction of risk into a programming model: Different criteria and the actual behavior of farmers. *European Econ. Rev.*, **1** (1969–70), 92–121.

Bowman, M. J. (ed.) *Expectations, uncertainty and business behavior*. Social Science Research Council, Committee on Business Enterprise Research: New York (1958).

Brainard, W. Uncertainty and the effectiveness of policy. *Am. Econ. Rev.*, Proceedings, **57** (1967), 411.

Breen, W. Specific *versus* general models of portfolio selection. *Oxford Econ. Papers*, **20** (1968), 361–368.

Breen, W. and Savage, J. Portfolio distributions and tests of security selection models. *Jour. Finance*, **23** (1968), 805–819.

Breiman, L. Optimal gambling systems for favorable games. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Univ. Calif. at Berkeley Press: Berkeley (1960).

Brennan, M. J. Capital market equilibrium with divergent borrowing and lending rates. *Jour. Fin. Quant. Anal.*, **6** (1971), 1197–1205.

Brock, W. A. On models of expectations that arise from maximizing behavior of economic agents over time. *Jour. Econ. Theory.*, **5** (1972), 348–376.

Brock, W. A. and Mirman, L. J. Optimal economic growth and uncertainty: The no discounting case, *Int. Econ. Rev.*, **74** (October 1973), 560–573.

Brock, W. A. and Mirman, L. J. *The stochastic modified golden rule in a one sector model of economic growth with uncertain technology*. Mimeographed, Rochester and Cornell (1970).

Brock, W. A. and Mirman, L. J. *Optimal economic growth and uncertainty: the Ramsey–Weizsacker case*. Working Paper No. 7, Dept. of Economics, Univ. of Calif. at Berkeley (1971).

Brockie, M. D. and Grey, A. L., Jr. The marginal efficiency of capital and investment programming. *Econ. Jour.*, **66** (1956), 662–676.

Budd, A. P. and Litzenberger, R. H. The market price of risk, size of market and investor's risk aversion: A comment. *Rev. Econ. Stat.*, **54** (1972), 204–206.

Buehler, R. J. Measuring information and uncertainty. In *Proceedings of the Symposium on Foundations of Statistical Inference*. Holt, Rinehart, and Winston: Toronto (1971).

Bush, R. R., Mosteller, F. and Thompson, G. L. A formal structure for multiple-choice situations. In *Decision Processes* (eds. Thrall, R. M., Coombs, C. H. and Davis, R. L.). John Wiley: New York (1954), 92–126.

Canada, J. R. The consideration of risk and uncertainty in capital investment analysis. *Management Intl. Rev.*, **7** (1967), 47–57.

Cass, D. and Stiglitz, J. E. The structure of investor preference and asset returns, and separability in portfolio allocation: A contribution to the pure theory of mutual funds. *Jour. Econ. Theory*, **2** (1970), 122–160.

Cass, D. and Stiglitz, J. E. Risk aversion and wealth effects on portfolios with many assets. *Rev. Econ. Stud.*, **39** (1972), 331–354.

Chacko, G. K. *Applied Statistics in Decision-Making*. American Elsevier Pub. Co.: New York (1971).

Champernowne, D. G. *Uncertainty and Estimation in Economics*. Holden Day: San Francisco, Calif. (1969).

Chance, W. A. *Statistical Methods for Decision Making*. Irwin: Homewood, Illinois (1969 and 1970).

Chang, J. C. and Holt, C. C. Optimal investment orders under uncertainty and dynamic costs: Theory and estimates. *South. Econ. Jour.* **39** (1973), 508–525.

Chase, S. B., Jr. Credit risk and credit rationing: Comment. *Quart. Jour. Econ.*, **75** (1961), 319–327.

Chernoff, H. Rational selection of decision functions. *Econometrica*, **22** (1954), 423–443.

Chernoff, H. and Moses, L. E. *Elementary Decision Theory*. John Wiley: New York (1959).

Chipman, J. S. Stochastic choice and subjective probability. In *Decisions, Values and Groups* (ed. Willner, D.). Vol. I. Pergamon Press: New York (1960), 70–95.

Chipman, J. S. The ordering of portfolios in terms of mean and variance. *Rev. Econ. Stud.*, **40**, No. 122 (April 1973), 167–190.

Churchman, C. W. Decision and value theory. In *Progress in Operation Research* (ed. Ackoff, R. L.), Vol. I. John Wiley: New York (1961), 35–64.

Churchman, C. W. *Prediction and Optimal Decision: philosophical issues of a science of values.* Prentice-Hall: Englewood Cliffs, N. J. (1961).

Cohen, J. and Hansel, C. E. M. The nature of decisions in gambling. *Acta Psychologica* (Amsterdam), **13** (1958), 357–370.

Collins, R. A nonproportional (spread loss) reinsurance pool. *Transactions,* 19th International Congress of Actuaries (1972).

Coombs, C. H. A study of decision making under risk. In *La Décision.* Editions du Centre Nationale de la Recherche Scientifique: Paris (1961), 63–72.

Coombs, C. H. and Beardslee, D. On decision-making under uncertainty. In *Decision Processes* (eds. Coombs, C. H., Davis, R. L. and Thrall, R. M.). John Wiley: New York (1954), 255–285.

Coombs, C. H., Bezembinder, T. G. G. and Goode, F. M. Testing expectation theories of decision making without measuring utility or subjective probability. *Jour. Math. Psychology,* **4** (1967), 72–103.

Coombs, C. H. and Komorita, S. S. Measuring utility of money through decisions. *Am. Jour. Psychology,* **71** (1958), 383–389.

Cootner, P. H. Stock prices: Random *vs.* systematic changes. *Industrial Management Rev.,* **3** (1962), 25–45.

Cootner, P. H. *The Random Character of Stock Market Prices.* MIT Press: Cambridge, Mass. (1964).

Cramer, H. On the mathematical theory of risk. In *Forsakringsaktiebolaget Skandia 1855–1930,* Vol. II. Centraltrycheriet: Stockholm (1930), 7–84.

Crew, M. Coinsurance and the welfare economics of medical care. *Am. Econ. Rev.,* **59** (1969), 906–908.

Currie, K. A. and Wu, S. Y. *Rate-of-return regulation with valuation and demand uncertainty.* Bureau of Business and Economic Research, Working Paper Series, No. 73–5, Univ. of Iowa (1973).

Darling, P. G. Surrogative measurement of expectations: An example in estimating the liquidity influence on investment. *Rev. Econ. Stat.,* **38** (1956), 413–426.

Davidson, D. and Marshak, J. Experimental tests of a stochastic decision theory. In *Measurement: Definitions and Theories* (eds. Churchman, C. W. and Ratoosh, P.). Wiley: New York (1959), 233–269.

Davidson, D. and Suppes, P. A finitistic axiomatization of subjective probability and utility. *Econometrica,* **24** (1956), 264–275.

Davidson, D., Suppes, P. and Siegel, S. *Decision Making: An Experimental Approach.* Stanford Univ. Press: Stanford (1957).

Davis, O. and Whinston, A. Externalities, welfare, and the theory of games. *Jour. Pol. Econ.,* **70** (1962), 241–262.

Day, R. H. Rational choice and economic behavior. *Theory and Decision,* **1** (1971), 229–251.

Day, R. H. and Kennedy, P. Recursive decision systems: An existence analysis. *Econometrica,* **38** (1970), 666–681.

De Ghellinck, G. T. and Eppen, G. D. Linear programming solutions for separable markovian decision problems. *Management Science,* **13** (1967), 371–394.

Debreu, G. Stochastic choice and cardinal utility. *Econometrica,* **26** (1958), 440–444.

Debreu, G. Cardinal utility of even-chance mixtures of pairs of sure prospects. *Rev. Econ. Stud.,* **71** (1959), 174–177.

Debreu, G. *Theory of Value: an axiomatic analysis of economic equilibrium.* John Wiley: New York (1959).

Debreu, G. Topological methods in cardinal utility theory. In *Mathematical Methods in the Social Sciences* (eds. Arrow, K. J., Karlin, S. and Suppes, P.). Stanford Univ. Press: Stanford (1960), 16–26.

Debreu, G. Preference functions on measure spaces of economic agents. *Econometrica,* **35** (1967), 111–122.

de Finetti, B. La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré,* **7** (1937), 1–68. English translation in Kyburg and Smokler, 1964.

Dhrymes, P. J. On the theory of the monopolistic multiproduct firm under uncertainty. *Intl. Econ. Rev.,* **5** (1964), 239–257.

Diamond, P. A. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility: comment. *Jour. Pol. Econ.,* **75** (1967), 765–766.

Diamond, P. A. The role of a stock market in a general equilibrium model with technological uncertainty. *Am. Econ. Rev.,* **57** (1967), 759–776.

Diamond, P. A. and Mirrlees, J. A. Optimum taxation and public production: I – production efficiency. *Am. Econ. Rev.,* **61** (1971), 8–27.

Diamond, P. A. and Mirrlees, J. A. Optimum taxation and public production: II – tax rules. *Am. Econ. Rev.,* **61** (1971), 261–278.

Diamond, P. A. and Stiglitz, J. E. *Increases in risk and risk aversion.* MIT, Dept. of Econ. Working Paper, No. 97 (Jan. 1973).

Diamond, P. A. and Yaari, M. Implications of the theory of rationing for consumer choice under uncertainty. *Am. Econ. Rev.,* **62** (1972), 333–343.

Dixit, A. K. and Goldman, S. M. Uncertainty and demand for liquid assets. *Jour. Econ. Theory,* **2** (1970), 368–382.

Domar, E. and Musgrave, R. Proportional income taxation and risk-taking. *Quart. Jour. Econ.,* **58** (1944), 388–422.

Doob, J. L. *Stochastic Processes.* John Wiley: New York (1953).

Douglas, G. W. Risk in the equity markets: An empirical appraisal of market efficiency. *Yale Econ. Essays,* **9** (1969), 3–45.

Dreze, J. H. *Individual decision making under partially controllable uncertainty.* PhD dissertation, Columbia Univ. (1958).

Dreze, J. H. Market allocation under uncertainty. *European Econ. Rev.,* **2** (1970–71), 133–165.

Dreze, J. H. Econometrics and decision theory. *Econometrica,* **40** (1972), 1–17.

Dreze, J. and Gabsewicz, J. *Demand fluctuations, capacity utilization, and prices.* CORE Dis. Paper 6607, Louvain (1967).

Dreze, J. H. and Modigliani, R. Consumption decisions under uncertainty. *Jour. Econ. Theory,* **5** (1972), 308–335.

Dubins, L. E. and Savage, L. J. *How to Gamble if You Must: Inequalities for Stochastic Processes.* McGraw-Hill: New York (1965).

Dunford, N. and Schwartz, J. T. *Linear Operators.* Interscience: New York (1964).

Durand, D. The cost of capital, corporation finance and the theory of investment: A comment. *Am. Econ. Rev.,* **49** (1959), 639–654.

Dusak, K. *Futures trading and investor returns: An investigation of commodity market risk premiums.* PhD Dissertation, Univ. Chicago (1971).

Edwards, W. The theory of decision making. *Psychological Bulletin,* **51** (1954), 380–417.

Edwards, W. Behavioral decision theory. In *Annual Review of Psychology* (eds. Farnsworth, P. E., McNemar, O. and McNemar, Q.). Annual Reviews, Inc.: Palo Alto, Calif. (1961), 473–498.

Edwards, W. Measurement of utility and subjective probability. In *Psychological Scaling: Theory and Applications* (eds. Bulliksen, H. and Messick S.). John Wiley: New York (1960), 109–127.

Edwards, W. Subjective probabilities inferred from decisions. *Psychological Rev.*, **69** (1962), 109–135.

Edwards, W. Utility, subjective probability, their interaction and variance preferences. *Jour. Conflict Resolution*, **6** (1962), 42–51.

Ehrlich, I. and Becker, G. S. Market insurance, self-insurance, and self-protection. *Jour. Pol. Econ.*, **80** (1972), 623–648.

Eisner, R. and Strotz, R. Flight insurance and the theory of choice, *Jour. Pol. Econ.*, **69** (1961), 355–68.

Eisner, R. and Strotz, R. The determinants of business investment. In *Impacts of Monetary Policy*. Prentice-Hall: Englewood Cliffs, N.J. (1963).

Ekern, S., Taxation, political risk and portfolio selection. *Economica*, N.S., **38** (1971), 421–430.

El Agizy, M. Two-stage programming under uncertainty with discrete distribution function. *Oper. Res.*, **15** (1967), 55–70.

Ellsberg, D. Risk, ambiguity, and the Savage axioms. *Quart. Jour. Econ.*, **75** (1961), 643–669.

Encarnacion, J. On decision under uncertainty. *Econ. Jour.*, **75** (1965), 442–444.

Encarnacion, J. On independence postulates concerning choice. *Intl. Econ. Rev.*, **10** (1969), 134–140.

Eppen, G. D., and Fama, E. F. Solution for cash balance and simple dynamic portfolio problems. *Jour. Bus.*, **14** (1968), 94–112.

Eppen, G. D. and Fama, E. F. Cash balance and simple dynamic portfolio problems with proportional costs. *Intl. Econ. Rev.*, **10** (1969), 119–133.

Evans, J. L. and Archer, S. H. Diversification and the reduction of dispersion: An empirical analysis. *Jour. Finance*, **23** (1968), 761–767.

Fama, E. F. The behavior of stock market prices. *Jour. Bus.*, **38** (1965), 34–105.

Fama, E. F. Portfolio analysis in a stable Paretian market. *Management Science*, **11** (1965), 404–419.

Fama, E. F. Risk, return and equilibrium: Some clarifying comments. *Jour. Finance*, **23** (1968), 29–40.

Fama, E. F. Efficient capital markets: A review of theory and empirical work. *Jour. Finance*, **25** (1970), 383–417.

Fama, E. F. Multiperiod consumption–investment decisions. *Am. Econ. Rev.*, **60** (1970), 163–174.

Fama, E. F. Risk, return, and equilibrium. *Jour. Pol. Econ.*, **79** (1971), 30–55.

Fama, E. F. Perfect competition and optimal production decisions under uncertainty. *Bell Jour. Econ. Management Science*, **3** (1972), 509–530.

Fama, E. F. and Roll, R. Some properties of symmetric stable distributions. *Jour. Am. Stat. Assn.*, **63** (1968), 817–836.

Fama, E. F., Fisher, L., Jensen, M. C. and Roll, R. The adjustment of stock prices to new information. *Intl. Econ. Rev.*, **10** (1969), 1–21.

Feather, N. T. Success probability and choice behavior. *Jour. Experimental Psychology*, **58** (1959), 150–164.

Feeney, G. J., McGlothlin, W. H. and Wolfson, R. J. *Risk aversion in incentive contracting: An experiment.* The RAND Corporation, RM–4231–PR (1964).

Feldstein, M. S. Uncertainty and forward speculation. *Rev. Econ. Stat.*, **50** (1968), 182–192.

Feldstein, M. S. Mean-variance analysis in the theory of liquidity preference and portfolio selection. *Rev. Econ. Stud.*, **36** (1969), 5–12.

Fellner, W. Distortion of subjective probabilities as a reaction to uncertainty. *Quart. Jour. Econ.*, **75** (1961), 670–689.

Ferguson, T. S. *Mathematical Statistics: a Decision Theoretic Approach.* Academic Press: New York (1967).

Fischer, S. A life cycle model of life insurance purchases, *Intl. Econ. Rev.*, **14** (1973), 132–152.

Fishburn P. C. *Decision and Value Theory.* John Wiley: New York (1964).

Fishburn, P. C. Independence in utility theory with whole produce sets. *Operations Research,* **13** (1965), 28–45.

Fishburn, P. C. Markovian dependence in utility theory with whole product sets. *Operations Research,* **13** (1965), 238–257.

Fishburn, P. C. Bounded expected utility. *Annals Math. Stat.*, **38** (1967), 1054–1060.

Fishburn, P. C. Preference-based definitions of subjective probability. *Annals Math. Stat.*, **38** (1967), 1605–1617.

Fishburn, P. C. Semiorders and risky choices. *Jour. Math. Psychology,* **5** (1968), 358–361.

Fishburn, P. C. A general theory of subjective probabilities and expected utilities. *Annals Math. Stat.*, **40** (1969), 1419–1429.

Fishburn, P. C. Intransitive indifference in preference theory: A survey. *Operations Research,* **18** (1970), 207–228.

Fishburn, P. C. Intransitive individual indifference and transitive majorities. *Econometrica,* **38** (1970), 482–489.

Fishburn, P. C. *Utility Theory for Decision Making.* John Wiley: New York (1970).

Fishburn, P. C. The theory of representative majority decision. *Econometrica,* **39** (1971), 273–284.

Fishburn, P. C. Lotteries and social choices. *Jour. Econ. Theory,* **5** (1972), 189–207.

Fisher, I. *The Theory of Interest*, revised edition. Augustus M. Kelley: New York (1965).

Fisher, I. N. and Hall, G. R. Risk and corporate rates of return. *Quart. Jour. Econ.*, **83** (1969), 79–92.

Fisher, W. Estimation in the linear decision model. *Intl. Econ. Rev.*, **3** (1962), 1–29.

Freimer, M. and Gordon, M. J. Investment behavior with utility a concave function of wealth. In *Risk and Uncertainty* (eds. Borch, K. and Mossin, J.). St. Martin's Press: New York (1968).

Fried, J. Forecasting and probability distribution for models of portfolio selection. *Jour. Finance,* **25** (1970), 539–554.

Friedman, M. Choice, chance, and the personal distribution of income. *Jour. Pol. Econ.*, **61** (1953), 277–290.

Friedman, M. and Savage, L. J. The utility analysis of choices involving risk. *Jour. Pol. Econ.*, **56** (1948), 279–304.
Reprinted in *Readings in Price Theory* (eds. Stigler, G. J. and Boulding, K. E.). Irwin: Homewood, Illinois (1952), 57–96.

Friend, I. and Blume, M. Measurement of portfolio performance under uncertainty. *Am. Econ. Rev.*, **60** (1970), 561–575.

Friend, I. and Kravis, I. B. Entrepreneurial income, saving, and investment. *Am. Econ. Rev.*, **47** (1957), 269–301.

Friend, I. and Vickers, D. Portfolio selection and investment performance. *Jour. Finance,* **20** (1965), 391–415.

Georgescu-Roegen, N. Choice, expectations, and measurability. *Quart. Jour. Econ.*, **68** (1954), 503–534. Reprinted in Georgescu-Roegen, N. *Analytical Economics* (1966), 184–215.

Georgescu-Roegen, N. The nature of expectation and uncertainty. In *Expectations, Uncertainty, and Business Behavior* (ed. Bowman, M. J.). Social Science Research Council: New York (1958). Reprinted in Georgescu-Roegen, N., *Analytical Economics* (1966), 241–275.

Georgescu-Roegen, N. The theory of choice and the constancy of economic laws. *Quart. Jour. Econ.*, **64** (1950), 125–138. Reprinted in Georgescu-Roegen, *Analytical Economics*. Harvard University Press: Cambridge (1966), 171–183.

Glustoff, E. and Nigro, N. Liquidity preference and risk aversion with an exponential utility function. *Rev. Econ. Stud.*, **39** (1972), 113–115.

Good, I. J. Rational decisions. *Jour. Royal Stat. Society*, **14** (1952), 107–114.

Gordon, M. J., Paradis, G. E. and Rorke, C. H. Experimental evidence on alternative portfolio decision rules. *Am. Econ. Rev.*, **62** (1972), 107–118.

Gorman, W. M. The structure of utility functions. *Rev. Econ. Stud.*, **35** (1968), 367–390.

Gould, J. P. The expected utility hypothesis and the selection of optimal deductibles for a given insurance policy. *Jour. Bus.*, **42** (1969), 143–151.

Grandmont, J. M. *Continuity properties of a von Neumann–Morgenstern utility*. Working Paper 301, Center for Research in Management Science, Univ. of Calif. at Berkeley (1970).

Green, H. A. J. Uncertainty and the 'expectations hypothesis'. *Rev. Econ. Stud.*, **34** (1967), 387–398.

Griswold, B. J. and Luce, R. D. Choices among uncertain outcomes: A test of decomposition and two assumptions of transitivity. *Am. Jour. Psychology*, **75** (1962), 35–44.

Gronhaug, K. Risk indicators, perceived risk and consumer's choice of information sources. *Swedish Jour. Econ.*, **74** (1972), 246–262.

Groves, T. *The allocation of resources under uncertainty in a team*. Working Paper No. 270, Center for Research in Management Science, Univ. of Calif. at Berkeley (1969).

Groves, T. *The team under varying degrees of uncertainty*. Working Paper No. 269, Center for Research in Management Science, Univ. of Calif. at Berkeley (1969).

Groves, T. and Radner, R. Allocation of resources in a team. *Jour. Econ. Theory*, **4** (1972), 415–441.

Grunfeld, Y. The determinants of corporate investment. In *The Demand for Durable Goods* (ed. Harberger, A. C.). Univ. Chicago Press: Chicago (1960).

Hadar, J. and Russell, W. R. Rules for ordering uncertain prospects, *Am. Econ. Rev.*, **59** (1969), 25–34.

Hadar, J. and Russell, W. R. Stochastic dominance and diversification. *Jour. Econ. Theory*, **3** (1971), 288–305.

Hagen, K. P. Taxation and investment behavior under uncertainty: A multiperiod portfolio analysis. *Theory and Decision*, **1** (1971), 269–295.

Hagen, O. Risk aversion and incentive contracting. *Econ. Record*, **42** (1966), 416–429.

Hahn, F. H. Equilibrium dynamics with heterogeneous capital goods. *Quart. Jour. Econ.*, **80** (1966), 633–646.

Hahn, F. H. Savings and uncertainty. *Rev. Econ. Stud.*, **37** (1970), 21–24.

Hahn, F. H. Equilibrium with transaction costs. *Econometrica*, **39** (1971), 417–439.

Hakansson, N. H. Optimal investment and consumption strategies under risk, an uncertain lifetime, and insurance. *Intl. Econ. Rev.*, **10** (1969), 443–466.
Hakansson, N. H. Friedman–Savage utility functions consistent with risk aversion. *Quart. Jour. Econ.*, **84** (1970), 472–487.
Hakansson, N. H. Optimal investment and consumption strategies under risk for a class of utility functions. *Econometrica*, **38** (1970), 587–607.
Hakansson, N. H. Multi-period mean-variance analysis: Toward a general theory of portfolio choice. *Jour. Finance*, **26** (1971), 857–884.
Hamada, K. *Alternative approaches to dynamic portfolio selection.* Paper presented at the 5th Far Eastern Meeting, Tokyo, Japan (June 1970).
Hamada, R. S. Investment decision with a mean-variance approach. *Quart. Jour. Econ.*, **85** (1971), 667–683.
Hammer, P. L. and Rudeanu, S. *Boolean Methods in Operations Research and Related Areas.* Springer-Verlag: Berlin (1968).
Hammer, P. L. and Shlifer, E. Applications of pseudo-Boolean methods to economic problems. *Theory and Decision*, **1** (1971), 296–308.
Hammond, J. S. III, Better decisions with preference theory. *Harvard Business Rev.*, **45** (1967), 123–141.
Hammond, P. J. *Optimal saving when the probability distribution function of the yield is unknown.* Mimeographed, Cambridge (1969).
Handa, J. C. A theory of risk preference in gambling. *Jour. Pol. Econ.*, **79** (1971), 1073–1083.
Hangen, R. A. Insurer risk under alternative investment and financing strategies. *Jour. Risk Ins.*, **38** (1971), 71–80.
Hanoch, G. and Levy, H. The efficiency analysis of choices involving risk. *Rev. Econ. Stud.*, **36** (1969), 335–346.
Hanson, D. L. and Menezes, C. F. Risk aversion and bidding theory. In *Papers in Quantitative Economics* (eds. Quirk, J. P. and Zarley, A. M.). Univ. Kansas Press: Lawrence (1968).
Hanson, D. L. and Menezes, C. F. On a neglected aspect of the theory of risk aversion. *Western Econ. Jour.*, **9** (1971), 211–217.
Hanssmann, F. Probability of survival as an investment criterion. *Management Science*, **15** (1968), 33–48.
Hansson, B. Choice structures and preference relations. *Synthese*, **18** (1968), 443–458.
Haring, J. E. and Smith, G. C. Utility theory, decision theory, and profit maximization. *Am. Econ. Rev.*, **49** (1959), 566–583.
Harsanyi, J. C. Approaches to the bargaining problem before and after the theory of games: a critical discussion of Zeuthen's, Hicks' and Nash's theories. *Econometrica*, **24** (1956), 144–157.
Harsanyi, J. C. On the rationality postulates underlying the theory of cooperative games. *Jour. Conflict Resolution*, **5** (1961), 179–196.
Harsanyi, J. C. A simplified bargaining model for the *n*-person cooperative game. *Intl. Econ. Rev.*, **4** (1963), 194–220.
Harsanyi, J. C. Bargaining and conflict situations in the light of a new approach to game theory. *Am. Econ. Rev.*, **55** (1965), 447–457.
Harsanyi, J. C. A general theory of rational behavior in game situations. *Econometrica*, **34** (1966), 613–634.
Harsanyi, J. C. Notes on the so-called incompleteness problem and on the proposed alternative concept of rational behavior. *Theory and Decision*, **2** (1972), 342–352.
Hart, A. G. Risk, uncertainty, and the unprofitability of compounding probabilities.

In *Studies in Mathematical Economics and Econometrics* (eds. Lange, O., McIntyre, F. and Yntema, F.). Univ. of Chicago Press: Chicago (1942), 110–118.

Hart, A. G. *Anticipations, Uncertainty, and Dynamic Planning.* Augustus Kelley: New York (1951).

Hartman, R. The effects of price and cost uncertainty on investment. *Jour. Econ. Theory,* **5** (1972), 258–266.

Hartman, R. Adjustment costs, price and wage uncertain, and investment. *Rev. Econ. Stud.,* **40** (April 1973), 293–296.

Hampenious, A. L. *Monopoly with random demand.* Econometric Inst. Report 6904, Netherlands School of Economics (1969).

Henderson, D. W. and Turnovsky, S. J. Optimal macroeconomic policy adjustment under conditions of risk. *Jour. Econ. Theory,* **4** (1972), 58–71.

Herstein, I. N. and Milnor, J. An axiomatic approach to measurable utility. *Econometrica,* **21** (1953), 291–297.

Hibbard, T. H. Insurance and the optimal distribution of medical care. *Western Econ. Jour.,* **9** (1971), 231–241.

Hildreth, C. Alternative conditions for social ordering. *Econometrica,* **21** (1953), 81–91.

Hirshleifer, J. On the theory of optimal investment decisions. *Jour. Pol. Econ.,* **66** (1958), 329–352.

Hirshleifer, J. Efficient allocation of capital in an uncertain world. *Am. Econ. Rev.,* **54** (1964), 77–85.

Hirshleifer, J. Investment decision under uncertainty: Choice-theoretic approaches. *Quart. Jour. Econ.,* **79** (1965), 509–536.

Hirshleifer, J. Investment decision under uncertainty: Applications of the state-preference approach. *Quart. Jour. Econ.,* **80** (1966), 252–277.

Hirshleifer, J. *Investment, Interest, and Capital.* Prentice-Hall: Englewood Cliffs, N. J. (1970).

Hirshleifer, J. and Shapiro, D. L. The treatment of risk and uncertainty. In *Public Expenditures and Policy Analysis* (eds. Havenman, R. and Margolis, J.). Markham Pub. Co.: Chicago, Ill. (1970).

Hodgman, D. R. Credit risk and credit rationing. *Quart. Jour. Econ.,* **74** (1960), 258–278.

Hodgman, D. R. Credit risk and credit rationing: Reply. *Quart. Jour. Econ.,* **75** (1961), 327–329.

Howard, R. A. The foundations of decision analysis. *IEEE Transactions on Systems Science and Cybernetics,* SSC-4 (1968), 1–9.

Hurst, P. M. and Siegel, S. Prediction of decision from a higher-ordered metric scale of utility. *Jour. Experimental Psychology,* **52** (1956), 138–144.

Hurwicz, L. The theory of economic behavior. *Am. Econ. Rev.,* **35** (1945), 919–925.

Hurwicz, L. *Optimality criteria for decision making under ignorance.* Cowles Commission Discussion Paper No. 370 (1950).

Hymans, S. H. The price-taker: uncertainty, utility, and the supply function. *Intl. Econ. Rev.,* **7** (1966), 346–356.

Ishii, Y. Credit creation and portfolio decisions of commercial banks under uncertainty. *Econ. Stud. Quart.,* **22** (1971), 13–24.

Jacob, N. L. The measurement of market similarity for securities under uncertainty. *Jour. Bus.,* **43** (1970), 328–340.

Jacob, N. L. The measurement of systematic risk for securities and portfolios: some empirical results. *Jour. Fin. Quant. Anal.*, **6** (1971), 815–833.

Jafee, D. M. and Modigliani, F. A theory and test of credit rationing. *Am. Econ. Rev.*, **59** (1969), 784–793.

Jean, P. *Optimal growth in a multi-sector economy with stochastic technology.* PhD dissertation, Univ. of Calif. at Berkeley (1970).

Jean, W. H. The extension of portfolio analysis to three or more parameters. *Jour. Fin. Quant. Anal.*, **6** (1971), 505–515.

Jeffrey, R. C. Ethics and the logic of decision. *Jour. Philosophy*, **62** (1965), 528–539.

Jeffrey, R. C. *The Logic of Decision.* McGraw-Hill: New York (1965).

Jeffrey, R. C. New foundations for Bayesian decision theory. In *Logic, Methodology and Philosophy of Science,* proceedings of the 1964 International Congress (ed. Bar-Hillel, Y.), North-Holland Pub. Co.: Amsterdam (1965), 289–300.

Jensen, M. C. The performance of mutual funds in the period 1945–1964. *Jour. Finance,* **23** (1968), 389–416.

Jensen, M. C. Risk, the pricing of capital assets, and the evaluation of investment portfolios, *Jour. Bus.*, **42** (1969), 167–248.

Jensen, M. C. Capital market: Theory and evidence. *Bell Jour. Econ. Management Science,* **3** (1972), 357–398.

Jensen, M. C. The foundations and current state of market theory. In *Studies in the Theory of Capital Markets* (ed. Jensen, M. C.). Praeger Publishers: New York (forthcoming).

Jensen, M. C. and Long, J. B., Jr. Corporate investment under uncertainty and Pareto optimality in the capital markets. *Bell Jour. Econ. Management Science,* **3** (1972), 151–174.

Johnson, G. E. Uncertainty and price theory: comment. *Quart. Jour. Econ.*, **78** (1964), 647–648.

Jorgenson, D. *Anticipations, appropriations and investment behavior in US business: quarterly 1947–1960.* Mimeographed, Univ. Calif. at Berkeley (1962).

Kamien, M. I. and Schwartz, N. L. Limit pricing and uncertain entry. *Econometrica,* **39** (1971), 441–454.

Kamien, M. I. and Schwartz, N. L. Some economic consequences of anticipating technical advance. *West. Econ. Jour.*, **10** (1972), 123–138.

Kamien, M. I. and Schwartz, N. L. Timing of innovations under rivalry. *Econometrica,* **40** (1972).

Kamien, M. I. and Schwartz, N. L. Uncertain entry and excess capacity. *Am. Econ. Rev.*, **62** (1972), 918–927.

Karlin, S. *Mathematical Methods and Theory in Games, Programming, and Economics.* Vol. I. Addison-Wesley: Reading, Mass. (1959).

Katona, G. Rational behavior and economic behavior. *Psychological Rev.*, **60** (1953), 307–318.

Kaufmann, A. *The Science of Decision Making.* George Weidenfeld and Nicolson: London (1968).

Keeney, R. L. Risk independence and multiattributed utility functions. *Econometrica,* **41** (January 1973), 27–34.

King, B. F. Market and industry factors in stock price behavior. *Jour. Bus.*, **39** (1966), 139–190.

Klevorick, A. K. The behavior of a firm subject to stochastic regulatory review. *Bell. Jour. Econ. Management Science,* **4** (1973), 57–88.

Klevorick, Alvin K. A note on 'the ordering of portfolios in terms of mean and variance', *Rev. Econ. Stud.*, **40** (April 1973), 293–296.

Knight, F. H. *Risk, Uncertainty and Profit.* Houghton Mifflin: Boston (1921).

Kogan, N. and Wallach, M. A. *Risk Taking: a study in cognition and personality.* Holt, Rinehart, and Winston: New York (1964).

Koopman, B. O. The axioms and algebra of intuitive probability. *Annals Math.*, **41** (1940), 269–292.

Koopmans, T. C. Stationary ordinal utility and impatience. *Econometrica*, **8** (1960), 287–309.

Koopmans, T. C. On flexibility of future preferences. In *Human Judgement and Optimality* (eds. Maynard, W. S. and Bryan, G. L.). John Wiley: New York (1964).

Krantz, D. H., Luce, R. D., Suppes, P. and Tversky, A. *Foundations of Measurement.* Vol. I. Academic Press: New York (1971).

Kraus, A. and Melnick, A. A sequential decision model of bargaining theory. *West. Econ. Jour.*, **10** (1972), 359–369.

Krelle, W. A theory on rational behavior under uncertainty. *Metroeconomica*, **11** (1959), 51–63.

Kroeber-Riel, W. Constructs and empirical basis in theories of economic behavior. *Theory and Decision*, **1** (1971), 337–349.

Kushner, H. J. *Stochastic Stability and Control.* Academic Press: New York (1967).

Kyburg, H. E. *Probability and the Logic of Rational Belief.* Wesleyan Univ. Press: Middletown, Connecticut (1961).

Kyburg, H. E. and Smokler, H. E. *Studies in Subjective Probability.* John Wiley: New York (1964).

Latane, H. A. Criteria for choice among risky ventures. *Jour. Pol. Econ.*, **67** (1959), 144–155.

Latane, H. A. and Tuttle, D. L. Criteria for portfolio building. *Jour. Finance*, **23** (1967), 359–373.

Lavalle, I. H. On cash equivalents and information evaluation in decisions under uncertainty: Part I, Basic theory; Part II, Incremental information decisions. *Jour. Am. Stat. Assn.*, **63** (1968), 252–284.

Lees, D. S. and Rice, R. G. Uncertainty and the welfare economics of medical care: comment. *Am. Econ. Rev.*, **55** (1965), 140–154.

Leinfellner, W. Generalization of classical decision theory. In *Risk and Uncertainty* (eds. Borch, L. and Mossin, J.). London (1968), 196–210.

Leland, H. E. *Dynamic portfolio theory.* PhD dissertation, Harvard Univ. (1968).

Leland, H. E. Savings and uncertainty: The precautionary demand for saving. *Quart. Jour. Econ.*, **82** (1968), 465–473.

Leland, H. E. Theory of the firm facing uncertain demand. *Am. Econ. Rev.*, **62** (1972), 278–291.

Leland, H. E. On the existence of optimal policies under uncertainty. *Jour. Econ. Theory*, **4** (1972), 35–44.

Levhari, D. and Srinivasan, T. N. Optimal savings under uncertainty. *Rev. Econ. Stud.*, **36** (1969), 153–164.

Levi, I. *Gambling with Truth.* Knopf: New York (1967).

Levy, H. and Hanoch, G. Relative effectiveness of efficiency criteria for portfolio selection. *Jour. Fin. Quant. Anal.*, **5** (1970), 63–76.

Levy, H. and Sarnat, M. The mean-variance criterion and the efficiency frontier. *Western Econ. Jour.*, **9** (1971), 46–51.

Levy, H. and Sarnat, M. Two-period portfolio selection and investors' discount rates. *Jour. Finance*, **26** (1971), 757–761.

Lintner, J., Dividends, earnings, leverage, stock prices, and the supply of capital to corporations. *Rev. Econ. Stat.*, **44** (1962), 243–269.

Lintner, J. Security prices, risk, and maximal gains from diversification. *Jour. Finance*, **20** (1965), 587–615.

Lintner, J. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Rev. Econ. Stat.*, **47** (1965), 13–37.

Lintner, J. The aggregation of investors' diverse judgements and preferences in perfectly competitive security markets. *Jour. Fin. Quant. Anal.*, **4** (1969), 347–400.

Lintner, J. The market price of risk, size of market and investor's risk aversion: A reply. *Rev. Econ. Stat.*, **54** (1972), 206–208.

Long, J. B., Jr. Consumption–investment decisions and equilibrium in the securities market. In *Studies in the Theory of Capital Markets* (ed. Jensen, M. C.). Praeger Publishers: New York (forthcoming).

Long, J. B., Jr. Wealth, welfare and the price of risk. *Jour. Finance*, **27** (1972), 419–433.

Lorie, J. H. and Savage, L. J. Three problems in rationing capital. *Jour. Bus.*, **28** (1955), 229–239.

Lucas, R. E., Jr. and Prescott, E. C. Investment under uncertainty. *Econometrica*, **39** (1971), 659–681.

Luce, R. D. Semiorders and a theory of utility discrimination. *Econometrica*, **24** (1956), 178–191.

Luce, R. D. *Individual Choice Behavior*. John Wiley: New York (1959).

Luce, R. D. Conditional expected extensive utility. *Theory and Decision*, **3** (1972), 101–106.

Luce, R. D. and Krantz, D. H. Conditional expected utility. *Econometrica*, **39** (1971), 253–271.

Luce, R. D. and Raiffa, H. *Games and Decisions: Introduction and Critical Survey*. John Wiley: New York (1957).

MacCrimmon, K. R. Descriptive and normative implications of the decision theory postulates. In *Risk and Uncertainty* (ed. Borch, K. H.). Macmillan: New York (1967).

MacCrimmon, K. R. *Decisionmaking among multiple-attribute alternatives: a survey and consolidated approach*. Memorandum RM-4823-ARPA, The Rand Corporation, Santa Monica (1968).

Malinvaud, E. Capital accumulation and efficient allocation of resources. *Econometrica*, **21** (1953), 233–268.

Malinvaud, E. First order certainty equivalence. *Econometrica*, **37** (1969), 706–718.

Malinvaud, E. The allocation of individual risks in large markets. *Jour. Econ. Theory*, **4** (1972), 312–328.

Mandelbrot, B. The variation of certain speculative prices. *Jour. Bus.*, **36** (1963), 394–419.

Mao, J. C. Essentials of portfolio diversification strategy. *Jour. Finance*, **25** (1970), 1109–1121.

Mao, J. C. Survey of capital budgeting: theory and practice. *Jour. Finance*, **25** (1970), 349–360.

Marglin, S. The social rate of discount and the optimal rate of investment. *Quart. Jour. Econ.*, **77** (1963), 95–111.

Markowitz, H. M. Portfolio selection. *Jour. Finance*, **7** (1952), 77–91.

Markowitz, H. M. The utility of wealth. *Jour. Pol. Econ.*, **60** (1952), 151–158.

Markowitz, H. *Portfolio Selection: efficient diversification of investments.* John Wiley: New York (1959).

Marley, A. A. J. Some probabilistic models of simple choice and ranking. *Jour. Math. Psychology*, **5** (1968), 311–332.

Marschak, J. Money and the theory of assets. *Econometrica*, **6** (1938), 311–325.

Marschak, J. Neumann's and Morgenstern's new approach to static economics. *Jour. Pol. Econ.*, **54** (1946), 97–115.

Marschak, J. Rational behavior, uncertain prospects and measurable utility. *Econometrica*, **18** (1950), 111–141.

Marschak, J. Towards a preference scale for decision-making. In *Readings in Game Theory and Political Behavior* (ed. Shubik, M.). Doubleday & Co.: New York (1954), 22–32.

Marschak, J. Norms and habits of decision making under uncertainty. In *Mathematical Models of Human Behavior* (ed. Dunlap, J. W.). Dunlap and Associates: Stanford, Connecticut (1955), 44–53.

Marschak, J. Actual *versus* consistent decision behavior. *Behavioral Science*, **9** (1964), 103–110.

Marschak, J. *Decision Making: Economic Economic Aspects* (*Intl. Encyclopedia of the Social Sciences*), Vol. 4. Crowell Collier and Macmillan: New York (1968), 42–55.

Marschak, J. Economics of inquiring, communicating, deciding. *Am. Econ. Rev.*, **58** (1968), 1–18.

Marschak, J. and Nelson, R. Flexibility, uncertainty and economic theory. *Metroeconomica*, **14** (1962), 42–58.

Marschak, J. and Radner, R. *The Economic Theory of Teams.* Yale Univ. Press: New Haven and London (1972).

Masson, R. T. The creation of risk aversion by imperfect capital markets. *Am. Econ. Rev.*, **62** (1972), 77–86.

Mayers, D. Non-marketable assets and capital market equilibrium under uncertainty. In *Studies in the Theory of Capital Markets* (ed. Jensen, M. C.). Praeger Publishers: New York (forthcoming).

Mayers, D. *Non-marketable assets and the determination of capital asset prices in the absence of a riskless asset.* Working Paper No. 7114, Graduate School of Management, Univ. of Rochester (1971).

McCall, J. J. Competitive production for constant risk utility functions. *Rev. Econ. Stud.*, **34** (1967), 417–420.

McCall, J. J. Probabilistic micro-economics. *Bell Jour. Econ. and Management Science*, **2** (1971), 403–433.

McClennen, E. F. An incompleteness problem in Harsanyi's general theory of games and certain related theories of non-cooperative games. *Theory and Decision*, **2** (1972), 314–341.

McGuigan, J. R. and King, W. R. Security option strategy under risk aversion: An analysis. *Jour. Fin. Quart. Anal.*, **8** (1973), 1–15.

McKean, R. N. and Moore, J. H. Uncertainty and the evaluation of public investment decisions: comment. *Am. Econ. Rev.*, **62** (1972), 165–167.

McKinnon, R. Future markets, buffer stocks, and income stability for primary producers. *Jour. Pol. Econ.*, **75** (1967), 844–861.

Menezes, C. F. and Hanson, D. L. On the theory of risk aversion. *Intl. Econ. Rev.*, **11** (1970), 481–487.

Menges, G. On subjective probability and related problems. *Theory and Decision*, **1** (1971), 40–60.

Merton, R. C. Lifetime portfolio selection under uncertainty: the continuous time case. *Rev. Econ. Stat.*, **51** (1969), 247–257.

Merton, R. C. *An analytic derivation of the efficient portfolio frontier.* Working Paper No. 493–70, Sloan School of Management, M.I.T. (1970).

Merton, R. C. Optimum consumption and portfolio rules in a continuous-time model. *Jour. Econ. Theory,* **3** (1971), 373–413.

Merton, R. C. The theory of rational option pricing. *Bell Jour. Econ. Management Science,* **4** (1973), 141–183.

Michalos, A. C. Decision-making in committees. *Am. Philosophical Quarterly,* **7** (1970), 91–106.

Michalos, A. C. Cost–benefit *versus* expected utility acceptance rules. *Theory and decision,* **1** (1971), 61–88.

Miller, D. W. and Starr, M. K. *Executive Decisions and Operations Research.* Prentice-Hall: Englewood Cliffs (1960).

Miller, M. H. and Scholes, M. Rates of return in relation to risk: a reexamination of some recent findings. In *Studies in the Theory of Capital Markets* (ed. Jensen, M. C.). Praeger Publishers: New York (forthcoming).

Mills, E. S. Uncertainty and price theory. *Quart. Jour. Econ.*, **73** (1959), 116–130.

Mills, E. S. *Prince, Output, and Inventory Policy.* John Wiley: New York (1961).

Milnor, J. Games against nature. In *Decision Processes* (eds. Thrall, R. M., Coombs, C. H. and Davis, R. L.). John Wiley: New York (1954), 49–59.

Mirman, L. J. Uncertainty and optimal consumption decisions. *Econometrica,* **36** (1971), 177–183.

Mirman, L. On the existence of steady state measures for one sector growth models with uncertain technology. *Int. Econ. Rev.*, **31** (1972), 271–286.

Mirman, L. J. The steady state behavior of a class of one sector growth models with uncertain technology. *Jour. Econ. Theory,* **6** (1973), 219–242.

Mishan, E. J. Public investment criteria: Some simplifying suggestions. *Jour. Pol. Econ.*, **75** (1967), 139–146.

Mishan, E. J. Uncertainty and the evaluation of public investment decisions: Comment. *Am. Econ. Rev.*, **62** (1972), 161–164.

Modigliani, F. and Miller, M. The cost of capital, corporate finance and the theory of investment. *Am. Econ. Rev.*, **48** (1958), 261–297.

Moore, A. *A statistical analysis of common stock prices.* PhD Dissertation, Univ. of Chicago (1962).

Morishima, M. *Theory of Demand Real and Monetary.* Clarendon Press: Oxford (1973).

Morrison, D. G. On the consistency of preference in Allais' paradox. *Behavioral Science,* **12** (1967), 373–383.

Mossin, J. Equilibrium in a capital asset market. *Econometrica,* **34** (1966), 768–783.

Mossin, J. Aspects of rational insurance purchasing. *Jour. Pol. Econ.*, **76** (1968), 553–568.

Mossin, J. Optimal multiperiod portfolio policies. *Jour. Bus.* **41** (1968), 215–229.

Mossin, J. Taxation and risk-taking: an expected utility approach. *Economica,* **35** (1968), 74–82.

Mossin, J. Security pricing and investment criteria in competitive markets. *Am. Econ. Rev.*, **59** (1969), 749–756.

Mosteller, F. and Nogee, P. An experimental measurement of utility. *Jour. Pol. Econ.*, **59** (1951), 371–404.

Myers, S. C. A time-state-preference model of security valuation. *Jour. Fin. Quant. Anal.*, **3** (1968), 1–33.

Nash, J. F. The bargaining problem. *Econometrica*, **17** (1950), 155–162.
Nash, J. F. Two-person cooperative games. *Econometrica*, **20** (1953), 128–140.
Neave, E. The stochastic cash balance problem with fixed costs for increases and decreases. *Bell Jour. Econ. Management Science*, **1** (1970), 472–490.
Nelson, R. Uncertainty, learning and the economics of parallel research and development efforts. *Rev. Econ. Stat.*, **43** (1961), 351–364.
Nelson, R. Uncertainty, prediction and competitive equilibrium. *Quart. Jour. Econ.*, **75** (1961), 41–62.
Nevins, A. Some effects of uncertainty: Simulation of a model of price. *Quart. Jour. Econ.*, **80** (1966), 73–87.
Nichols, A. Uncertainty and the evaluation of public investment decisions: Comment. *Am. Econ. Rev.*, **62** (1972), 168–169.
Nikaido, H. *Convex Structures and Economic Theory*. Academic Press: New York (1968).

Obst, N. P. On organized markets under uncertainty. *Western Econ. Jour.*, **10** (1972), 182–192.
Oi, W. Y., The desirability of price instability under perfect competition. *Econometrica*, **29** (1961), 58–64.
Oi, W. Y. The desirability of price instability under perfect competition: Rejoinder. *Econometrica*, **31** (1963), 248.
Oi, W. Y. The economics of product safety. *Bell Jour. Econ. and Management Science*, **4** (1973), 3–28.
Okuguchi, K. A comparison of optimal outputs in risky situations (unpublished paper, 1970).
Olivera, J. H. B. On Bernoullian production sets. *Quart. Jour. Econ.*, **87** (1973), 112–120.

Pankoff, L. D. Market efficiency and football betting. *Jour. Bus.*, **41** (1968), 203–214.
Papandreou, A. G. A test of a stochastic theory of choice. *Univ. of California Publication in Econ.*, **16** (1957), 1–18.
Parkin, J. M. and Wu, S. Y. Choice involving unwanted risk events and optimal insurance. *Am. Econ. Rev.*, **62** (1972), 982–987.
Parthasarathy, K. *Probability Measures on Metric Spaces*. Academic Press: New York (1967).
Pattanaik, P. K. Risk, impersonality, and the social welfare function. *Jour. Pol. Econ.*, **76** (1968), 1152–1169.
Pauly, M. The economics of moral hazard: Comment. *Am. Econ. Rev.*, **58** (1968), 531–537.
Pauly, M. Efficiency, incentives and reimbursement for health care. *Inquiry*, **7** (1970), 114–131.
Pen, J. A general theory of bargaining. *Am. Econ. Rev.*, **42** (1952), 24–42.
Penner, R. G. Uncertainty and the short-run shifting of the corporation tax. *Oxford Econ. Papers*, **19** (1967), 99–110.
Pfanzagl, J. A general theory of measurement: Application to utility. *Naval Research Logistics Quarterly*, **6** (1959), 283–294.
Pfanzagl, J. Subjective probability derived from the Morgenstern–von Neumann utility concept. In *Essays in Mathematical Economics in Honor of Oskar Morgenstern* (ed. Shubik, M.). Princeton Univ. Press: Princeton, N. J. (1967), 237–251.
Phelphs, E. S. The accumulation of risky capital: A sequential utility analysis. *Econometrica*, **30** (1962), 729–743.
Pollak, R. A. Additive von Neumann–Morgenstern utility functions. *Econometrica*, **35** (1967), 485–494.

Pollak, R. A. Homogeneous von Neumann–Morgenstern utility functions. *Intl. Econ. Rev.*, **11** (1970), 117–130.

Porter, R. B. *Application of stochastic dominance principles to the problem of asset selection under risk.* PhD dissertation, Purdue Univ. (1971).

Porter, R. B. and Gaumnitz, J. E. Stochastic dominance *vs* mean-variance portfolio analysis: an empirical evaluation. *Am. Econ. Rev.*, **62** (1972), 438–446.

Potzsch, R. Decision-theoretic aspects of risk-taking behavior, *Theory and Decision*, **3** (1972), 107–125.

Pratt, J. W. Risk aversion in the small and in the large. *Econometrica*, **32** (1964), 122–136.

Pratt, J. W., Raiffa, H. and Schlaifer, R. The foundations of decision under uncertainty: An elementary exposition. *Jour. Am. Stat. Assn.*, **59** (1964), 353–375.

Pratt, J. W., Raiffa, H. and Schlaifer, R. *Introduction to Statistical Decision Theory.* McGraw-Hill: New York (1965).

Preston, M. C. and Baratta, P. An experimental study of the auction value of an uncertain outcome. *Am. Jour. Psychology*, **61** (1948), 183–193.

Pruitt, D. G. Pattern and level of risk in gambling decisions. *Psychological Rev.*, **69** (1962), 187–201.

Pruzan, P. M. Is cost–benefit analysis consistent with the maximization of expected utility? In *Operational Research and the Social Sciences* (ed. Lawrence, J. R.). London (1966), 319–336.

Pyle, D. H. and Turnovsky, S. J. Safety-first and expected utility maximization in mean-standard deviation portfolio analysis. *Rev. Econ. Stud.*, **52** (1970), 75–81.

Quirk, J. P. The capital structure of firms and the risk of failure. *Intl. Econ. Rev.*, **2** (1961), 210–228.

Quirk, J. P. and Saposnik, R. Admissibility and measurable utility functions. *Rev. Econ. Stud.*, **29** (1962), 140–146.

Radner, R. Team decision problems. *Annals Math. Stat.*, **33** (1962), 857–881.

Radner, R. Mathematical specification of goals for decision problems. In *Human Judgments and Optimality* (eds. Shelly, M. W., II and Bryan, G.). John Wiley: New York (1964), 178–216.

Radner, R. Competitive equilibrium under uncertainty. *Econometrica*, **36** (1968), 31–58.

Radner, R. *Existence of equilibrium of plans, prices, and price expectations in a sequence of markets.* Mimeographed, Univ. of Calif. at Berkeley (1970).

Radner, R. Problems in the theory of markets under uncertainty. *Am. Econ. Rev.*, **60** (1970), 454–460.

Radner, R. Balanced stochastic growth at the maximum rate. In *Contributions to the von Neumann growth model* (eds. Bruckman, G. and Weber, W.). Springer-Verlag: Vienna (1971), 39–62.

Radner, R. Some characterizations of optimal growth under uncertainty. Paper presented at the 1971 Fall Meeting of the Society for Industrial and Applied Mathematics, Univ. of Wisconsin-Madison (11–13 October 1971). Also in *Mathematical Topics in Economic Theory and Computation* (eds. Day, R. H. and Robinson, S. M.) Society of Industrial and Applied Mathematics: Philadelphia (1972).

Radner, R. Optimal stationary consumption with stochastic production and resources. *Jour. Econ. Theory*, **6** (1973), 68–90.

Radner, R. and Marschak, J. Note on some proposed decision criteria. In *Decision Processes* (eds. Thrall, R. M., Coombs, C. H. and Davis, R. L.). John Wiley: New York (1954), 61–68.

Raiffa, H. Risk, ambiguity, and the Savage axioms: comment. *Quart. Jour. Econ.*, **75** (1961), 690–694.
Raiffa, H. Decision analysis. In *Introductory Lectures on Choices Under Uncertainty*. Addison Wesley: Reading, Massachusetts (1968).
Raiffa, H. and Schlaifer, R. *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Administration, Harvard University: Boston (1961).
Ramsey, F. P. *The Foundations of Mathematics and Other Logical Essays*. Harcourt, Brace and Co.: New York (1931).
Ramsey, F. P. Truth and probability. In Ramsey, F. P., *The Foundation of Mathematics and Other Logical Essays*. Harcourt, Brace: New York (1931), 156–198.
Restle, F. *Psychology of Judgement and Choice: a Theoretical Essay*. John Wiley: New York (1961).
Richard, S. The theory of games. *Econ. Jour.*, **58** (1948), 185–201.
Richardson, G. B. Equilibrium, expectations and information. *Econ. Jour.*, **69** (1959), 223–227.
Richter, M. K. Cardinal utility, portfolio selection and taxation. *Rev. Econ. Stud.*, **27** (1960), 152–166.
Richter, M. K. Rational choice under uncertainty: revealed probability and revealed expectations. Paper presented at the 2nd World Congress of the Econometric Society, Cambridge, England (September 1970). Abstract in *Econometrica*, **39** (1971), 225.
Robbins, H. The empirical Bayes approach to statistical decision problems. *Annals Math. Stat.*, **35** (1964), 1–20.
Roberts, H. V. Risk, ambiguity, and the Savage axioms: Comment. *Quart. Jour. Econ.*, **77** (1963), 327–336.
Rockafellar, R. T. *Convex Analysis*. Princeton Univ. Press: Princeton, New Jersey (1970).
Roll, R. Investment diversification and bond maturity. *Jour. Finance*, **26** (1971), 51–66.
Rosett, R. N. Gambling and rationality. *Jour. Pol. Econ.*, **73** (1965), 595–607.
Ross, S. *Risk and efficiency*. Discussion Paper No. 196, Dept. of Economics, Univ. of Pennsylvania (1971).
Ross, S. The separation theorem: A distributional approach (unpublished manuscript, 1971).
Rothenberg, T. J. *The Measurement of Social Welfare*. Prentice-Hall: Englewood Cliffs, N. J. (1961).
Rothenberg, T. J. and Smith, K. R. The effect of uncertainty on resource allocation in a general equilibrium model. *Quart. Jour. Econ.*, **85** (1971), 440–459.
Rothschild, M. The risk analysis of choices involving uncertainty. Part II of *Essays in economic theory*, unpublished PhD dissertation, M.I.T. (1969).
Rothschild, M. Prices, information and market structure. Paper presented at the 2nd World Congress of the Econometric Society, Cambridge, England (September 1970).
Rothschild, M. and Stiglitz, J. E. Increasing risk I: A definition, *Jour. Econ. Theory*, **2** (1970), 225–243.
Rothschild, M. and Stiglitz, J. E. Increasing risk II: Its economic consequences. *Jour. Econ. Theory*, **3** (1971), 66–84.
Rothschild, M. and Stiglitz, J. E. Addendum to 'Increasing risk: I. A definition'. *Jour. Econ. Theory*, **5** (1972), 306.
Roy, A. D. Safety first and the holding of assets. *Econometrica*, **20** (1952), 431–449.
Royden, H. L. Suppes, P. and Walsh, K. A model for the experimental measurement of the utility of gambling. *Behavioral Science*, **4** (1959), 11–18.

Salazar, R. C. and Sen, S. K. A simulation model of capital budgeting under uncertainty. *Management Science,* **15** (1968), B161–B179.

Samuelson, P. A. Some aspects of the pure theory of capital. *Quart. Jour. Econ.,* **51** (1937), 469–496.

Samuelson, P. A. The St. Petersburg paradox as a divergent double limit. *Intl. Econ. Rev.,* **1** (1960), 31–37.

Samuelson, P. A. *Foundations of Economic Analysis.* Atheneum: New York (1965).

Samuelson, P. A. General proof that diversification pays. *Jour. Fin. Quant. Anal.,* **2** (1967), 1–13.

Samuelson, P. A. Lifetime portfolio selection by dynamic stochastic programming. *Rev. Econ. Stat.,* **51** (1969), 239–246.

Samuelson, P. A. The fundamental approximation theorem of portfolio analysis in terms of means, variances, and higher moments. *Rev. Econ. Stud.,* **37** (1970), 537–542.

Samuelson, P. A. The mathematics of speculative price. 1971 John von Neumann Lecture presented at the 1971 Fall Meeting of the Society for Industrial and Applied Mathematics, Univ. of Wisconsin-Madison (11–13 October 1971). Also in *Mathematical Topics in Economic Theory and Computation* (eds. Day, R. H. and Robinson, S. M.). Society of Industrial and Applied Mathematics: Philadelphia (1972).

Sandmo, A. Capital risk, consumption, and portfolio choice. *Econometrica,* **37** (1969), 586–599.

Sandmo, A. The effect of uncertainty on savings decisions. *Rev. Econ. Stud.,* **37** (1970), 353–360.

Sandmo, A. On the theory of the competitive firm under price uncertainty. *Am. Econ. Rev.,* **61** (1971), 65–73.

Sandmo, A. Discount rates for public investment under uncertainty. *Intl. Econ. Rev.,* **13** (1972), 287–302.

Sandmo, A. 'Fixed costs' and the competitive firm under price uncertainty: Reply. *Am. Econ. Rev.,* **62** (1972), 194–195.

Saraydar, E. Uncertainty, the bargaining problem, and the Nash–Zeuthen solution. *Theory and Decision,* **1** (1971), 309–319.

Savage, L. J. The theory of statistical decision. *Jour. Am. Stat. Assn.,* **46** (1951), 55–67.

Savage, L. J. *The Foundations of Statistics.* John Wiley: New York (1954).

Savage, L. J. The foundations of statistics reconsidered. In *Proceedings of the Fourth Berkeley Symposium.* Univ. of Calif. Press: Berkeley (1961), 575–586.

Savage, L. J. Elicitation of personal probabilities and expectations. *Jour. Am. Stat. Assn.,* **66** (1971), 783–801.

Schilbred, C. M. An experiment with bonds and risk. *Swed. Jour. Econ.,* **74** (1972), 344–355.

Schilbred, C. The market price of risk. *Rev. Econ. Stud.,* **40**, No. 122 (April 1973), 293–296.

Schlaifer, R. *Probability and Statistics for Business Decisions: an introduction to managerial economics under uncertainty.* McGraw-Hill: New York (1959).

Schlaifer, R. *Analysis of Decisions under Uncertainty.* McGraw-Hill: New York (1967).

Schmalensee, R. Option demand and consumer's surplus: valuing price changes under uncertainty. *Am. Econ. Rev.,* **62** (1972), 813–823.

Schneeweiss, H. Note on two dominance principles in decision theory. *Unternehmens,* **12** (1968), 213–216.

Schwartz, T. On the possibility of rational policy evaluation. *Theory and Decision,* **1** (1971), 89–106.

Seal, H. L. *Stochastic Theory of a Risk Business.* John Wiley: New York (1969).

Sengupta, J. K. Safety-first rules under chance-constrained linear programming. *Operations Research,* **17** (1969), 112–132.

Sengupta, S. S. *Operations Research in Sellers' Competition.* John Wiley: New York (1967).

Shackle, G. L. S. *Expectation in Economics.* The Univ. Press: Cambridge (1949).

Shackle, G. L. S. *Uncertainty in Economics.* The Univ. Press: Cambridge (1955).

Shapley, L. S. and Shubik, M. Quasi-cores in a monetary economy with non-convex preferences. *Econometrica,* **34** (1966), 805–827.

Sharpe, W. F. A simplified model for portfolio analysis. *Jour. Management Science,* **9** (1963), 277–293.

Sharpe, W. F. Capital asset prices: A theory of market equilibrium under conditions of risk. *Jour. Finance,* **19** (1964), 425–442.

Sharpe, W. F. Risk aversion in the stock market: Some empirical evidence. *Jour. Finance,* **20** (1965), 416–422.

Sharpe, W. F. Mutual fund performance. *Jour. Bus.,* **39** (1966), 119–138.

Sharpe, W. F. *Portfolio Theory and Capital Markets.* McGraw-Hill: New York (1970).

Shell, K. and Stiglitz, J. E. The allocation of investment in a dynamic economy. *Quart. Jour. Econ.,* **81** (1967), 592–609.

Shubik, M. (ed.) *Readings in Game Theory and Political Behavior.* Doubleday: New York (1954).

Shubik, M. Studies and theories of decision making. *Administrative Science Quart.,* **3** (1958), 289–306.

Shubik, M. A note on decision making and replacing sure prospects with uncertain prospects. *Management Science,* **19** (1973), 711–712.

Siegel, S. *Decision and Choice: Contributions of Sidney Siegel* (eds. Messick, S. and Brayfield, A. H.). McGraw-Hill: New York (1964).

Siegel, S. and Fouraker, L. E. *Bargaining and Group Decision Making: experiments in bilateral monopoly.* McGraw-Hill: New York (1960).

Siegel, J. J. Risk, interest, and forward exchange. *Quart. Jour. Econ.,* **86** (1972), 303–309.

Simon, H. A. Dynamic programming under uncertainty with a quadratic criterion function. *Econometrica,* **24** (1956), 74–81.

Simon, H. A. A behavioral model of rational choice. *Quart. Jour. Econ.,* **69** (1955), 99–118. Reprinted in Simon, H. A., *Models of Man, Social and Rational.* John Wiley: New York (1957), 241–260.

Simon, H. A. Rational choice and the structure of the environment. *Psychological Rev.,* **63** (1956), 129–138. Reprinted in Simon, H. A., *Models of Man, Social and Rational.* John Wiley: New York (1957), 261–273.

Simon, H. A. Theories of decision making in economics and behavioral science. *Am. Econ. Rev.,* **49** (1959), 253–283.

Simon, H. A. The logic of heuristic decision making. In *The Logic of Decision and Action* (ed. Rescher, N.). Univ. of Pittsburgh Press: Pittsburgh (1967), 1–20.

Simpson, P. B. Risk allowances for price expectations. *Econometrica,* **18** (1950), 253–259.

Smith, C. A. B. Consistency in statistical inference and decision. *Jour. Royal Stat. Society,* Ser. B, **23** (1961), 1–25.

Smith, C. A. B. Personal probability and statistical analysis. *Jour. Royal Stat. Society,* Ser. A, **128** (1[illegible]), 469–489.

Smith, K. R. The effect of uncertainty on monopoly price, capital stock and utilization of capital. *Jour. Econ. Theory,* **1** (1969), 48–59.

Smith, K. R. Risk and the optimal utilization of capital. *Rev. Econ. Stud.,* **37** (1970), 253–259.

Smith, P. F. Measuring risk on consumer instalment credit. *Management Science,* **11** (1964), 327–340.
Smith, V. L. Optimal insurance coverage. *Jour. Pol. Econ.,* **76** (1968), 68–77.
Smith, V. L. Measuring nonmonetary utilities in uncertain choices: the Ellsberg urn. *Quart. Jour. Econ.,* **83** (1969), 324–329.
Smith, V. L. Corporate financial theory under uncertainty. *Quart. Jour. Econ.,* **84** (1970), 451–471.
Smith, V. L. The borrower–lender contract under uncertainty. *Western Econ. Jour.,* **9** (1971), 52–56.
Smith, V. L. Economic theory of wager markets. *Western Econ. Jour.,* **9** (1971), 242–255.
Smith, V. L. Investment decision, uncertainty and the incorporated entrepreneur. In *Papers in Quantitative Economics II* (eds. Quirk, J. and Zarley, A.). University of Kansas Press: Kansas City (1971).
Smith, V. L. Default risk, scale, and the homemade leverage theorem. *Am. Econ. Rev.,* **62** (1972), 66–76.
Sneed, J. D. Strategy and the logic of decision. *Synthese,* **16** (1966), 270–283.
Sondermann, D. Temporary competitive equilibrium under uncertainty. Presented at European Research Conference on Economic Theory, Bergen, Norway (1971).
Soper, C. S. and Webb, L. Tisdell in uncertainty and Pareto optimality. *Econ. Record,* **40** (1964), 581–590.
Stapleton, R. C. Portfolio analysis, stock valuation and capital budgeting decision rules for risky projects. *Jour. Finance,* **26** (1971), 95–117.
Starr, R. M. Quasi-equilibria in markets with non-convex preferences. *Econometrica,* **37** (1969), 25–39.
Starr, R. Optimal production and allocation under uncertainty. *Quart. Jour. Econ.,* **87** (1973), 81–95.
Stigler, G. J. Imperfections of the capital market. In *The Organization of Industry* (ed. Stigler, G. J.). Irwin: Homewood (1968).
Stiglitz, J. E. Behavior towards risk with many commodities. *Econometrica,* **37** (1969), 660–667.
Stiglitz, J. E. The effects of income, wealth and capital income gains taxation on risk taking. *Quart. Jour. Econ.,* **83** (1969), 263–283.
Stiglitz, J. E. A re-examination of the Modigliani–Miller theorem. *Am. Econ. Rev.,* **59** (1969), 784–793.
Stiglitz, J. E. Perfect and imperfect capital markets. Paper presented at the meeting of the Econometric Society, New Orleans (1971).
Stiglitz, J. E. *On the irrelevance of corporate financial policy.* Cowles Foundation Discussion Paper No. 339, Yale Univ. (1972).
Stiglitz, J. E. On the optimality of the stock market allocation of investment. *Quart. Jour. Econ.,* **86** (1972), 25–60.
Stiglitz, J. E. Some aspects of the pure theory of corporate finance: Bankruptcies and take-overs. *Bell Jour. Econ. Management Science,* **3** (1972), 458–482.
Stigum, B. P. Competitive equilibria under uncertainty. *Quart. Jour. Econ.,* **83** (1969), 533–561.
Stigum, B. P. Entrepreneurial choice over time under conditions of uncertainty. *Intl. Econ. Rev.,* **10** (1969), 426–442.
Stigum, B. P. Balanced growth under uncertainty. *Jour. Econ. Theory,* **5** (1972), 42–68.
Stigum, B. P. Resource allocation under uncertainty. *Intl. Econ. Rev.,* **13** (1972), 431–459.
Stone, B. K. *Risk, Return and Equilibrium.* M.I.T. Press: Cambridge (1970).

Suppes, P. and Winet, M. An axiomatization of utility based on the notion of utility differences. *Management Science,* **2** (1955), 259–270.

Suppes, P. The role of subjective probability and utility in decision making. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955,* Vol. 5. Univ. of California Press: Berkeley (1956), 61–73.

Swalm, R. O. Utility theory – insights into risk taking. *Harvard Bus. Rev.,* **47** (1966), 123–136.

Telser, L. G. Safety first and hedging. *Rev. Econ. Stud.,* **23**·(1955–56). 1–16.

Theil, H. A note on certainty equivalence in dynamic planning. *Econometrica,* **25** (1957), 346–349.

Theil, H. *Economic Forecasts and Policy,* 2nd edition. North-Holland: Amsterdam (1961).

Theil, H. Some reflections on static programming under uncertainty. *Weltwirtschaftliches Archiv,* **87** (1961), 124–138.

Theil, H. On the symmetry approach to the committee decision problem. *Management Science,* **9** (1963), 380–393.

Theil, H. *Optimal Decision Rules for Government and Industry.* North-Holland: Amsterdam (1964).

Tintner, G. A pure theory of production under technological risk and uncertainty. *Econometrica,* **9** (1941), 305–312.

Tintner, G. The theory of choice under subjective risk and uncertainty. *Econometrica,* **9** (1941), 298–304.

Tintner, G. A contribution to the non-static theory of choice. *Quart. Jour. Econ.,* **9** (1942), 274–306.

Tintner, G. Foundations of probability and statistical inference. *Jour. Royal Stat. Society,* Ser. A, **112** (1949), 251–279.

Tintner, G. Time series (general). In *The International Encyclopedia of the Social Sciences* (ed. Sills, D. L.). Macmillan: New York (1968), **16**, 47–59.

Tintner, G. and Sengupta, J. K. Stochastic linear programming and its application to economic planning. On political economy and econometrics. *Essays in Honor of Oskar Lange.* PWN: Warsaw (1964), 601–618.

Tisdell, C. A. Decision making and the probability of loss. *Australian Econ. Papers,* **1** (1962), 109–118.

Tisdell, C. A. Uncertainty and Pareto optimality. *Econ. Record,* **39** (1963), 405–412.

Tisdell, C. A. Uncertainty, instability, expected profit. *Econometrica,* **31** (1963), 243–247.

Tisdell, C. A. Uncertainty and Pareto optimality: A rejoinder. *Econ. Record,* **40** (1964), 590–596.

Tisdell, C. A. Some bounds upon the Pareto optimality of group behavior. *Kyklos,* **19** (1966), 81–105.

Tisdell, C. A. *The Theory of Price Uncertainty, Production and Profit.* Princeton Univ. Press: Princeton, N. J. (1966).

Tobin, J. A survey of the theory of rationing. *Econometrica,* **20** (1952), 521–553.

Tobin, J. Liquidity preference as behavior toward risk. *Rev. Econ. Stud.,* **25** (1958), 65–68. Reprinted in *Risk Aversion and Portfolio Choice* (eds. Hester, D. D. and Tobin, J.). John Wiley: New York (1957), 1–26.

Tobin, J. The theory of portfolio selection. In *The Theory of Interest Rates* (eds. Hahn, F. H. and Brechling, F. P. R.). Macmillan: London (1965).

Tobin, J. Comment on Borch and Feldstein. *Rev. Econ. Stud.,* **36** (1969), 13–14.

Tropper, R. The consequences of investment in the process of conflict. *Jour. Conflict Resolution,* **16** (1972), 97–98.

Tsiang, S. C. The rationale of the mean-standard deviation analysis, skewness preference, and demand for money. *Am. Econ. Rev.*, **62** (1972), 354–371.
Tversky, A. Utility theory and additivity analysis of risky choices. *Jour. Experimental Psychology*, **75** (1967), 27–36.
Tversky, A. Intransitivity of preferences. *Psychological Rev.*, **76** (1969), 31–48.

Uzawa, H. Note on preference and axioms of choice. *Annals Institute Stat. Math.*, **8** (1957), 35–40.

van Horne, J. Capital budgeting decisions involving combinations of risky investments. *Management Science*, **13** (1966).
Venttsel, E. S. *An Introduction to the Theory of Games*. D. C. Heath and Co.: Boston (1963).
Vickrey, W. Measuring marginal utility by reaction to risk. *Econometrica*, **13** (1945), 319–333.
Vickrey, W. Utility, strategy, and social decision rules. *Quart. Jour. Econ.*, **74** (1960), 507–535.
Vickrey, W. Risk, utility, and social policy. *Social Research*, **28** (1961), 205–217.
von Neumann, J. and Morgenstern, O. *Theory of Games and Economic Behavior*, third edition. Princeton Univ. Press: Princeton (1953).

Wald, A. *Statistical Decision Functions*. John Wiley: New York (1950).
Weddepohl, H. N. *Axiomatic Choice Models and Duality*. Rotterdam Univ. Press: Rotterdam (1970).
Weinstein, A. A. Individual preference intransititivy. *Southern Econ. Jour.*, **34** (1968), 335–343.
Weiss, Y. The risk element in occupational and educational choices. *Jour. Pol. Econ.*, **80** (1972), 1203–1213.
Wellington, D. Uncertainty and the evaluation of public investment decisions: Comment. *Am. Econ. Rev.*, **62** (1972), 170.
White, W. H. The rate of interest, marginal efficiency of capital and investment programming. *Econ. Jour.*, **68** (1958), 51–59.
Whitmore, G. A. Third-degree stochastic dominance. *Am. Econ. Rev.*, **60** (1970), 457–459.
Williamson, O. E. and Sargent, T. J. Social choice: A probabilistic approach. *Econ. Jour.*, **77** (1967), 797–813.
Wilson, C. Z. and Alexis, M. Basic frameworks for decisions. *Jour. of the Academy of Management*, **5** (1962), 151–164.
Wilson, R. B. On the theory of syndicates. *Econometrica*, **36** (1968), 119–132.
Wilson, R. B. The structure of incentives for decentralization under uncertainty. In *La Décision*. Editions du Centre National de la Recherche Scientifique: Paris (1969).
Wilson, R. B. The game-theoretic structure of Arrow's general possibility theorem. *Jour. Econ. Theory*, **5** (1972), 14–20.
Winkler, R. L. The assessment of prior distributions in Bayesian analysis. *Jour. Am. Stat. Assn.*, **62** (1967), 776–800.
Winkler, R. L. The consensus of subjective probability distributions. *Management Science*, **15** (1968), B61–B75.
Wippern, R. F. Utility implications of portfolio selection and performance appraisal models. *Jour. Fin. Quant. Anal.*, **6** (1971), 913–924.
Wolfowitz, J. Bayesian inference and axioms of consistent decision. *Econometrica*, **30** (1962), 470–479.

Wu, S. Y. The effects of vertical integration on price and output. *Western Econ. Jour.*, **2** (1964), 117–133.

Wu, S. Y. Sales curve in a competitive market under the conditions of dynamic uncertainty. *Metroeconomica,* **23** (1971), 66–80.

Yaari, M. E. On the consumers' lifetime allocation process. *Intl. Econ. Rev.,* **5** (1964), 304–317.

Yaari, M. E. Uncertain lifetime, life insurance and the theory of the consumer. *Rev. Econ. Stud.,* **32** (1964), 137–150.

Yaari, M. E. Convexity in the theory of choice under risk. *Quart. Jour. Econ.*, **79** (1965), 278–290.

Yaari, M. E. Some remarks on measure of risk aversion and on their uses. *Jour. Econ. Theory,* **1** (1969), 315–329.

Yaari, M. *A law of large numbers in the theory of consumer's choice under uncertainty.* Working Paper No. CP–330, Center for Research in Management Science, Univ. of Calif. at Berkeley (1971).

Zabel, E. The competitive firm and price expectations. *Intl. Econ. Rev.,* **10** (1969), 467–478.

Zabel, E. Monopoly and uncertainty. *Rev. Econ. Stud.,* **37** (1970), 205–219.

Zabel, E. Risk and the competitive firm. *Jour. Econ. Theory,* **3** (1971), 109–133.

Zeckhauser, R. Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives. *Jour. Econ. Theory,* **2** (1970), 10–26.

Zucker, A. On the desirability of price instability: An extension of the discussion. *Econometrica,* **33** (1965), 437–441.

APPENDIX: PAPERS PRESENTED AT THE NSF–NBER CONFERENCES ON DECISION RULES AND UNCERTAINTY

1. Akerlof, G. A theory of information in labor markets. April 6–7, 1973, Princeton University.
2. Arrow, K. and Sheshinski, E. A stochastic production function for the repairman problem. May 27–29, 1971, MIT.
3. Brock, W. A. On the modeling of expectations and perfect foresight. May 23–27, 1972, The University of Iowa.
4. Chipman, J. On ordering of portfolios in terms of mean and variance. May 27–29, 1971, MIT.
5. Diamond, P. and Stiglitz, J. Increases in risk aversion. May 23–27, 1972, The University of Iowa.
6. Diamond, P. Single activity accidents. April 6–7, 1973, Princeton University.
7. Fields, G. Educational systems and labor markets. April 6–7, 1973, Princeton University.
8. Fishburn, P. On the foundations of decision making under uncertainty. May 23–27, 1972, The University of Iowa.
9. Green, J. Stability of the stochastic edgeworth barter process. January 7–9, 1971, MIT.
10. Green, J. Temporary general equilibrium in a sequential trading model with spot and futures transactions. May 23–27, 1972, The University of Iowa.
11. Hadar, J. and Rusell, W. Stochastic dominance in choice under uncertainty. May 23–27, 1972, The University of Iowa.
12. Hamada, K. Ability spread and the optimal linear income tax. May 23–27, 1972, The University of Iowa.
13. Hildreth, C. Effects of current prospects on decisions. May 23–27, 1972, The University of Iowa.
14. Jaffe, D. and Russell, T. Information asymmetry, self-selection, and credit rationing. April 6–7, 1973, Princeton University.
15. Kirman, A. and Sobel, M. Dynamic oligopoly with inventory. May 23–27, 1972, The University of Iowa.
16. Kurz, M. Arrow–Debreu equilibrium of an exchange economy with transactions costs. May 23–27, 1972, The University of Iowa.
17. Leland, H. Optimal growth in a stochastic environment. May 23–27, 1972, The University of Iowa.
18. McFadden, D., and Richter, M.K. Revealed stochastic preference. January 7–9, 1971, MIT.
19. Merton, R. Continuous time general equilibrium models of portfolio selection. January 7–9, 1971, MIT.

20. Mirrlees, J. Notes on welfare economics, information and uncertainty. May 23–27, 1972, The University of Iowa.
21. Mirrlees, J. Optimal growth under uncertainty. May 23–27, 1972, The University of Iowa.
22. Radner, R. Optimal stationary consumption with stochastic production and resources. May 23–27, 1972, The University of Iowa.
23. Reddy, B., Choice of assets under uncertainty. May 23–27, 1972, The University of Iowa.
24. Richter, M. K. Revealed von Neumann–Morgenstern utility. January 7–9, 1971, MIT.
25. Ross, S. Economic theory of agency. May 23–27, 1972, The University of Iowa.
26. Rothschild, M. Market organization models with incomplete information and uncertainty. May 27–29, 1971, MIT.
27. Rothschild, M. and Stiglitz, J. Competitive insurance markets. April 6-7, 1973, Princeton University.
28. Salop, J. and Salop, S. Self-selection devices and turnover in the labor market. April 6–7, 1973, Princeton University.
29. Spence, M. Competitive and optimal responses to signals: an analysis of efficiency and distribution. April 6–7, 1973, Princeton University.
30. Stiglitz, J. The monopolistically competitive nature of equilibrium under uncertainty when the number of securities is less than the number of states. January 7–9, 1971, MIT.
31. Stiglitz, J. The theory of 'screening', education, and the distribution of income. April 6–7, 1973, Princeton University.
32. Stigum, B. Balanced growth under uncertainty in decomposable economies. May 23–27, 1972, The University of Iowa.
33. Tversky, A. Judgment under uncertainty. May 23–27, 1972, The University of Iowa.
34. Winter, S. Stochastic decision rules and survival of the firm. May 27–29, 1972, MIT.
35. Wold, H. Planning *vs.* research innovations. May 23-27, 1972, The University of Iowa.
36. Yaari, M. Consumer theory with random earnings. January 7–9, 1971, MIT.
37. Yaari, M. A price characterization of efficient random variables. April 6–7, 1973, Princeton University.
38. Zabel, E. Consumption and portfolio choices with transaction costs. May 23–27, 1972, The University of Iowa.

AUTHOR INDEX

SUBJECT INDEX

ESSAYS ON ECONOMIC BEHAVIOR
UNDER UNCERTAINTY

CONTRIBUTIONS
TO
ECONOMIC ANALYSIS

88

Honorary Editor

J. TINBERGEN

Editors

D. W. JORGENSON

J. WAELBROECK

NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM · OXFORD

AMERICAN ELSEVIER PUBLISHING COMPANY, INC. – NEW YORK