Foundations and Trends® in Accounting

Accounting Theory as a Bayesian Discipline

Suggested Citation: David Johnstone (2018), "Accounting Theory as a Bayesian Discipline", Foundations and Trends® in Accounting: Vol. 13, No. 1-2, pp 1–266. DOI: 10.1561/1400000056.

David Johnstone
School of Accounting Economics and Finance
University of Wollongong
and
Discipline of Finance
University of Sydney
Australia
djohnsto@uow.edu.au

now
the essence of knowledge
Boston — Delft

Contents

Accounting Theory as a Bayesian Discipline

David Johnstone

Professor of Accounting, University of Wollongong
Honorary Professor of Finance, University of Sydney

ABSTRACT

The Bayesian logic of probability, evidence and decision is the presumed rule of reasoning in analytical models of accounting disclosure. Any rational explication of the decades-old accounting notions of "information content", "value relevance", "decision useful", and possibly conservatism, is inevitably Bayesian. By raising some of the probability principles, paradoxes and surprises in Bayesian theory, intuition in accounting theory about information, and its value, can be tested and enhanced. Of all the branches of the social sciences, accounting information theory begs Bayesian insights. This monograph lays out the main logical constructs and principles of Bayesianism, and relates them to important contributions in the theoretical accounting literature. The approach taken is essentially "old-fashioned" normative statistics, building on the expositions of Demski, Ijiri, Feltham and other early accounting theorists who brought Bayesian theory to accounting theory. Some history of this nexus, and the role of business schools in the development of Bayesian statistics in the 1950–1970s, is described. Later developments in accounting, especially noisy rational expectations models under which the information reported by firms is endogenous, rather than unaffected or "drawn from nature", make the task of Bayesian inference more

David Johnstone (2018), "Accounting Theory as a Bayesian Discipline", Foundations and Trends® in Accounting: Vol. 13, No. 1-2, pp 1–266. DOI: 10.1561/1400000056.

difficult yet no different in principle. The information user must still revise beliefs based on what is reported. The extra complexity is that users must allow for the firm's perceived disclosure motives and other relevant background knowledge in their Bayesian models. A known strength of Bayesian modelling is that subjective considerations are admitted and formally incorporated. Allowances for perceived self-interest or biased reporting, along with any other apparent signal defects or "information uncertainty", are part and parcel of Bayesian information theory.

1

Introduction

This monograph introduces Bayesian theory and its role in statistical accounting information theory. Its intended audience includes accounting PhD students and researchers. The Bayesian statistical logic of probability, evidence and decision lies at the historical and modern epicenter of accounting thought and research. It is not only the presumed rule of reasoning in analytical models of accounting disclosure but also the default position for empiricists when hypothesizing about how the users of financial statements think:

> Based on Bayesian decision theory research (e.g. DeGroot, 1970) that shows that loss-minimizing investors place less weight on noisier (i.e. more uncertain) information, we expect to observe more muted initial market reactions to unexpected earnings signals that have higher information uncertainty. (Francis *et al.*, 2007, p. 408)

Bayesian logic comes to light throughout accounting research. It is the soul of most strategic disclosure models, for the reason that any other model of investor behavior implies an incoherence or inconsistency in beliefs and actions by which the investor will overall surely lose to a more coherent market or opponent:

> In theory-based, economic analyses, reliance on Bayes rule is so routinized an assumption as rarely to warrant any justification. The compelling feature of Bayes rule is that it implies the most efficient use of information. Consequently, in market settings, investors who use information more efficiently (i.e. Bayesians) should be able to exploit and dominate their less efficient counterparts. (Verrecchia, 2001, p. 123)

Bayesianism is similarly a large part of the stated and unstated motivation of empirical studies of how market prices and their implied costs of capital react to better financial disclosure. Investors are taken to impose discount rates or costs of capital consistent with their best possible (i.e. most rational) probability assessments.

Summarizing their philosophical position, Chen and Schipper (2016) argued for theory to play a greater part in accounting PhD programs and in empirical research designs. Their view of accounting is overtly Bayesian. They highlight the role of accounting measurements as information for fundamental analysis, which is understood as the formation of beliefs about the firm's cash flows and risks, culminating in financial investment decisions:

> Analyses of different accounting measurement attributes (for example, fair value and historical cost) illustrate the potential benefit of using theory to discipline empirical analysis. A general question that accounting researchers are interested in is whether accounting measurements matter, in the sense of whether different accounting measurement attributes for the same item lead to differences in investors' assessment of firms' fundamentals and therefore affect investors' decision-making. (Chen and Schipper, 2016)

Similarly, Barth (2006b) notes an array of market effects that are indicative of accounting information having met its objective, namely to alter investors' beliefs and thus actions:

> Some [empirical research] designs use capital market metrics, other than equity market value, such as trading volume,

> cost of capital estimates, and bond ratings. These studies help to provide insights into the role of accounting in capital markets. Beaver (1968) is the seminal paper in this literature and shows that accounting information changes investors' beliefs by showing that trading volume increases at earnings announcement dates. (Barth, 2006b, p. 95)

It could be argued that using information for decision-making — and hence logical (i.e. Bayesian) reasoning — all goes without saying. The retrospective provided by Chen and Schipper suggests otherwise. They explain that even theoretically formal and rigorous valuation models, like the Ohlson residual income model, are essentially non-Bayesian, because they feed accounting information into a finance-based valuation model rather than feeding Bayes theorem. Any implicit belief revision upon the Ohlson framework is not brought to light:

> This valuation approach does not model how investors use accounting information to update their beliefs about firms' future dividends. Therefore, the value relevance literature circumvents what some might view as a basic question to be asked about differences in accounting measurement attributes, namely, do the different measurements indeed result in differences in information used by investors. Furthermore, because the valuation model is silent on what "information content" and "value relevance" mean and how they are affected by different measurements, it has limited ability to guide research designs and to help researchers draw meaningful inferences. Consequently, much of the existing literature has relied on ad hoc specifications, and focused on assessments of explanatory power and assessments of regression coefficients linking accounting outcomes such as earnings to market outcomes such as price or return. Absent a theory or at least an analytical structure explicitly considering investors' use of information (e.g., investors' prior, Bayes updating), the interpretations of these results must of necessity be ad hoc. . . . We are not implying that the residual

> income frameworks revived by Ohlson (1995) and others have no value. In fact, we believe this research provides useful insights on the role of accounting measurement. Our point is that this research is not suitable to answer questions related to how investors use accounting data to update their assessments of estimates of future cash flows. (Chen and Schipper, 2016)

Any attempt to explicate the decades-old accounting notions of "information content", "value relevance", "decision useful" and the like, is inevitably a Bayesian task. It is fair to say that in the human logic of reasoning under uncertainty, probability theory (and thus Bayes' theorem) is the only candidate (we would not draw balls from an urn, and make inferences about its contents, on any formal understanding other than the laws of probability).

Frequentist or "classical" statistics, which we have probably all studied, refuses to play that game. It is not permitted under frequentist statistical theory to put a probability of any description on a proposition or "hypothesis". We can write $f(data|hypothesis)$, provided that we interpret f as frequency, but we cannot write $f(hypothesis|data)$ on any interpretation of f. So, for example, we cannot use accounting data to come to an assessed probability of a firm going bankrupt, which of course means that we cannot revise that probability when new accounting data arrives.

Subjectivist Bayesian inference supports inferences drawn from accounting "measurements" or "numbers" and does not need input observations/signals to have any substantive meaning other than as merely a "signal". Just as we can use a barometer to give an "indicator" of what weather to expect, while not necessarily giving that reading of barometric pressure any deeper scientific interpretation, Bayesian theory shows that extensibly "meaning-free" or merely "hard to interpret" accounting disclosures (and non-disclosures) can be decision-useful indicators of economic fundamentals. That understanding of Bayesian belief revision and decision-making was brought to accounting theory by Feltham, Demski and others in the 1960s and 1970s, and mirrored the rise of neo-Bayesianism in other fields in the 1950s–1960s, which in turn followed a burst of statistical work in decision theory, operations research and code breaking during WWII.

The approach taken in this monograph is a Demski-like treatment of "accounting numbers" as "signals" rather than as "measurements". It should be of course that "good" measurements like "quality earnings" reports make generally better signals. However, to be useful for decision-making under uncertainty, accounting measurements need to have more than established accounting measurement virtues, of the types that early theorists like Paton, Bell and Sterling might have advocated, and which recently resurfaced in the 1960s/1970s-like normative discussion in Hodder *et al.* (2014) and Dechow *et al.* (2010). Chen and Schipper's view is that accounting measurements need to possess enough technical Bayesian information attributes to materially influence users' beliefs and consequent investments. This monograph is really about explaining what those Bayesian information attributes are, where they come from in Bayesian theory, and how they apply in statistical accounting information theory.

2

Bayesianism Early in Accounting Theory

Bayesian accounting theory emerged remarkably early in the rise of modern Bayesianism in statistics. That wider Bayesian movement arose partly from work done during WWII and took root in US Universities during the 1950s and 1960s.

The following brief history first describes how Bayesian methods were rediscovered in statistics, and how the newly enthused and influential Bayesian school, including leaders like L.J. Savage, went into dispute with the conventional frequentist doctrine that had risen and dominated statistical theory in the period between the wars. Then comes some detail of how a Bayesian movement arose in US business schools, most clearly at the Harvard Business School.

The final piece of this short historical picture remarks on how the rise of Bayesianism in a few very influential US statistics departments and business schools, quite simultaneously, gave rise to a new and rich view of accounting theory, whereby accounting disclosures, or accounting "numbers", were interpreted not only as "measurements" but ultimately as "signals" evaluated by their relevance and effect on users' Bayesian beliefs and Bayesian decision-making.

2.1 Rise of Bayesian statistics

Historians of statistics note that once upon a time there was only Bayesian statistics.[1] Inductive inference was called "inverse probability" and Bayes theorem was the only way.[2] That changed strongly in the 1920s and 1930s with the ingenious mathematical developments of frequentist theoretical tools, mainly by R.A. Fisher, Jerzy Neyman and Egon Pearson.

Fisher developed his theory of "significance tests" and Neyman altered and re-interpreted that theory as one of "hypothesis tests". The testing method that emanated out of a fierce disagreement between the two avowedly non-Bayesians was a hybrid model of testing which nowadays is known interchangeably by the two names and mixes up their methods and interpretations in ways that camouflage their historical disagreement, see Johnstone (1986), Johnstone (1987b), and Johnstone (1987a). There are many excellent surveys of this philosophical dispute. Its importance to real-world statistical practice brought it to the attention of philosophers of science, see particularly Howson and Urbach (2005) for an overview. See also Berger (2003) for a retrospective.

The next great shift came after WWII when other legendary names in the field returned to Bayes theorem as the logical law of inference under uncertainty and rejected on principle the Fisherian and Neyman–Pearson logic of statistical tests. That rebellion caused a deep philosophical and political divide in statistics as a discipline, and gave rise to disputes that were both personal and professional, and which influenced the way statistics was taught all over the world.

Neyman had fallen out with Fisher and gone to the US from Cambridge. At Berkeley he founded the first statistics department in the US and through his PhD students, who did the same at other US schools, Neyman in effect taught generations of American students strictly frequentist, strictly Neyman–Pearson, methods.

---

[1] For some detailed history, see Fienberg (2006).

[2] Keynes's lesser known treatise, "A Treatise on Probability" (1921), gives a detailed history of Bayes theorem and induction before the rise of frequentist statistics.

David Blackwell, whose name is well known in accounting theory, was also at Berkeley and was in today's terms a staunch "Bayesian", at least by the time that he wrote his introductory textbook, Blackwell (1969).[3] A measure of the unfamiliarity and likely also contentiousness of Bayesianism at the time is that this book, which advocates the elements of Bayesian reasoning, did not use the word Bayes (DeGroot, 1986, p. 44).

The UK remained less strictly anti-Bayesian, beginning from Fisher's interpretation of statistical tests being substantially more Bayesian than frequentist. He interpreted a p-level as a measure of inductive evidence regarding a null hypothesis, which is the same "type of interpretation" as Bayesians put on their posterior probability of that hypothesis (although the calculations themselves are very different). That interpretation was a large part of how he and Neyman disagreed so deeply. Neyman had another interpretation that disallowed any mention of weight of evidence or belief.[4]

The net result of Neyman's move and prodigious abilities was that statistics departments in the US were originally dominated by frequentist teachings. But in the midst of this long period of mathematical advances in frequentist statistics, the famous names of von Neumann, Savage, Blackwell, Pratt, Roberts, Raiffa, Schlaifer, Zellner and others rediscovered Bayesianism and reformed the theory of Bayesian inference and decision analysis as a mathematical calculus deduced from axioms of rationality.

In the UK the same neo-Bayesian movement already existed in the work of now celebrated, yet still widely unknown, Harold Jeffreys, and around the same time I.J. Good and D.V. Lindley, both of whom came to Bayesianism on "conceptual grounds" and were influenced by working at Bletchley Park with "the man who won the war", Alan

[3] Fienberg (2006, p. 18) writing about Blackwell: "David Blackwell's pioneering work with Girshick brought him close to Bayesian ideas, but he [Blackwell] has observed: 'Jimmie [Savage] convinced me that the Bayes approach is absolutely the right way to do statistical inference'."

[4] At the end of this chapter, I describe that interpetation, which was essentially "non-scientific" in Fisher's view, and has often been disparaged by Neyman's Bayesian critics.

Turing.[5] Underlying all this, and later to become perhaps the most celebrated Bayesian theorist of all, was de Finetti (1937/1964/1964), first published in French and later in English. See also de Finetti's (1974/5) textbook. Before them, and later to gain great respect, the Cambridge mathematician Ramsey (1926) had already discovered a set of axioms of rational betting that would prompt later proofs by Savage, DeGroot and others of the principle of maximizing expected utility as the subjectivist rule of economic rationality.[6]

2.2 Bayes in US business schools

> Summing up, we may say that Bayesianism was at least "in the air" [in 1960–1961] in the Department of Statistics of places like Carnegie, Chicago and Stanford which also happened to host the leading business schools of the new, "scientific" kind. Harvard Business School may also be added to the list, thanks to the teaching of scholars like Raiffa, Schlaifer and, later, John Pratt. While of course the conjecture awaits further confirmation (say, by examining MBA curricula and reading lists, or by investigating personal relations between economists and Bayesian statisticians in those very universities), it may provisionally be concluded that the dramatic change underwent by management teaching in the second half of the 20th century has probably played a major role in the spreading of Bayesianism in general, and of SEUT [subjective expected utility theory] in particular, within contemporary economics. To reword Marschak's dictum once again, it was not that homo economicus directly became a Bayesian statistician, but, possibly, that, first,

[5] See Banks (1996) for a fascinating account of how Bayesian calculations were invented by Turing and others in war time and while working on Enigma. Works by Good that resulted include Good (1950), Good (1952), and Good (1965). Lindley, who was younger, came later with Lindley (1965, Parts 1 and 2). A popular account is Mcgrayne (2011).

[6] Ramsey died at only 26, which along with Savage's early death, and later DeGroot, are regarded as greatly affected what would have been the development and takeup of subjective Bayesian thinking.

> homo managerialist was taught to behave that way, and, then, that economic agents came to be modeled as Bayesian corporate managers… As far as Stanford is concerned, it may be added that Robert Wilson, the giant of contemporary game and decision theory who has been at Stanford Business School since 1964, learned his Bayesian skills under Raiffa at HBS. The list of Wilson's students and colleagues at Stanford reads like a who's who in the application of Bayesian methods to modern economics. (Giocoli, 2013, p. 92)

Business schools in the US can rightfully claim to have been close to the forefront of the Bayesian uprising in the 1960s. Several of the most influential ever Bayesian theorists spent their careers employed at Harvard Business School. Harvard had Raiffa and Schlaifer and Pratt. L.J. Savage, whose work was as much economics as statistics, worked with Friedman and Pratt and had immense cross-disciplinary influence.[7] Savage (1954) published his *Foundations of Statistics* and Schlaifer (1959) followed with an applied book, ostensibly for MBA students, called *Probability and Statistics for Business Decisions*.

> The basic ideas of statistical decision theory were conceived by Schlaifer independently of the work of L. J. Savage or de Finetti, and early on he saw that those ideas were broadly applicable to problems in decision making under uncertainty. He was a pioneer in the practical assessment of subjective probabilities and utilities. (Robert Winkler quoted; presenting the 1992 *Decision Analysis Society* Frank P. Ramsey Medal to Schlaifer)

Another Bayesian founder who spent a career at the University of Chicago's Business School was Arnold Zellner. The Bayesian activity at HBS drew Lindley from the UK to spend sabbaticals with Savage and Pratt, and later, although not in the business school, with Blackwell

[7] See Lindley (1980) and Fienberg (2006).

at Berkeley.[8] Much of the Bayesian history of the Harvard Business School is recounted in Fienberg's (2008) conversation with Howard Raiffa. Fienberg, another celebrated Bayesian statistician, remembered the business school being ahead of the statistics department in its development of Bayesian thinking:

> In 1964 I arrived as a graduate student at Harvard and in my first class on statistical inference, a faculty member, whose name I will not mention, began teaching inference from a Neyman–Pearson perspective, that is, hypothesis testing and confidence intervals. ... I asked around the department about what alternatives were available to me and someone said: "On Monday they have a seminar at the business school across the Charles river." ... I recall that it was one of the most animated and heated discussions I had engaged in ... (Fienberg, 2008, p. 137)

Accounting theorists will be interested in how, at the business school, Raiffa came to see himself as a Bayesian decision analyst rather than mere statistician. Raiffa explained as follows:

> In the academic year 1961–1962, ... I was a second reader of a thesis proposal by Jack Grayson, a student in the Business School ... Grayson was interested in financial decisions of oil wildcatters. ... The drilling of an exploratory well was simultaneously a terminal-action and an information-gathering move. How should they form syndicates for sharing risks? This leads to decision problems galore, and the problems did not easily conform to the classical statistical paradigm ... Schlaifer agreed with me and we began to think of ourselves more as decision analysts than as statisticians. With my new orientation I saw problems all over the place in business, in medicine, in engineering, in public policy, where the decision problems under uncertainty did not fit comfortably into

[8] "In the early 1960s we had a series of distinguished Bayesians (Lindley, Box and Tiao), who each spent a semester at the Harvard Business School." (Howard Raiffa quoted in Fienberg, 2008, p. 145).

> the classical mold. . . . and we started the Decision Under Uncertainty Seminar that you, Steve, referred to earlier. (Raiffa quoted in Fienberg, 2006, pp. 12–13)

The Bayesian crusade at Harvard Business School was not all a success, which might explain why it is only now, more than 50 years later, that Bayesianism has become in many circles almost conventional and easily accepted:

> In the mid-1960s Robert [Schlaifer] introduced a required course in the first year of the MBA entitled Managerial Economics. All 800 students were exposed to cases that featured decision making under uncertainty. It was a heroic effort that was not universally appreciated . . . At semester's end the students burned one of Robert's books, one that I believed deserved a prize for innovation. (Raiffa quoted in Fienberg, 2006, p. 13)

The Bayesians at Harvard are credited by Fienberg as the second major force in the development of the Bayesian school, following, in both time and importance, the statistics department at Chicago, which in the 1950s included Roberts and Wallace, and Pratt. DeGroot was a graduate student, and Lindley was a visitor. It was here and in this period that the word "Bayesian" first began to appear. Fienberg (2006, p. 18) holds along with others that the Bayesian movements at Chicago and Harvard Business School arose largely independently, and it is widely claimed that Schlaifer at Harvard came to invent Bayesian decision theory from the ground up, after being dissatisfied with the frequentist principles' inapplicability to business decision-making under uncertainty.

From the start, the Bayesian paradigm and the classical or frequentist orthodoxy did not warm to each other, and that dispute was part of Raiffa's ending up in the Harvard Business School in 1964. In his personal account Raiffa wrote as follows:

> It's not surprising that the rumblings against the Neyman–Pearson school were most pronounced at the University of Chicago, the home of the Cowles Commission, which mixed

> together mathematical economists (like Jacob Marschak and Tjalling Koopmans) with statisticians (like Jimmy Savage, Herman Chernoff, and Herman Rubin). It was the articulation of Rubin's sure-thing principle in a paper by Chernoff (1954) that led me to embrace the subjective school. My religious-like conversion did not come lightly, since all I was teaching about (tests of hypotheses, confidence intervals, and unbiased estimation) was, in my newly held opinion, either wrong or not central. But my colleagues in the statistics departments were so violently opposed to using judgmental probabilities that I became a closet subjectivist. To them, statistics belonged in the scientific domain, and the introduction of squishy judgmental probabilities where opinions differed did not belong in this world of hard science.
>
> The seminal book by Savage (1954) did not so much convert me to the subjectivist camp — I was already converted intellectually by the time I read this bible — but it convinced me that I was on the right track. (Raiffa, 2002, p. 180)

2.3 Early Bayesian accounting theorists

The early (circa 1945–1960) accounting subjects at HBS were oriented to management control and accounting and statistical methods were treated as "scaffolding" or technical foundations (Zeff, 2008). Theorists including Raiffa, Schlaifer and Harry Roberts, who are known today as founders in Bayesian statistics, were directly involved in the teaching of financial decision-making:

> But it was a difficult job to find instructors who could teach both accounting and statistics. Robert K. Jaedicke, who was an assistant professor at Harvard Business School in 1958–61, recalls, "Howard Raiffa and Bob Schlaifer used to teach us statistical theory so we could teach it to the first years!!" (Zeff, 2008, p. 186).

Statistics as a discipline has a way of celebrating its pioneers and personal favorites. The journal *Statistical Science* was set up with that partly in mind, and perhaps because of its early strong orientation towards Bayesian statistics, under the editorship of DeGroot, whose textbook on decision theory is widely cited in Accounting, we have learnt a great deal about the influences and personalities that affected celebrated Bayesians like those at Harvard.

There is not yet quite as much detail available about our early Bayesians decision theorists in accounting, although it seems clear that many of those who studied in MBA and PhD programs at Harvard and equally ranked US business schools must have gained much of their Bayesian direction indirectly, or directly, from the Bayesian statisticians and economists who were on the rise in the 1960s and 1970s.

From my reading of the accounting literature, the first "modern Bayesian" works in accounting information theory were, in no particular order, by Feltham (1968), Demski (1969), Feltham and Demski (1970), Demski and Feltham (1972), Gonedes and Ijiri (1974), Gonedes (1975) and Gonedes (1976), Ng (1978), Lee and Bedford (1969) and Theil (1969).

Swieringa *et al.* (1976) note the simultaneous rise of Bayesian work in other branches of accounting research, listing many papers and reinforcing the impression that there was much evangelism about the potential of Bayesian methods in the 1970s:

> Research in financial accounting, managerial accounting, and auditing has reflected increased use of Bayesian decision theory. . . . A major feature of Bayesian decision theory is its reliance on subjective probability, which it regards as the quantified opinion of an idealized person faced with uncertainty. The subjective probability of an event is defined by the set of bets about the event such a person is willing to accept, and an internally consistent (or coherent) subjective probability can be derived for the person if his choices among bets about this event satisfy the axioms of the theory. A major contribution of Bayesian decision theory to accounting research is that by embedding a subjective interpretation of probability in a general theory of rational decision making, it provides for explicit recognition of uncertainty. Instead of

> focusing on models in which decision variables have known values or are treated as certainty equivalents, accounting researchers have tried to focus on models that incorporate the random nature of decision variables. (Swieringa *et al.*, 1976, p. 159)

This quote is essentially saying, correctly by the modern Bayesian outlook, that Bayesian methods treat unknown parameters as uncertain or "random" quantities rather than as fixed unknowns, and therefore make them subjects of subjective probability distributions. Another early indicator of the invention of statistical accounting theory came in the application of probabilistic information theory to accounting by Theil (1969) and Bedford and Baladouni (1962). See Ross (2016) for details.

Bayesian expositions in the US accounting literature appeared very early in the rise of modern Bayesianism in statistics and information economics, and anticipated much of what was yet to become a much more widely appreciated theoretical framework.[9] It could be argued that the paper by Demski and Feltham (1972) sets out essentially all of the same Bayesian subjective probability and decision paradigm as still obtains, including a prelude to the current initiative of stating forecasts not as points or intervals but as probability statements. It is another interesting part to the history of Bayesian ideas in accounting theory that both Gonedes and Ijiri (1974) and Scott (1979) wrote about the use of probability scoring rules. Scoring rules are decidedly Bayesian in their philosophy and origins, and were largely invented by de Finetti. They are only now becoming well known and widely applied in economic forecasting and forecast evaluation (see later discussion). Indicative of the overlap during this period between Bayesian theorists and accounting, Gonedes and Ijiri (1974) is essentially a study on Bayesian subjective probability and forecast evaluation, citing all the Bayesian "names" listed in this monograph.

[9]Beck *et al.* (2012, p. 72) claim that the first articles using Bayes theorem in the *Journal of Finance* appeared in 1972. Formal Bayesian portfolio theory began in the 1970s with Winkler (1973), Winkler and Barry (1975) and Barry and Winkler (1976). Other early Bayesian studies on parameter risk and portfolio choice include Brown (1979) and Klein and Bawa (1976). Bawa *et al.* (1979) summarized this literature.

Feltham (2007), in his tribute to Demski, cites his own PhD dissertation and those of Butterworth and Mock as the first works at Berkeley applying decision analysis in accounting theory. See Butterworth (1972) and Mock (1971) and Mock (1973). Mattessich (2003) credited Feltham's own 1967 PhD thesis as the first of the modern information economics ilk.[10]

All of the early Bayesian work in accounting appears to have been in the US and done by US-trained researchers, which can be attributed to the roles and influence of Savage and other Bayesian statistical theorists within the major US Business Schools. There was no parallel outbreak of Bayesian accounting theory in the UK or elsewhere.

The Bayesian revolution in accounting theory is summed up in one conceptual shift, namely, the treatment of accounting reports or "accounting numbers" not as measurements, but as statistical signals. Rather than a variable like earnings being treated as a measure or representation of some underlying physical reality, it becomes merely an "information event" from which to form revised Bayesian beliefs, upon which to rest current decisions.

When viewed this way, the quality of the information is not a matter of whether it is an accurate measurement of something "true", not that we would ever know for a quantity like accounting earnings. Rather, it is just one signal or input among others on which to update probability beliefs. Its "accuracy" is never known or possibly even of interest, apart from being reflected vaguely in the accuracy of the user's probability beliefs, as assessed after the event, or in the realized profits of the investor's actions based on those beliefs.

> The information perspective, the notion that accounting is designed to provide information, views accounting as using the language and algebra of valuation but for the purpose of conveying information. The distinction is subtle but profound. . . . The information content school . . . views

[10] See Mattessich (2006) for a fascinating history and overview of the development of accounting information theory, including Bayesian decision analysis. He notes also the overlooked quality of the books in German by Ewart and Wagenhofer, which are equivalents to Christensen and Feltham (2003). Further history is provided by Verrecchia (1982).

> the financial measures as measures of information events, not value. (Christensen and Demski, 2003, pp. 4–5)

Under such an instrumentalist decision–theoretic model of financial accounting "numbers", normative accounting theory virtually disappears. The only normative framework is the rational Bayesian decision model, and obeyance of the laws of probability, assumed of the user.[11] The input information can be a valuation, cost, depreciated cost or whatever most assists the Bayesian user to reach accurate posterior beliefs, from which to make hopefully profitable investment decisions. A priori directions, or any pragmatic thoughts, about how best to value assets or match costs and revenues, or any other traditional accounting debate, are not disallowed or irrelevant. They are encouraged and later evaluated where possible by whether the resulting information assists Bayesian probability revision and ensuing investment outcomes.

My own first introduction to Bayesian thinking in an accounting textbook was via George Foster's (1978) seminal textbook on financial statement analysis. Its early chapter on Bayesian inference and the value of imperfect information was written to set the grounds for all that follows to do with markets and investors rationally construing information about the firm's uncertain business prospects from the limited disclosures in its financial statements. I see this chapter as a carryover of the influence on accounting academics and PhD programs of the early Bayesians in US Business Schools.

2.4 Postscript

The strict theoretical interpretation of Neyman–Pearson methods is most easily understood when directed at confidence intervals, which are easier to pin down than hypothesis tests. According to Neyman, a 95% confidence interval is not an interval that can be considered to be true with probability or degree of belief 0.95. Instead, very long windedly, it

[11]The "rationality" generally presumed of Bayesian inference and decision traces more deeply to the axioms of logically coherent choice and probability beliefs. Such axiomatizations trace to at least Ramsey (1926) and de Finetti (1937/1964) and were made better known by Savage (1954) and DeGroot (1970). See Fishburn (1986).

is an interval generated by a mechanism that produces 95% confidence intervals that are indeed true 95% of the time (on the assumption of a true underlying model). Although that distinction between the interval and the mechanism generating it sounds obscure, and not very helpful, there is a reason for it. The interval in question, perhaps because it is based on an apparently "unrepresentative" but still strictly random sample, may be clearly unreliable and not worth 95% confidence in any natural sense. That would be the case when the random sample in question happens by bad luck to include a clearly "biased looking" subset of the population in question. To avoid this and the closely related issue of ancillary statistics,[12] Neyman focussed on producing intervals and test results that are justified solely in terms of long-run error rates in repeated applications of the same procedure. Despite much criticism, Neyman stuck tight to his rule that the single confidence interval or test result is just one of a long run, and is not important of itself or worth any special interpretation in its own single case. That roundabout and evasive interpretation has caused many a new student of statistics bemusement and "wrong" answers in exams, see Perezgonzalez, 2015 for a tutorial on interpreting hypothesis and significance tests. See Berger and Wolpert (1988), Berger (2003), Howson and Urbach (2005) and Johnstone (1987a) on the interpretation of Neyman–Pearson hypothesis tests.

[12]An ancillary statistic a is a sample statistic that of itself does not carry information about the unknown population parameter, but the test statistic is x has a different distribution under the unknown parameter θ when it is conditioned on the observed value of a; that is, $f(x|\theta, a)$ changes with a. The most obvious ancillary statistic is the sample size, but other more subtle examples occur and cause breakdowns in Neyman–Pearson statistical logic.

3

Survey of Bayesian Fundamentals

There is a vast literature in statistics, decision analysis, philosophy and applied fields on Bayesian logic and philosophy. This chapter introduces the language and philosophy of Bayesian methods, covering the main points of what makes "Bayesian". Concepts and principles are gathered from key references in Bayesian statistics and decision theory. Published overviews of Bayesian theory are largely of one mind, which can be attributed to the internal consistency, general applicability and intuitive appeal of Bayesian logic and method.

Source references are provided and categorized at the end of this chapter, for readers interested in the origins and a better explanation of the various "Bayesianisms" is discussed.

3.1 All probability is subjective

There has long been philosophical dispute about how to interpret probability. Is it a physical entity, is it a relative frequency in a reference set, does that set have to be existent or infinite, is it a deeply hidden logical relation between proposition and the evidence.[1]

[1] De Finetti famously in Bayesian theory said that "probability does not exist". The usual interpretation of this statement is that there is no such thing as a "true"

The elegant and workable Bayesian solution to this is to regard all probability as subjective. Some probabilities are merely "less subjective" than others in the sense that more people would have a roughly similar belief. For example, we might all believe that a die bought at the first shop we pass will have a probability of coming up "six" of about 1/6, but we will not agree as closely on whether company profits will increase.

3.2 Inference comes first

A Bayesian starts with an inference problem, even when ultimately the task is to make a decision or choice between actions (one of which might be inaction or "to wait"). She[2] has an "unknown" in mind, like a population parameter, or an outcome like whether it will rain today. She wants to "know" (i.e. judge subjectively) its probability distribution.

If the unknown in question is θ, she forms a probability distribution $f(\theta|x)$ based on observation x. In fact, she forms $f(\theta|x, BK)$ using both x and all her relevant background knowledge BK. From now on, BK is presumed but generally suppressed in the notation, because it is always implicitly there. It is essential because it includes information concerning the source or reliability of x.

A distinctively Bayesian step is to treat population parameters as uncertain, "stochastic" or "random" quantities, rather than as fixed "deterministic" unknowns, and to make parameters the subject rather than always the condition in probability statements.

3.3 Bayesian learning

To form a new belief or "learn" from x, Bayes theorem updates a prior distribution $f(\theta)$ to a posterior distribution

$$f(\theta|x) \propto f(\theta)f(x|\theta),$$

or "physical" probability. Think of the probability of red on the spin of a wheel. The wheel has physical properties like its size and weight, but the probability of it coming up red does not exist as a physical attribute.

[2]Bayesian inference is "personal" and it helps in exposition to use the personal pronoun "she" or "he", or sometimes "you" or "we".

or, more specifically,

$$f(\theta|x) = \frac{f(\theta)f(x|\theta)}{\int f(\theta)f(x|\theta)d\theta}.$$

That calculation is usually not the hardest part of the inference. Instead, the hardest part is specifying a subjective "likelihood function" for x

$$f(x|\theta),$$

which is a statistical description of how x is affected by θ.

3.4 No objective priors

Attempts to let the data "speak for themselves" or to be "objective" suggest that inference should start from a prior distribution that is uniform over the whole parameter space. A uniform prior is meant to represent ignorance, but in fact it does not. If the parameter in question is θ, the uniform prior over θ implies that $1/\theta$ is almost certainly close to zero, because that holds for most of the possible θ values on the real line. If the parameter is a Bernoulli probability, the uniform prior says that $E[\theta] = 0.5$ and also that values of θ above and below are equally likely. And then there is the issue of whether we are really interested in θ per se or more interested practically in say $1/\theta$ or $\sqrt{\theta}$. If the latter, then the "objective" prior might be uniform over $1/\theta$ or say $\sqrt{\theta}$ which entails a different belief about θ. So no objectivity is achieved. The uniform prior is justified subjectively when that's what's believed, but not as an "objective" starting point.

3.5 Independence is subjective

Outcomes A and B are independent if $p(A|B) = p(A)$ and $p(B|A) = p(B)$. Like everything else in Bayesian probability, independence is a subjective judgment rather than a physical fact. Obviously when making that judgment, there will often be consideration of apparent physical or causal connections between A and B, but ultimately that connection will exist only in the decision maker's subjective judgment, and under current background knowledge BK.

Similarly, A is conditionally independent from B, given C, if $p(A|B,C) = p(A|C)$. That says merely that we believe that, when C

holds, the probability of A is not influenced by B. That is a judgment under BK, and may prove wrong with further knowledge.

3.6 No distinction between risk and uncertainty

No distinction occurs between risk and uncertainty. Rather, all probability distributions are treated as subjective, to a greater or lesser degree, even when they appear to have an objective basis in nature (e.g. spins of a wheel) or in empirical evidence (e.g. the frequency of bankruptcy). Part of the reason is that even in the case of physical apparatus, the physics can still be wrong or misapplied (e.g. a mechanical or electronic casino game may have a bug or have been tampered with), and even if physical probabilities truly "exist" they are unobservable and can only be assessed by subjective (human) means. Similarly, observed frequencies must be conditioned on subjectively relevant factors (e.g. in the case of solvency — industry, inventory levels, debt levels, etc.) and are subject also to sampling error.

There can be no objective probability distribution for something like a cash dividend, which is not to say of course that, in forming a probability distribution subjectively, a decision maker cannot absorb all "objective" looking sample evidence; e.g. recent empirical estimates of distributional shapes and parameters. However, when making empirical considerations that seem objective, there are usually many embedded subjective inputs (e.g. selection of a sample period that seems best representative of the future). The Bayesian position is that subjectivity is inevitable and is best treated when made visible and used to advantage by integrating them into the model.

3.7 The likelihood function (i.e. model)

The likelihood function $f(x|\theta)$, or really $f(x|\theta, BK)$,[3] is "the model". It describes the probability process by which x is understood to have

[3] BK is an unwritten condition in all Bayesian probabilities. Any valid probability expression remains valid when the same condition is added to every term in that expression.

occurred. For example, x might be understood to have a normal distribution with mean μ and variance σ^2, in which case the likelihood function is

$$f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right].$$

In Bayesian inference, if any of the parameters is unknown, it will also be given a probability distribution, which often involves one or more other parameters, and on it goes. That is the start of what is called a hierarchical model and is why Bayesian computation often requires numerical methods like Monte Carlo simulation of posterior distributions.

3.8 Sufficiency and the likelihood principle

It follows directly from Bayes theorem that evidence x is summarized sufficiently by its likelihood function $f(x|\theta)$. To a Bayesian, the actual observed x is irrelevant once we know $f(x|\theta)$ for that observed x. Put another way, the "likelihood principle" holds that two observations x_1 and x_2 carry exactly the same evidence about θ if they have proportional likelihood functions

$$f(x_1|\theta) = c\ f(x_2|\theta). \quad (c > 0)$$

Information x might be reported in the form of a summary or statistic $g = g(x)$. That statistic is sometimes a sufficient summary of x for inference about θ. By expansion

$$f(x|\theta) = f(g|\theta) f(x|g, \theta).$$

But the distribution of x given g might not depend on θ, implying

$$f(x|\theta) = f(g|\theta) f(x|g),$$

in which case $g = g(x)$ is a sufficient statistic for x with respect to θ, in the sense that the posterior distribution of θ is the same given either x or $g(x)$.

The proof is as follows:

$$\begin{aligned} f(\theta|x) &\propto f(\theta)f(x|\theta) \\ &= f(\theta)f(g|\theta)f(x|g) \\ &\propto f(\theta)f(g|\theta) \\ &\propto f(\theta|g). \end{aligned}$$

The idea of a sufficient summary is that no evidence in x about θ is lost, so inferences are the same whether made from x or from its summarized expression g.

3.9 Coherence

A Bayesian can "have" whatever probability beliefs she feels — on one condition. That condition is "coherence" and requires simply that the agent's probability beliefs are mutually consistent, they must relate to each other correctly in the laws of probability.

The basic laws are the addition law, the multiplication law and Bayes law, which itself is easily obtained from the multiplication law. The only other basic law, which follows from the others, is the law of complete probability. It says simply that the probability distribution of x is its weighted average distribution across all possible states θ

$$f(x) = \int f(\theta)f(x|\theta)d\theta.$$

Part of the essence of Bayesianism is that the unknowns include both parameters (e.g. a population mean) and variables (e.g. a sample observation). So it makes no difference whether θ or x is the parameter or the observation, and hence we can equally write

$$f(\theta) = \int f(x)f(\theta|x)dx,$$

simply interchanging the two in the expression above. Both expressions hold, and coherence means that results found from one can be substituted into the other and cross-checked. In a coherent Bayesian calculation, everything ties together. See for example the inter-related calculations in the later illustration of Simpson's paradox.

3.10 Coherent means no "Dutch book"

Coherence is not merely for personal internal consistency or for some normative ideal. The practical reason for coherence is that coherent probabilities cannot be subjected to a Dutch Book.

Imagine a bookmaker who sets prices or probabilities that add up to less than one. If the gambler bets cleverly across all the horses in the race, she can make a book (portfolio) of bets by which she will necessarily win, whichever horse wins. Similarly, if the bookmaker sets prices (probabilities) that add to more than one, then by selecting which bets he will accept and which he won't, he can bet against the market so as to be sure to make a profit overall, whichever horse wins.

A book or set of bets that is guaranteed to win is a Dutch book, and if you are not coherent, you can have a Dutch book made against you.

The Bayesian principle of "no Dutch book", or coherence, is essentially also the principle of no arbitrage. Like coherent probabilities, asset prices that prevent arbitrage need not be "accurate" in the sense that they approximate any fundamental value or truth, since accuracy is not required for coherence. Coherence is a necessary but not sufficient condition for avoiding losses when betting or investing.[4]

3.11 Coherent is not necessarily accurate

Coherent beliefs can be inaccurate and even ridiculous, their coherence makes them merely mutually consistent, and not open to a Dutch book (arbitrage). The ideal of Bayesian inference is to reach probabilities that are both coherent and accurate.

3.12 Accuracy is relative

A common but misplaced criticism of Bayesian thinking occurs when the weather forecaster reports that the probability of rain tomorrow is 0.5.

[4]The notion of coherence is the linchpin of Bayesian theory and the theory of economic rationality. The idea was developed by Ramsey (1926) and taken up by Savage (1954). Rigorous further development came with DeGroot (1970).

That sounds on the surface like a coin toss and the height of vacuousness or ignorance. A Bayesian perspective is much more forgiving. If rain at this time of year, and in these conditions, is subjectively unlikely, then it is quite a statement to predict rain with 50% probability. Moreover, if in fact it rains, that probability will have proven very accurate, at least relative to a model or forecaster who gave rain say a 5% chance. And if it does not rain, that single statement will be proven relatively inaccurate.

Similarly, a model that produces bankruptcy probabilities might be relatively accurate even when it gives firms that went bankrupt only small probabilities of failing. If the average rate of bankruptcy is say 5% per year, then a probability of even 10% attached to a bankrupt firm represents a more accurate probability than the average.

3.13 Odds form of Bayes theorem

Emphasizing the role of the likelihood function in Bayesian learning, Bayes theorem is sometimes written in odds form. The odds expression of a probability p is $p/(1-p)$ or "p to $(1-p)$ in favor". Let θ belong to a parameter space Θ and suppose that the problem is to assess the probability of $H_0 : \theta \in \Theta_0$ where the alternative is $H_1 : \theta \in \Theta_1$, and $\Theta = \Theta_0 \cup \Theta_1$. By Bayes theorem,

$$\frac{p(\Theta_0|x)}{p(\Theta_1|x)} = \frac{p(\Theta_0)}{p(\Theta_1)} \frac{p(x|\Theta_0)}{p(x|\Theta_1)},$$

which can be summarized as

$$\text{posterior odds} = \text{prior odds } \times \text{ likelihood ratio.}$$

Thus, information x alters the Bayesian's odds only through its likelihood ratio, here $\frac{p(x|\Theta_0)}{p(x|\Theta_1)}$.

The log odds form of Bayes theorem is also insightful, showing how the statistical information content in the prior and the likelihood ratio can be understood as additive

$$\log \frac{p(\Theta_0|x)}{p(\Theta_1|x)} = \log \frac{p(\Theta_0)}{p(\Theta_1)} + \log \frac{p(x|\Theta_0)}{p(x|\Theta_1)}.$$

3.14 Data can't speak for itself

Bayes theorem says that if the task is to infer the probability of a proposition of general form $\theta \in \Theta_0$, then the evidence carried by x favors $\theta \in \Theta_0$ over its complement $\theta \in \Theta_1$ if and only if the likelihood ratio

$$\frac{f(x|\Theta_0)}{f(x|\Theta_1)} > 1.$$

Evidence x supports, albeit does not prove, whichever proposition or hypothesis gives it greater probability of having occurred. There is always a subjectively chosen model and some temporal background knowledge lurking within the likelihood function, $f(x|\theta) \equiv f(x|\theta, BK)$. That function could equally be written as $f(x|\theta, model \cap BK)$ so as to emphasize that it is dependent on a given model. The mere fact that we see x as depending on the parameter θ shows that we have a some model in mind.

Hence, data does not have any particular Bayesian meaning without a model and without conditioning its interpretation on any relevant related evidence and background knowledge. Thus, data does not easily "speak for itself".

3.15 Ancillary information

Sometimes to find the likelihood $f(x|\theta)$ we know only $f(x|\theta, y)$, but y of itself is uninformative about θ. Being uninformative means that $f(y|\theta) = f(y)$, that is, the distribution of y is unaffected by θ so the observed of y says nothing about θ. So by Bayes theorem,

$$f(\theta|x) = \frac{f(\theta)f(x|\theta, y)}{\int f(\theta)f(x|\theta, y)d\theta},$$

and thus there is no difficulty caused by knowing only $f(x|\theta, y)$. The information brought by y is used in the calculation, but it is "ancillary". Sometimes y is the "source" of x, like a set of experimental conditions or a state of nature, and implies nothing about θ of itself, but does make a difference to the probability of x given θ.

3.16 Nuisance parameters "integrate out"

When modelling the effect of population parameter θ on observation x, is typical that another parameter α plays a role in producing x. The likelihood function is then $f(x|\theta, \alpha)$ and by Bayes theorem

$$f(\theta, \alpha|x) \propto f(\theta, \alpha)f(x|\theta, \alpha).$$

To reach an inference about θ unconditional on α, or, in other words, averaged over α, the "nuisance parameter" α is "integrated out",

$$f(\theta|x) = \int f(\theta, \alpha|x)d\alpha.$$

Illustrating the internal consistency of Bayesian probability calculations, α could equally be integrated out of the likelihood function $f(x|\theta, \alpha)$, giving

$$f(x|\theta) = \int f(x|\theta, \alpha)f(\alpha|\theta)d\alpha.$$

The Bayesian method of integrating out nuisance parameters is essential in any inference problem where interest concerns the (average) effect of just one particular parameter, or say a particular pair of parameters, within a model containing many parameters. Typically, for example, inference might focus on an unknown population mean, but in the context of an equally unknown population variance.

3.17 "Randomness" is subjective

Since probability is not physically existent, neither is randomness. Randomness, like probability, is a subjective assessment, and of course will be more easily made in some situations than others.

The Bayesian term equivalent to "random" in conventional statistics is "exchangeable". Rather than saying that a sample observation is a random draw from a population, the Bayesian term is that it is exchangeable with any other element of that population. In effect, that is like saying that it is viewed as iid with any other member of the population. Tosses of a given coin are typically viewed as exchangeable with each other, but the events of whether two different stocks go up or down are not. The difference is that we typically regard the coin as having a fixed Bernoulli parameter θ, and regard every toss as iid given θ. Stocks on the other hand are not usually regarded as iid all with the same θ.

According to the Bayesian subjective notion of exchangeability, an observation or element of a population does not need to be drawn using a physical randomizer to be regarded as having the same probability distribution as other elements of the sequence or population, or to be viewed as validly “representative” of that population. Rather, exchangeability requires only a subjective judgment that the observation in question is “unbiased” or “representative”, or not distributed in a way other than the rest of the population.

Two main points follow, both of which depart substantially from much statistical convention. First, if the observation in question does not seem exchangeable or representative, then it is no help that it was drawn with a randomizer.

Second, a subjectively representative looking observation is exchangeable however it was drawn. For example, the proceeds from a bet on number 13 in roulette are subjectively representative of any other bet on the roulette wheel, regardless of whether number 13 was chosen using a randomizer (such as by spinning another roulette wheel) or perhaps because it is the gambler’s lucky number. This is not saying that every number on the roulette wheel has the same physical probability of occurring, rather it says that in usual circumstances we view the wheel that way, and have no evidence suggesting that the probabilities are any different. Hence, a random bet on one segment of the wheel is exchangeable in our judgment with a bet on any equally large segment.

In Bayesian theory, randomization is neither sufficient nor necessary as the basis for probability calculations. But nor is it harmful of itself. Indeed, some Bayesians see randomizers as a good way to ease the difficulty of the main task, namely the making of a judgment of exchangeability. A randomizer takes away the sampler’s discretion, which might otherwise have introduced a hidden bias.

Remember again, all Bayesian probabilities are conditioned on what background knowledge BK is known, they are not claimed as “true” or necessarily even “close”, they are merely coherent under and with current BK.

Interestingly, this Bayesian position ratifies much of what happens in empirical research using frequentist statistics. Samples are not in fact often drawn with physical randomizers, but they are usually considered for whether the seem representative. Another related point is to do with “ambiguity aversion”, which is not a Bayesian notion. People are reluctant to bet on a binary event when they don’t “know the probabilities”, but in that case bets on “A” or “B” are exchangeable, in the same way as if we knew for sure that they each had probability of 0.5 (which we never do).

3.18 “Exchangeable” samples

A Bayesian spinning a well-engineered red and black roulette wheel would see a sample {R, R, R, B, B, B} as exchangeable with {R, B, B, R, B, R}. When trying to infer the probable value of $\theta = p(\text{Black})$, the Bayesian would not care about the order of the Reds and Blacks, because that ordering, and any apparent pattern, is viewed under BK as an irrelevant chance outcome from a wheel with “no memory”.

Consider a process that switches unobservably between two regimes, like two different “wheels” with unequal parameters θ_1 and θ_2. If we judge that the two regimes or wheels have probabilities $p(\theta_1) = p$ and $p(\theta_2) = 1 - p$, then any sample of size n is exchangeable with any other sample of size n because every observation is, on what (little) we know, an independent draw from a process with average parameter $\theta = p\ \theta_1 + (1 - p)\ \theta_2$. That assessment holds because we have no indication of which wheel our realized observations were drawn from. In that case, if we want to infer the value of θ, every sample of size n is “the same” and exchangeable in our eyes, even though we know that in any given sample there might be more draws under θ_1 or more under θ_2.

Even though the individual observations in the realized sample are not known to have come from the two wheels in the given mixture proportions p and $1 - p$, the sample is nonetheless subjectively exchangeable with any other, and remains so until we learn otherwise (as would happen if it was revealed that one particular sample came from a different mixture of θ_1 and θ_2).

To see the practical importance of exchangeability, consider the following inference problem. Suppose that we use the words occurring in text as a way to infer who wrote them. If the "bag of words" in the sample is {economy, growth, arbitrage, ..., asset} we might well see that sample as exchangeable with another sample with common business words, like {company, tax, bond ..., invest}. If so, then we are saying that we would be equally happy to use either sample, and also that we would be happy to merge the two samples much like tosses of the same coin. But if the sample contained {business, multinationale, client, ..., vendeur}, we would not believe that the two samples are reflective of the same population/author.

Note that the fact that two samples are "iid" in the frequentist sense does not make them exchangeable subjectively. The judgment that they are exchangeable requires a readiness to see them as representative of the same population. A sample drawn using a randomizer and thus "iid" from all sales transactions in August might contain sales of mainly one product and hence not be exchangeable with another sample of August sales that contains a more representative looking sample of sales of all the firms' products.

3.19 The Bayes factor

Bayes theorem in the case of hypothesis $H_0 : \theta \in \Theta_0$ versus complement $H_1 : \theta \in \Theta_1$ can be written as,

$$
\begin{aligned}
f(\Theta_0|x) &= \frac{f(\Theta_0)\ f(x|\Theta_0)}{f(\Theta_0)\ f(x|\Theta_0) + f(\Theta_1)\ f(x|\Theta_1)} \\
&= \left[1 + \frac{f(\Theta_1)}{f(\Theta_0)}\ \frac{f(x|\Theta_1)}{f(x|\Theta_0)}\right]^{-1} \\
&= \left[1 + \frac{f(\Theta_1)}{f(\Theta_0)}\ \frac{1}{B}\right]^{-1},
\end{aligned}
$$

where

$$
B = \frac{f(x|\Theta_0)}{f(x|\Theta_1)} = \frac{\int_{\theta\in\Theta_0} f(\theta) f(x|\theta) d\theta}{\int_{\theta\in\Theta_1} f(\theta) f(x|\theta) d\theta}
$$

is the "Bayes factor" or "weighted likelihood ratio" of the evidence x supporting Θ_0.

Evidence x enters calculations only via the "Bayes factor" B, which is the ratio of the average probability of x under hypothesis $\theta \in \Theta_0$ to the average probability of x under the alternative $\theta \in \Theta_1$. Bayes factors are widely suggested by Bayesians as a coherent alternative to conventional measures of evidence, particularly p-values.

3.20 Conditioning on all evidence

Bayesian coherence requires that all probability assessments be made under all of the evidence. The principle of conditioning on all available information is sometimes stated as "no relevant subsets", and amounts to the same requirement as "exchangeability". A relevant subset occurs when the observation can be described more fully, in a way that changes its subjective probability. For example, $x \geq x_c$ is one description of $x = x_c$, but the more specific description narrows the reference set of observations within which it is viewed as exchangeable, and gives the observation a different subjective probability based on a new but still not necessarily "true" reference set.

The rule of no relevant subsets has a long history in the logic of inductive inference, and is formalized mathematically in Bayesian statistics by the requirement of exchangeability. "Exchangeability" and "no relevant subsets" are essentially the same requirement and are both subjective assessments. For example, an insurance company might see a client as exchangeable in risk of motor accident with other 40–50-year-old males living in the same city, but if he seems to belong to an apparently relevant subset, like those males with good driving records, the subjective probability of accident will change and so of course might the insurance premium (see later analogy with the cost of capital).

3.21 Bayesian versus conventional inference

To sum up Bayesian theory, it helps to identify the main ways that Bayesian logic departs from "conventional" statistical thinking.

The debate between Bayesian and "frequentist" statistical methods is decades old and has lately, especially in disciplines like computer science, genetics and forensic science, swung the Bayesian way. Interestingly, economic decision theory (à la Savage) and subfields like noisy rational expectations models of information disclosure have from their onset been Bayesian, albeit not by any ideological choice other than the economic assumption that individuals act rationally in their own self-interest — and thus internally consistently or "coherently". Bayesian inference is seen by many statistical theorists to be based on a more defensible logic than frequentist methods. That conclusion stems from simple methodological comparisons like those set out below.

Likelihood ratios, not p-levels

The p-level of observation x_{obs} is by definition the probability $\Pr(x \geq x_{obs}|H_0)$ of sample x falling in tail-area $x \geq x_{obs}$, conditional on hypothesis H_0, where the symbol $\geq$ implies "as or more discrepant with". That probability cannot be interpreted in terms of evidence under Bayes theorem without knowing the probability of the same result under the alternative hypothesis. If the associated likelihood ratio,

$$\frac{p(x \geq x_{obs}|H_0)}{p(x \geq x_{obs}|H_1)},$$

is less than one, then the observation summarized as $x > x_c$ supports H_0 rather than H_1, even when its p-value or p-level, $p(x \geq x_c|H_0)$, is arbitrarily low. According to Bayes theorem, therefore, the p-level of observation $x = x_{obs}$ cannot be interpreted without knowing the other half of the relevant likelihood ratio,

$$p(x \geq x_{obs}|H_1).$$

Conditional probabilities, not long-run frequencies

An application of the unconditional long-run frequency approach happens when a randomized mixture of two hypothesis tests, each with its own conditional error frequencies $\{\alpha, \beta\}$ has apparently "better" error frequencies than either of the two tests individually. For example, suppose that the randomizer is a coin toss and the tests have

$\{\alpha = 0.05, \beta = 0.1\}$ and $\{\alpha = 0.005, \beta = 0.4\}$. The mixture test then has $\{\alpha = 0.0275, \beta = 0.25\}$, which for some users or loss functions is a better compromise between α and β than either constituent single test.

To claim those preferred error frequencies, the user must intentionally ignore the fact that the test selected by the randomizer, and thus actually run, is known. If it's not known which test was run, a Bayesian would use the mixture probabilities too, and that is why it is often suggested that frequentist statistics works best logically when some information is ignored.

The conditional versus long-run frequency distinction comes to light in many contexts. Suppose that a random sample is drawn, correctly randomly by spinning a wheel, and yet the resulting sample turns out to be "not a typical looking random sample". From a conditional perspective, some post-sample reconsideration is necessary, but from a long-run frequency approach there is no issue, because in the long run all such biased samples average each other out. In fact, all such samples must be included in the reference set so as not to introduce a bias that changes those long-run error frequencies.

The two famous frequentist statisticians, Fisher and Neyman, and their schools, fell out over this very issue. In essence, Fisher adopted a Bayesian-like conditional single-case approach and Neyman took refuge in long-run frequencies under repeated applications.

Likelihood function versus error frequencies

We wish to compare two possible test signals x_1 and x_2, as would occur when evaluating their ex ante values of information. Each test can return a signal $x_i \in \{U, F, N\}$, indicating unfavorable, favorable or neutral, and the state of nature is $V = 0$ or $V = 1$.

$x_1 \rightarrow$	U	F	N
$\Pr(x_1 \mid V = 0)$	0.9	0.05	0.05
$\Pr(x_1 \mid V = 1)$	0.09	0.055	0.855

$x_2 \rightarrow$	U	F	N
$\Pr(x_2 \mid V = 0)$	0.26	0.73	0.01
$\Pr(x_2 \mid V = 1)$	0.026	0.803	0.171

The Bayesian view of the two signals is that whatever the signal observed, U, F or N, the likelihood functions of the two test signals are proportional, and hence each leaves the same posterior probability.

For example, if the signal x_1 is F, the likelihood ratio is

$$\frac{p(x_1 = F|V = 0)}{p(x_1 = F|V = 1)} = \frac{0.05}{0.055} = 0.90909,$$

and the corresponding likelihood ratio of signal $x_2 = F$ is the same

$$\frac{p(x_2 = F|V = 0)}{p(x_2 = F|V = 1)} = \frac{0.73}{0.83} = 0.90909.$$

It does not matter which signal is observed, the evidence is the same under both tests, and hence the two test signals are of equal value ex ante and ex post.

That conclusion does not agree with a frequentist interpretation. The frequentist interpretation has both null and alternative hypotheses, say

$$H_0 : V = 0 \text{ and } H_1 : V = 1,$$

and error probabilities

$$\alpha = p(\text{reject } H_0|H_0) \text{ and } \beta = p(\text{accept } H_0|H_1).$$

The error characteristics of a test depend on its "rejection region" which is discretionary. Suppose that we regard $x \in F \cup N$ as favoring the higher value $V = 1$. If x_i falls in that region, we reject $H_0 : V = 0$. On that test design, signal x_1 has error probabilities

$$\alpha_1 = p(x_1 \in F \cup N|V = 0) = 0.05 + 0.05 = 0.1$$

and

$$\beta_1 = p(x_1 \in U|V = 1) = 0.09.$$

These seem like "good" error probabilities, which raises the question of whether the second signal, which has the same likelihood function, is equally good?

To a frequentist, the answer is no, because its error probabilities are easily seen to be $\alpha_2 = 0.74$ and $\beta_2 = 0.026$. The Bayesian and frequentist depictions of the same two tests are therefore contradictory.

To seek some remedy, perhaps, it is possible to change the rejection region of the test, to reject $V = 0$ only if $x = F$. The two tests error probabilities are then $\{\alpha_1 = 0.05, \beta_1 = 0.945\}$ and $\{\alpha_2 = 0.73, \beta = 0.197\}$ respectively, which puts the "same test" in a yet another frequentist light but does not solve the problem that the Bayesian interpretation views the two tests as intrinsically identical in every way.

Stopping rules

Let θ be the probability of "heads" in a coin toss, and suppose that we use data from actual tosses to test whether the coin is fair $H_0 : \theta = 0.5$ against the alternative that it is biased upwards $H_1 : \theta > 0.5$. Two experiments are run. The first takes pre-fixed $n = 12$ observations and produces 9 heads and 3 tails. The second draws randomly until 3 heads are observed, and coincidentally produces 3 heads also in 12 tosses. Is there any difference in the evidence coming from these two tests?

This is a well-known problem, and is revealing because the Bayesian and frequentist answers are so at odds. A quick summary goes as follows.[5]

Take the Bayesian approach first. In the first test, with predesignated $n = 12$, the likelihood function is the usual binomial probability

$$f_1(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} = 220 \times \theta^9 (1-\theta)^3.$$

In the second test, which requires sampling until exactly 3 tails, which happens to occur at $n = 12$ tosses, the likelihood function is found using the negative binomial distribution, and is given as

$$f_2(x|\theta) = \binom{n+x-1}{x} \theta^x (1-\theta)^n = 55 \times \theta^9 (1-\theta)^3.$$

So it follows Bayesianly that the two test outcomes offer the same evidence exactly, because their respective likelihood functions are proportional. The evidence is simply 9 heads in 12 tosses, regardless of why sampling stopped at that point.

The frequentist approach is more complicated. Its p-level of $x = 9$ in the first test is

$$p\text{-level}_1 = p(x \geq 9|\theta = 0.5, n = 12) = 0.075.$$

Similarly, the p-level of $x = 9$ in the second test is

$$\begin{aligned} p\text{-level}_2 &= p(x \geq 9|\theta = 0.5) \\ &= f_2(9|\theta = 0.5) + f_2(10|\theta = 0.5) + \cdots = 0.0325. \end{aligned}$$

[5] See Lindley and Phillips (1976) for the original exposition, and especially the Berger (1985) explanation.

Note that the numbers were chosen in this example so as to show how the same observation, $x = 9$ in $n = 12$, can be either "significant at 5%" or "not significant", depending only on how sampling came to stop.

Illustrations like this one are used to show how frequentist methods interpret the very same data differently depending on how they were drawn, which raises the disconcerting possibility that an experimenter seeking a given p-level might fudge, not by faking the data, but by faking the stopping rule. Other similar criticisms are that data should not be interpreted on the basis of the experimenter's unobservable intentions, and how can data be interpreted if no one knows or can remember why sampling stopped?[6]

A good illustration by which to understand the Bayesian position is to think of red and black balls dropping apparently "randomly" from a bucket. The Bayesian watches this process and sees a flow of perfectly acceptable exchangeable ("iid") observations, and hence evidence accumulating. Each draw is valuable and is merely one more observation. But suppose that a frequentist, who has been watching and calculating a p-level, demands a halt in the sampling at the point of say N observations. The Bayesian sees the result as merely N exchangeable observations falling from the bucket, no different to any other draw of N observations. The frequentist might have obtained the p-level desired and wanted a stop, but the Bayesian is interested only in the posterior probability from what evidence exists, quietly wishing for a few more observations, as might have been obtained had the frequentist used a different stopping rule (e.g. a lower desired level of significance).

The frequentist has managed to manufacture the desired p-level but had no influence over the observations themselves. They were affected only by random sampling and the true proportions of red and black balls in the bucket, neither having been tampered with, thus leaving N untainted observations from the Bayesian viewpoint.

[6]Criticisms like this are obviously highly damaging or at least contentious, and help explain why there was once much animosity between some Bayes and non-Bayes practitioners.

Lindley's paradox

Consider a simple test of $H_0 : \mu = 0$ versus $H_1 : \mu = 1$, where the population is $X \sim N(\mu, \sigma = 1)$. The sample size is n and the pre-test perspective is that higher n is preferred because a test of given size α (say $\alpha = 0.05$) has greater power when n is higher, and therefore has greater chance of rejecting H_0 if indeed it is wrong. The test is two-sided and has fixed $\alpha = 0.05$, so the critical level of the sample mean is $\overline{x}_c$ such that $f(|\overline{x}| \geq |\overline{x}_c|\,|H_0) = 0.05$.

Now suppose that the test observation $\overline{x}_{obs}$ is marginally significant at 5%, so $\overline{x}_{obs} = \overline{x}_c$. Hence, in each case the null hypothesis is rejected at $\alpha = 5\%$, in frequentist terms. The question that arises is how to interpret that rejection of H_0 with regard to the sample size.

One answer is that the sample size makes no difference to the meaning of the result, because it is accounted for in the calculations already. That answer does not hold up well, however, once we look at the orthodox frequentist confidence interval that corresponds to "rejection at 5%" (or *p-level* = 5%) with given n. The confidence intervals shown in Table 3.1 below become tighter and "nearer in mass" to $\mu = 0$ as n increases, suggesting intuitively that the evidence is less and less supportive of values of μ well away from the null value of $\mu = 0$.[7]

Hence, it can be argued that if we want to truly discredit or "reject" H_0, we might prefer ex post (after the test) to have run a test and obtained a rejection with lower power, not higher power. Rejection of H_0 is made more likely ex ante with higher n, but "means less" evidentially ex post with higher n.[8]

[7]Frequentist confidence intervals often show evidence more revealingly than *p*-levels, but unlike *p*-levels are not widely reported in empirical research. Confidence intervals reveal more of the size or "economic significance", or otherwise, of the effect observed. Critics of "statistical significance" as an evidential standard have often advocated a shift to reporting observed confidence intervals rather than significance levels.

[8]The distinction between pre-test and post-test interpretations of the same test was made by Savage (1962) under the heading "initial versus final precision", and is explained in principle and by examples in Berger (1985). A short explanation is that while a random sample might promise to be informative ex ante, the drawn sample x might have an observed characteristic or sample statistic $a(x)$ that makes it different from other mechanically "random" samples (of the same size) in the sense that knowing its property $a(x)$ changes its likelihood function $f(x|\theta, \cdot)$. See the discussion on ancillary statistics.

Table 3.1: Frequentist confidence intervals conditional on fixed *p-level* = 0.05.

n	$\overline{x}_{obs} = \overline{x}_c$	Power $f(\lvert\overline{x}\rvert \geq \lvert\overline{x}_c\rvert \,\lvert H_1)$	95% CI	$f(\overline{x}_{obs}\lvert H_0)/$ $f(\overline{x}_{obs}\lvert H_1)$	$\Pr(H_0\lvert\overline{x}_{obs}, n)$
1	1.96	0.1685	(0, 3.92)	0.232	0.188
5	0.8765	0.6087	(0, 1.75)	0.152	0.132
10	0.6198	0.8854	(0, 1.24)	0.302	0.232
20	0.4383	0.9940	(0, 0.88)	3.44	0.775
50	0.2772	1	(0, 0.55)	68920	1

This disturbing result is usually known as Lindley's paradox, and was shown in Bayesian terms by Lindley, and others including Jeffreys, by comparing the achieved frequentist significance level of 5% with the corresponding Bayesian posterior probability of H_0, allowing for n.

In the simple test described, the likelihood ratio is the ratio of two normal densities, and simplifies to

$$\frac{f(\overline{x}_c|H_0)}{f(\overline{x}_c|H_1)} = \frac{\exp\left[-\frac{n}{2}\overline{x}_c^2\right]}{\exp\left[-\frac{n}{2}(\overline{x}_c - 1)^2\right]} = \exp\left[n\left(\frac{1}{2} - \overline{x}_c\right)\right].$$

Note how this ratio, which carries all the Bayesian evidence, depends on just the critical value $\overline{x}_c$ obtained and the sample size n, thus showing how the Bayesian interpretation does call on n, even though n was used to find $\overline{x}_c$.

The value of the likelihood ratio is shown in the table for each sample size n, along with the resulting posterior probability, $\Pr(H_0|\overline{x}_{obs})$, given $\Pr(H_0) = 0.5$. Observe how the posterior probability of the "rejected" null hypothesis tends to one when n becomes large enough.[9] That is the famous paradox. See Lindley (1957) and Berger and Sellke (1987) and Robert (2007) for a fuller explanation and some qualifications.[10]

This clash of statistical cultures is largely overcome when frequentists report confidence intervals rather than statements like "reject the null

[9] As explained by Lindley (1957, p. 189): "5% in to-day's small sample does not mean the same as 5% in to-morrow's large one".

[10] Harvey (2017) in his presidential address to the AFA emphasized Lindley's paradox as a fundamental issue for logical statistical inference.

hypothesis". In terms of confidence intervals, higher n is unambiguously desirable, because the interval is narrower for any given confidence level, $100(1-\alpha)\%$. That "solution" goes only part way towards a Bayesian test interpretation, because the resulting confidence interval remains to be interpreted, and Neyman did not allow confidence intervals any Bayes-like interpretation in terms of evidence or beliefs.

3.22 Simpson's paradox

Simpson's paradox is not really a paradox, but rather a natural occurrence when conditioning probabilities to allow for "relevant subsets". Consider the data below, which is taken from a very large sample and shows the recovery rate of patients when they are treated and when they are not.

Patient recovery rates

	Yes (R)	No (N)		Recovery rate
Treatment (T)	20	20	40	50%
Control (C)	16	24	40	40%
	36	44	80	

The data above suggests that the treatment works in the sense that a patient has a 50% chance of recovery if treated and only 40% if not treated. That is, $p(R|T) > p(R|C)$. To narrow down and possibly better identify why the treatment seems to work, it helps to break the data into two sub-populations;, Males (M) and Females (F).

	Yes (R)	No (N)		Recovery rate
Males				
Treatment (T)	18	12	30	60%
Control (C)	7	3	10	70%
	25	15	40	
Females				
Treatment (T)	2	8	10	20%
Control (C)	9	21	30	30%
	11	29	40	

The curious fact is that although treatment seems to work overall, raising the unconditional recovery rate from 40% to 50%, treatment does not seem to work for males or for females, both having lower recovery rates when treated.

The first point to make before explaining how such a paradoxical looking result can occur is that the Bayesian and commonsense perspective is to take the conditional probabilities as the relevant ones. For males, without treatment, 70% recover, and with treatment only 60% recover. And for females, 30% of those without treatment recover, but only 20% of those with treatment recover.

So why does treatment appear to work when there is no conditioning on gender? The answer is that the treatment group includes $30/40 = 75\%$ males, most (60–70%) of whom would have recovered whether treated or not, thus making treatment "look" relatively effective. The overall recovery rate under treatment is a weighted average

$$\begin{aligned} &p(\text{Male}|T)p(R|T,\text{Male}) + p(\text{Female}|T)p(R|T,\text{Female}) \\ &\quad = 0.75 \times 0.6 + 0.25 \times 0.2 \\ &\quad = 0.50. \end{aligned}$$

So the 60% recovery rate of the treated males biases the overall recovery rate of the population up to 50%, which of course hides the fact that treated males recover in lower proportion than untreated males.

In terms of exchangeability, males and females are usually seen as "relevant subsets" and hence not exchangeable. It is wrong therefore in Bayesian inference to apply the 50% overall population recovery rate to either males or to females specifically. Conditioning is required on gender, or so it clearly seems on what is known (the required level of conditioning is always knowledge dependent, and subjective).

Bayesian conclusions are sensitive to further conditioning and hence to relevant background knowledge. It may be that if we were to condition on not just gender but on some other covariate like age, a relevant sub-subset of say young healthy females might be found to respond very favorably to treatment.

The ability of a finer level of conditioning to reverse previous conclusions shows why randomization has a role, at least after stratification.

Suppose that we believe that males with some suspected albeit yet unidentified characteristic are very different in their response rates to treatment than females generally. By selecting the males in the experiment randomly, and allocating them to treatment and control groups randomly, the hope is to avoid having a hidden unrepresentatively large proportion of those "unresponsive" males in the treatment group. That approach holds even when we do not know which males might have that physical characteristic, nor even what it is. Similarly, of course for females.

It may always be that more knowledge removes what previously seemed to be a relevant covariate. For example, gender might be irrelevant to recovery rates once those rates are conditioned on some deeper human genetic characteristic, but until then gender defines a subjectively relevant subset. That is how probabilities change, up and down, taking certainty up and down.

3.23 Data swamps prior

Typical of the intuitively agreeable logic of Bayesian inference is the way that Bayes theorem reacts to new information. Ultimately, as the amount of observation or information increases, different Bayesians who have the same model but different prior beliefs about the values of the parameters in that model, all come to the same posterior beliefs. Enough information about a parameter can bring strong agreement, even across different models involving that same parameter.

The Bayesian process by which "the data swamps the prior" is a response to claims that Bayesianism is "too subjective", but it also shows the inbuilt Bayesian ability to "go backwards", in the sense that the arrival of more data will sometimes reduce the Bayesian prior certainty about a parameter's value, or even about the model itself, and can lead to new results quite contrary to prior beliefs.

Bayesian inference is open to new information changing beliefs in any way at all. Ideally, there is enough data to reach strong conclusions, but a large set of data might merely alter prior beliefs just enough to leave the user much less certain, and hence more in need of still more data.

3.24 Stable estimation

Robust conclusions are obviously assisted by more data, but another property of Bayesian inference that makes the job easier is "stable estimation". In brief, if the likelihood function is relatively peaked, then the posterior density has the same peak and conclusiveness, provided that the prior distribution is fairly flat in the region where the likelihood function is peaked. That follows from the fact that the posterior density is proportional to the product of the prior density and the likelihood. Hence, in any region where the likelihood density is very small, the prior density usually makes little difference to the posterior. Thus, the general shape and location of the prior mass outside the region where the likelihood function is peaked is largely immaterial.

More data remains, of course, desirable. In fact, the likelihood function will tend to be sharply peaked only with a sufficiently large sample. A robust posterior distribution is then a combination of the two effects.

3.25 Cromwell's rule

Bayesian methodology does not welcome any probability assessment of precisely 0 or 1, because utter certainty can never be undone. With prior probability of 0 or 1, no evidence or likelihood ratio is strong enough to change that belief. When there exists virtual certainty, the inference maker will invoke a probability of very slightly less than 1, so as to admit Bayesian learning but still have little effect on any conclusion or decision. That method follows Cromwell's advice to the Church of Scotland "to always allow that it might be wrong".

3.26 Decisions follow inference

Coherence is required of both inference and decision-making. The simplest depiction of rational decision-making assumes that there is a quantity or parameter θ (like sales units) and an action d (like inventory) which as a pair produce utility $u(d, \theta)$. Since θ can sometimes be affected by d, the probability distribution of θ is written as $f(\theta|d)$.

The decision maker's expected utility from action d is then

$$\int f(\theta|d)\ u(d, \theta)d\theta,$$

which is maximized by choice of d.

Design of experiments

Before taking any action d, the decision maker might design an experiment e to produce sample information x, so as to update her probability distribution for parameter θ. The expected utility from action action $d = d(x)$ is

$$\int f(\theta|d(x))\ u(d(x), \theta)d\theta,$$

which allows for the possibility that d affects θ. This expectation is maximized by setting $d = D$, where D is best written as $D(x)$ because it is the best action, given evidence x.

The evidence x comes from an experiment e. If the consequence of running experiment e and choosing action $D(x)$ is utility $u(D(x), \theta, e)$, then the decision maker's expected utility from experiment e given observation x is

$$\overline{u}(e, x) = \int f(\theta|e, x, D(x))\ u(e, D(x), \theta)d\theta.$$

Note that the utility $u(e, D(x), \theta)$ allows for the cost of running experiment e.

Experiment e is chosen so as to maximize expected utility, allowing for all possible experimental results x,

$$\overline{u}(e) = \int f(x|e)\ \overline{u}(e, x)dx.$$

In effect, we choose e that produces x that prompts the decision $D(x)$ that yields on average, according to ex ante beliefs, higher utility than any other e.

3.27 Inference, not estimation

Bayesian theory gives little direct importance to the frequentist properties of estimators. Even the fact that an estimator is unbiased is not always desirable of itself. Rather, the estimate with highest expected utility, when it is used to revise beliefs and make a decision, can sometimes be a very biased estimator in the frequentist sense.

Similarly, the role of maximum likelihood estimation is played down. In frequentist methods, the maximum likelihood estimator $\widehat{\theta}_{mle}$ has been shown to have many good frequentist properties, like being unbiased and efficient, but those properties of themselves are not appealing to a Bayesian. Rather, if the likelihood function $f(x|\theta)$ is quite flat, the maximum likelihood estimate $\widehat{\theta}$ of parameter θ is easily over-ruled by even relatively vague prior knowledge. If the decision maker were to act "as if" $\widehat{\theta}_{mle}$ is the true parameter, she will often take an action far from the action that a Bayesian would take on the same data x.

A good summary is that both Bayesian and frequentist statistical methods see the likelihood function $f(x|\theta)$ as the natural expression of the data in the context of the model, but the two frameworks make vastly different use of that function. Frequentists might use it to find a point estimate $\widehat{\theta}_{mle}$ but Bayesians use it to find its inverse, the posterior probability distribution $f(\theta|x)$.[11]

3.28 Calibration

A level of calibration is an ex post quality of a set or probability assessments relative to a set of outcomes. Probabilities are "well calibrated" if proportion p of all propositions given ex ante subjective probability p are found to be true, for all p. Calibration is emphasized in Bayesian literature as a desirable attribute, but is not required in the way that coherence is. When combined with high resolution, calibration implies highly "accurate" probabilities. Of itself calibration is not compelling. For example, if it rains on average on 10% of days, a forecaster can merely state $p = 0.1$ everyday and hence be well calibrated over a long enough run, yet provide no incremental information.

In decision-making under uncertainty, the investor who acts on well-calibrated highly-resolved probabilities has an "accuracy" edge over any

[11]It should be noted that there is a camp of statisticians who leave the likelihood function as the end result, interpreting it as evidence of parameter values that best explain the data. See for example Royall (1997) and Aitkin (2010).

lesser combination of those two well-defined attributes. One without the other is generally a great disadvantage in investment contexts (see later discussion on economic Darwinism).

3.29 Economic scoring rules

Economic decision-making based on subjective probability assessments, which, in turn, are based on information, is generally more successful when those probabilities are more "accurate" relative to actual outcomes. To assess the accuracy of past probability assessments, meteorologists and Bayesian statisticians developed formal mathematical probability score functions, or "scoring rules".

One well-known scoring rule is the log score. If the event $E \in \{B, NotB\}$ is binary like B ("bankrupt") or $NotB$ ("not bankrupt"), then the score attached to a bankruptcy probability p is

$$S(p) = \begin{cases} \log(p) & \text{if } E = B \\ \log(1-p) & \text{if } E = NotB. \end{cases}$$

This is score is known to be "strictly proper" in the sense that someone who believes bankruptcy probability r but states probability p, will maximize her personal expected score,

$$r\log(p) + (1-r)\log(1-p),$$

if and only if she reports her true belief. That is, she reports truthfully $p = r$. Proper scoring rules are akin to conventional utility functions in the sense that decision makers wanting to maximize their own ex ante expected utilities must act "as if indeed they do believe their own beliefs".

Probability scoring rules can have direct economic interpretation, for example, consider a decision maker qua gambler who believes probability p and bets against a market maker (bookmaker) who quotes market probability q. If the gambler has utility $\log(w)$ for money wealth w, her realized utility if she begins with wealth w_0 is

$$\log(w_0) + \begin{cases} \log(p/q) & \text{if } E = B \\ \log\left[(1-p)/(1-q)\right] & \text{if } E = NotB. \end{cases}$$

So the economic quality of her score is captured by the difference between her log score and the bookmaker's log score,

$$\begin{cases} \log(p) - \log(q) & \text{if } E = B \\ \log(1-p) - \log(1-q) & \text{if } E = NotB. \end{cases}$$

3.30 Market scoring rules

The theory of scoring rules developed by Bayesian statisticians overlaps directly with betting markets, which are essentially markets for "binary options". Suppose that there are n probability forecasters with respective probability beliefs $\{p_1, p_2, \ldots, p_n\}$, respective wealth endowments $\{w_1, w_2, \ldots, w_n\}$ and all with log utility $\log(w)$. It is easily shown that if each investor "bets her belief", so as to maximize her personal expected log utility, her realized utility is

$$\log(w) + \begin{cases} \log(p) - \log(\overline{p}) & \text{if } E = B \\ \log(1-p) - \log(1-\overline{p}) & \text{if } E = NotB, \end{cases}$$

where the equilibrium (market-clearing) asset price is the wealth-weighted belief of the n gamblers

$$\overline{p} = \frac{(w_1p_1 + w_2p_2 + \cdots + w_np_n)}{\sum p_i}.$$

Note that $\overline{p}$ might be interpreted loosely as a "market consensus" probability, but does not have any Bayesian justification. It is a weighted "average" of the different Bayesian beliefs of the gamblers, but is not itself formally Bayesian, and may not coincide with anyone's Bayesian belief. This is interesting in terms what constitutes an efficient market, because it shows how "market beliefs" can be driven by individuals' rational Bayesian beliefs and yet not themselves be Bayesian. That is, a market of strictly Bayesian investors is not by that fact itself Bayesian.

3.31 Measures of information

Shannon invented a measure of the information contained in the user's current probability distribution. Consider a distribution over unknown X which can take possible values $X = x_j$. Current information makes us believe probability distribution $\Pr(X = x_j) = p_j$. Our stock of information or "negative entropy" is defined as

$$\sum_j p_j \log(p_j),$$

which is minimum when p_j is uniform, and is maximized when one of the p_j equals 1 (and the others zero). Importantly, note that the stock of information can increase or decrease. If the user's probability distribution over X becomes less peaked or flatter with new information arrival, its "entropy" or level of uncertainty

$$-\sum_j p_j \log(p_j),$$

increases. That will occur when the new information about X is unexpected or improbable based on our previous beliefs about X (i.e. when the new information contradicts the user's previous beliefs).

There is an intimate relationship between Shannon's entropy (with base e) and natural log utility. Information measures could be based on other utility functions, but the log measure has been found to have many desirable theoretical and intuitive properties. In particular, log information is additive in the sense that the log odds gained by learning something in two (or more) steps is the sum of the two increments (e.g. we might learn the result of rolling a die by first learning that it is an even number, then that it is not "four", ...).

3.32 Ex ante versus ex post accuracy

Any new information source or proposed experiment has positive expected utility merely because new beliefs always make the old ones seem wrong. Put another way, when the user has new beliefs, she does not expect on those beliefs that her earlier beliefs are more accurate. They nonetheless might be, as measured ex post by comparison with the event actually realized. Suppose, for example, that the new belief is that it will rain with probability 0.6, whereas that probability was previously only 0.8. If in fact it rains, then the earlier belief is more accurate relative to the outcome, and would generally have brought higher profits or lower losses for investors.

3.33 Sampling to forgone conclusion

An established property of most conventional significance tests is that they can hypothetically be driven to a forgone conclusion. Suppose that the model says that variable X is normally distributed with unknown mean θ and known unit variance. By the law of the iterated logarithm, it is guaranteed that by drawing enough observations the sample mean will sooner or later obtain statistical significance at any chosen "critical level" (greater than zero). Thus, if we set out to obtain 5% statistical significance in a two-sided test, we can draw until $|\overline{x}| \geq 1.96\sigma (= 1.96)$, and that will eventually occur simply by continued sampling.

In Bayesian eyes, if a frequentist adopts this method of "sampling to a foregone conclusion",[12] she is not tampering with the data, she is merely drawing more of it, and is thus paying the cost of more sampling rather than altering anything about the data's relevance or validity.

No Bayesian foregone conclusions

While a frequentist can always in theory (with an infinite budget) sample to a foregone significance level, it is not possible to sample to a guaranteed small posterior probability. The two objectives are of different evidential consequence. If a desirably small significance level is achieved only with a very large sample, it is effectively evidence supportive of the hypothesis tested (see Lindley's paradox), so a true null hypothesis cannot truly be discredited merely by obtaining more and more data.

By comparison, a small posterior probability, with any sample size, does discredit the hypothesis tested (the sample size is taken into account in the calculation of that probability). It is logical and reassuring therefore that an arbitrarily small posterior probability cannot be guaranteed by prolonging sampling, but instead can only come by the evidence falling that way.

Examples exist where the experimenter sets out to sample to a given small posterior probability. Depending on the input assumptions, that

[12] Such an approach actually breaks frequentist rules, because it affects the sample space in a way that is not used in the calculation of the p-level. All samples in the sample space have p-level less than or equal to the target.

result is usually possible ex ante but far from guaranteed, the task is much like someone who decides to hitchhike and just walk until picked up; a ride might come early or never.

Suppose that in Bernoulli sampling the Bayesian tests $H_0 : \theta = 0.5$ against $H_1 : \theta = 0.6$, assuming equal prior probabilities $f(H_0) = f(H_1) = 0.5$. Let θ be the probability of "red". To obtain a posterior probability of γ for H_0, the required observation must be a pair (n, s) for which the likelihood ratio

$$\frac{0.5^s(1-0.5)^{n-s}}{0.6^s(1-0.6)^{n-s}} \leq \frac{\gamma}{1-\gamma}.$$

The process is to draw from the null distribution $\theta = 0.5$ until the first (n, s) arrives for which $p(H_0|n, s) \leq \gamma$, and then stop and report. If that target is not reached, sampling continues in the hope that it will, until in effect n becomes "infinite". Simulation of this process shows that the chance of "succeeding" at level γ is as shown in the following table.

γ	p(Target Met)
0.05	4.54%
0.10	9.95%
0.20	23.4%

So there is only a 4.54% chance of obtaining a posterior belief of 5% in H_0, when in fact H_0 is correct. In the other 94.6% of attempts, the sample size becomes too huge without success to allow any possible turnaround, and the probability of H_0 goes to one. The Bayesian view of this unorthodox sampling method is again that there is no problem with the observations, they are all genuine, untainted and exchangeable. A problem occurs however if there is no report issued whenever the target posterior is not met, since in effect that conceals data strongly supportive of H_0.

Non-reporting is an instance not of falsifying data but of not reporting data that is both known and highly informative, which is another form of what has long been called the "file drawer problem".[13]

[13]The file drawer is reputed to contain data that did not suit the experimenter's preferred position, and can be viewed as a widespread principal–agent problem in statistics.

3.34 Predictive distributions

A common problem is to use the Bayesian posterior distribution $f(\theta|x_1, x_2, \ldots, x_n)$ to forecast, in the form of a probability distribution, the next observation x_{n+1}. That requires "integrating out θ" or "averaging" over the possible values of parameter θ,

$$\begin{aligned} &f(x_{n+1}|x_1, x_2, \ldots, x_n) \\ &\quad = \int f(x_{n+1}|\theta, x_1, x_2, \ldots, x_n) f(\theta|x_1, x_2, \ldots, x_n) d\theta \\ &\quad = \int f(x_{n+1}|\theta) f(\theta|x_1, x_2, \ldots, x_n) d\theta, \end{aligned}$$

since generally $f(x_{n+1}|\theta, x_1, x_2, \ldots, x_n) = f(x_{n+1}|\theta)$.

Mathematical difficulties typically arise when there is more than one parameter, and the available solutions are only by numerical (simulation) methods.[14]

3.35 Model averaging

Models are said to be always wrong. Bayesian inference sees the model as just another unknown and treats it like a parameter by attaching a probability to each possible model and using data to update beliefs about not only the parameters in the model but also the whole model. Ultimately, the probability distributions attached to predictions of variables like whether it will rain are the averages of the within-model beliefs across those models. That is called model averaging, and avoids the more conventional approach in statistics of "assuming" a model and either staying with it or discarding it completely for another model.

Suppose that the model is $m \in M$ and the parameter specified with m (the same in all m) is $\theta \in \Theta$, and the information observed to date is x_{obs}, then the next value of unknown X has probability distribution,

$$f(X|x_{obs}) = \sum_M \Pr(m|x_{obs}) \int_\Theta f(\theta|m, x_{obs}) f(X|m, \theta, x_{obs}) d\theta,$$

[14] Advancements in numerical computation techniques in recent years have made Bayesian statistics practical in very realistic contexts, which is a large part of why there has been a boom in Bayesian applications in "big data" science, robotics and other fields.

which is called the predictive distribution of X, and amounts to the "average" probability of X across the different models considered, where each model is weighted by its posterior probability given past observation x_{obs}. Under this approach, any new data allows the user to update her within-model beliefs about parameter θ, $f(\theta|m, x_{obs})$, and also her beliefs, $f(m|x_{obs})$, about which model is more probable on all the information received so far. That is essentially a model for all statistical inference, because it allows for learning not only about a model's uncertain parameter(s) but also about the model as a whole, as if m is merely another uncertain parameter.

3.36 Definition of a subjectivist Bayesian

It helps to sum up the most fundamental distinguishing features of subjectivist Bayesianism. This summary follows directly from the crystal clear writing of philosopher/statistician Seidenfeld (1985).[15] Subjectivism requires at a minimum: (i) coherence, i.e. all probabilities understood as degrees of belief are mutually consistent under existing knowledge K; (ii) the law of total evidence, i.e. all knowledge K existing at time t is included in the coherent beliefs formed at that time; and (iii) conditionalization, i.e. if new evidence e arrives, all probabilities are re-conditioned onto $K \cap e$.

These three requirements may seem too obvious to state, but if that were the case there would not have been so much antagonism towards the openly subjectivist theory of Bayesian statistics advanced by Savage and others. For a similar depiction of Bayesian statistical rationality in information economics, see Hirshleifer and Riley (1992).

3.37 What makes a Bayesian?

The following quotes from distinguished Bayesians will be of interest to anyone whose understanding of statistical theory and its rival schools

[15] It may come as no surprise that the fundamental issues of Bayesianism versus frequentism, all the way from inference to decision, have their own long standing and highly developed literature in the logic and philosophy of science. See Seidenfeld (1979) and Howson and Urbach (2005) for an introduction.

and origins is somewhat patchy or accidental, rather than learned by specific attention to the foundations of statistical methods and the literature in statistics and philosophy on methodological and philosophical foundations. They give a very frank and easy to understand idea of why someone might regard themselves as "Bayesian" and how they show that. The quotes are from a statement by Cowles, Kass and O'Hagan posted on the authoritative Bayesian website operated by the *International Society for Bayesian Analysis (ISBA)*. See *https://Bayesian.org/what-is-Bayesian-analysis/*

> There are many reasons for adopting Bayesian methods, and their applications appear in diverse fields. Many people advocate the Bayesian approach because of its philosophical consistency. Various fundamental theorems show that if a person wants to make consistent and sound decisions in the face of uncertainty, then the only way to do so is to use Bayesian methods. Others point to logical problems with frequentist methods that do not arise in the Bayesian framework. On the other hand, prior probabilities are intrinsically subjective — your prior information is different from mine — and many statisticians see this as a fundamental drawback to Bayesian statistics. Advocates of the Bayesian approach argue that this is inescapable, and that frequentist methods also entail subjective choices, but this has been a basic source of contention between the 'fundamentalist' supporters of the two statistical paradigms for at least the last 50 years. In contrast, it is more the pragmatic advantages of the Bayesian approach that have fuelled its strong growth over the last 20 years, and are the reason for its adoption in a rapidly growing variety of fields. Powerful computational tools allow Bayesian methods to tackle large and complex statistical problems with relative ease, where frequentist methods can only approximate or fail altogether. Bayesian modelling methods provide natural ways for people in many disciplines to structure their data and knowledge, and they yield direct and intuitive answers to the practitioner's questions.

It is essential to the honesty of this monograph that Bayesianism is not understood as a completely unified practice. Its strictest or most developed tools go all the way from subjective priors to subjective utility functions and personal optimal decisions under the rule of maximizing expected utility. The utility axioms in economics show that rationality requires the individual to act *as if* he or she is maximizing a proper utility function. The axioms however do not necessitate that probabilities are subjective. Subjectivity occurs only because there is no alternative when decisions have to be made in contexts where there are no probabilities that even resemble "objective" or "physical" probabilities. Of the different creeds of Bayesiansim described in the following quote from Cowles, Kass and O'Hagan, the one that has been put into effect by Demski, Feltham, Dye and others in accounting theory, and is my approach in this monograph, is the uninhibited version of subjective expected utility maximization (often abbreviated to SEU in the Bayesian literature):

> There are many varieties of Bayesian analysis. The fullest version of the Bayesian paradigm casts statistical problems in the framework of decision making. It entails formulating subjective prior probabilities to express pre-existing information, careful modelling of the data structure, checking and allowing for uncertainty in model assumptions, formulating a set of possible decisions and a utility function to express how the value of each alternative decision is affected by the unknown model parameters. But each of these components can be omitted. Many users of Bayesian methods do not employ genuine prior information, either because it is insubstantial or because they are uncomfortable with subjectivity. The decision-theoretic framework is also widely omitted, with many feeling that statistical inference should not really be formulated as a decision. So there are varieties of Bayesian analysis and varieties of Bayesian analysts. But the common strand that underlies this variation is the basic principle of using Bayes' theorem and expressing uncertainty about unknown parameters probabilistically.

After a lifetime's Bayesian work, Kadane's (2011) summary of the subjectivist Bayesian ethos is that Bayesian probability theory, or really just probability theory, of which Bayes theorem happens to be integral, is merely a language with rules by which to tie together and express mutually consistent beliefs:

> In my view, probability is like a language. Just as grammar specifies what expressions follow the rules that make thoughts intelligible, the rules of coherence specify what probability statements are intelligible. That sentences are grammatical says nothing about the wisdom of what is expressed. (Kadane, 2011, p. 447)

3.38 Rise of Bayesianism in data science

Big data and machine learning has boosted awareness and practical application of Bayesian logic (Murphy, 2012). Computer science is essentially now a Bayesian field. One obvious reason for this is that when looking for empirical flags or indicators about some unknown outcome or parameter, the intuitive way to grasp what data x "suggests" regarding H is to think of whether x would likely occur if H were true, and whether it is equally likely, or more/less likely, under *not-*H. The ratio of those two assessments is the likelihood ratio used in Bayes theorem, and the observations that are the clearest suspects to be good machine indicators are ones with empirical likelihood ratios approaching 0 or ∞.

Machine learning allows a search, not for individual indicators, but for joint indicators $(x, y, z, \ldots)$ that have pronounced (high or low) likelihood ratios, either of themselves directly,

$$\frac{f(x, y, z, \ldots|H)}{f(x, y, z, \ldots|\textit{not-}H)},$$

or in some functional combination defined by a model, $m(x, y, z, \ldots)$,

$$\frac{f[m(x, y, z, \ldots)|H)]}{f[m(x, y, z, \ldots)|\textit{not-}H)]}.$$

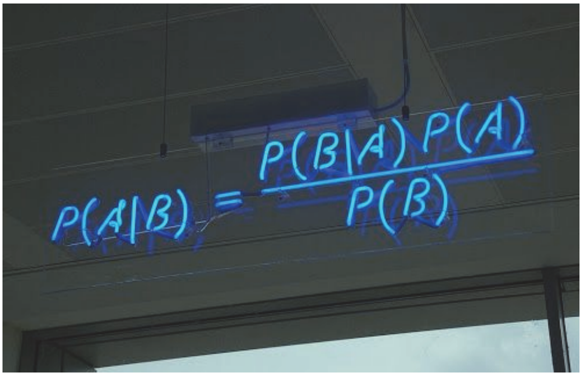
The best indicator is usually not any one factor of itself but its combination with others, very much like the way that variables contribute partial correlation to a regression equation. Such combinations of indicators are effectively searched out by algorithms designed for that purpose. Their more or less explicit use of the likelihood ratio is their Bayesian base.

Different families of machine-learning algorithms are more or less Bayesian internally, depending on their "logic", and they all offer different results. By averaging beliefs over the different models, there is a much reduced risk of over-fitting or "model risk". Frequentist statistics does not have the Bayesian capacity to formally merge the results that are found under different models, nor do they produce probabilities of propositions. Results therefore tend to be model-specific and "unhedged". Machine-learning methods do not commit to any particular "causal" view of the world. The Bayesian approach keeps all models "in play" and allows each model to gain more or less influence as data arrive and are better explained by some models than others. This approach is philosophically consistent with a "wisdom of crowds" or combined opinion approach, supported by the success of averages over individuals in many real-world contexts.

> Bayesian model averaging has been used to study heart attacks in medicine, traffic congestion in transportation economy, hot hands in basketball, and economic growth in the macroeconomy literature. In finance, Bayesian model averaging facilitates a flexible modeling of investors' uncertainty about potentially relevant predictive variables in forecasting models. In particular, it assigns posterior probabilities to a wide set of competing return-generating models (overall, 2^M models). It then uses the probabilities as weights on the individual models to obtain a composite-weighted model. This optimally weighted model is then employed to investigate asset allocation decisions. Bayesian model averaging contrasts sharply with the traditional classical approach of model selection. In the latter approach, one uses a specific criterion (e.g., adjusted R^2) to select a single

> model and then operates as if that selected model is correct. Implementing model-selection criteria, the econometrician views the selected model as the true one with a unit probability and discards the other competing models as worthless, thereby ignoring model uncertainty. Accounting for model uncertainty, Avramov (2002) shows that Bayesian model averaging outperforms, ex post out-of-sample, the classical approach of model-selection criteria, generating smaller forecast errors and being more efficient. Ex ante, an investor who ignores model uncertainty suffers considerable utility loses. (Avramov and Zhou, 2010, p. 38)

Bayes Neon Sign at Autonomy Corporation PLC, Cambridge, UK 1996 (now HP Autonomy)



Textbooks on Bayesian Theory

The following is a list of some important Bayesian reference texts, categorized according to their general content and style.

Historically Important Texts

Box, G.E.P. and Tiao, G.C. (1973) *Bayesian Inference in Statistical Analysis*. Reading, Mass: Addison-Wesley.

Blackwell, D. (1969) *Basic Statistics.* New York: McGraw-Hill.

de Finetti, B. (1974/5) *Theory of Probability.* Vols 1 and 2. (Translated 1970 book). New York: Wiley.
Good, I.J. (1950) *Probability and the Weighing of Evidence.* London: Charles Griffen.
Jeffreys, H. (1939) *The Theory of Probability.* Oxford University Press.
Lindley, D.V. (1965) *Introduction to Probability and Statistics*: *From a Bayesian Viewpoint.* Parts 1 and 2. Cambridge University Press.
Lindley, D.V. (1972) *Bayesian Statistics: A Review.* Philadelphia: Society for Industrial and Applied Mathematics.
Lindley, D. V. (1985) *Making Decisions*. New York: Wiley.
Raiffa, H. (1968) *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. New York: Random House.
Savage, L.J. (1954) *The Foundations of Statistics*. New York: Wiley.
Schlaifer, R.O. (1959) *Probability and Statistics for Business Decisions: An Introduction to Managerial Economics under Uncertainty.* New York: McGraw-Hill.
Schlaifer, R.O. (1969) *Analysis of Decisions under Uncertainty.* New York: McGraw-Hill.
Zellner A. (1971) *An Introduction to Bayesian Inference in Econometrics.* New York: Wiley.

Introductory Textbooks (Philosophical)

Berry, D.A. (1996) *Statistics: A Bayesian Perspective.* New York: Duxbury.
Thompson, B. (2007) *The Nature of Statistical Evidence.* New York: Springer.

Advanced Texts (Technical and Philosophical)

Aitkin, M. (2010) *Statistical Inference: An Integrated Likelihood/Bayesian Approach.* Boca Raton: Chapman and Hall.
Bernardo, J.M. and Smith, A.F.M. (1994) *Bayesian Theory.* New York: Wiley.
Berger, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis.* 2nd ed. New York: Springer-Verlag.
Berger, J.O. and Wolpert, R.L. (1988) *The Likelihood Principle.* 2nd

ed. Haywood, California: IMS Lecture Notes — Monograph Series B.
Cyert, R.M. and DeGroot, M.H. (1987) *Bayesian Analysis and Uncertainty in Economic Theory.* London: Chapman and Hall.
DeGroot, M.H. (1970) *Optimal Statistical Decisions.* New York: McGraw-Hill.
French, S. (1988) *Decision Theory: An Introduction to the Mathematics of Rationality.* Chichester, UK: Ellis Horwood.
Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004) *Bayesian Data Analysis.* 2nd ed. Boca Raton, FL.: CRC Press.
Jaynes, E.T. (2003) *Probability Theory: The Logic of Science.* Cambridge University Press.
Kadane, J.B. (2011) *Principles of Uncertainty.* Boca Raton, FL.: Chapman and Hall.
O'Hagen, A. (1994) *Kendall's Advanced Theory of Statistics. Volume 2B: Bayesian Inference.* London: Edward Arnold.
Pratt, J.W., Raiffa, H. and Schlaifer, R. (1995) *Introduction to Statistical Decision Theory.* Cambridge, Mass.: The MIT Press.
Press, S.J. (2003) *Subjective and Objective Bayesian Statistics.* 2nd ed. New York: Wiley.
Robert, C. (2007) *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation.* 2nd ed. New York: Springer.
Royall, R.M. (1997) *Statistical Evidence: A Likelihood Approach.* London: Chapman and Hall.
Smith, J.Q. (2010) *Bayesian Decision Analysis: Principles and Practice.* Cambridge University Press.
Winkler, R.L. (2003) *An Introduction to Bayesian Inference and Decision.* 2nd ed. Sugar Land, Texas: Probabilistic Publishing.

Highly Philosophical

Bovens, L. and Hartmann, S. (2003) *Bayesian Epistemology.* Oxford University Press.
Cooke, R.M. (1991) *Experts in Uncertainty: Opinion and Subjective Probability in Science.* Oxford university Press.

Good, I.J. (1983) *Good Thinking: The Foundations of Probability and Its Applications.* Minneapolis: University of Minnesota Press.
Howson, C. and Urbach, P. (2005) *Scientific Reasoning: The Bayesian Approach.* 3rd ed. Chicago: Open Court.
Lad, F. (1996) *Operational Subjective Statistical Methods: A Mathematical, Philosophical and Historical Introduction.* New York: Wiley.
Lindley, D.V. (2014) *Understanding Uncertainty.* New York: John Wiley.
Seidenfeld, T. (1979) *Philosophical Problems of Statistical Inference: Learning from R.A. Fisher.* Boston: D. Reidel.
Kadane, J., Schervish, M.J. and Seidenfeld, T. (1999) *Rethinking the Foundations of Statistics.* Cambridge University Press.

4

Case Study: Using All the Evidence

This case study re-examines Burgstahler's (1987) method for drawing Bayesian inferences from frequentist empirical test results. The findings apply not only to the original application of interpreting the results in empirical research studies, but also far more generally to interpreting any ill-defined or incomplete signal or statement of evidence.

The following analysis reveals how Bayesian interpretations of data, or of the translation of data that is actually reported to the user, are not merely subjective, but are also often highly sensitive to the Bayesian user's probability model, background knowledge or basic assumptions. In general, the more subjective the analysis, the wider its range of possible inferences, yet the more realistic its approach. The antidote to subjectivity is usually "get more data", but often there is a decision to make that cannot wait, or there is no possibility of more data, or other researchers want a conclusion.

That limitation is deeply understood in the information economics models used in accounting theory. Accounting, of all applications, will often leave the information user with a less than complete report, and yet needing to act or form beliefs on what is reported, despite its perceived weaknesses. In the "worst" cases, the user is left to act in the face of no express report at all, and to interpret that non-report for what it implies. See for example Dye (2017), Dye and Hughes (2018) and the corresponding analysis set out later in this monograph.

4.1 Interpreting "*p-level* $\leq \alpha$"

The Bayesian theory of experimental design applies to the ex ante planning of experiments or "signal design". It is recognized however that often the user or decision maker does not design the signal, nor know all about how it was produced, yet must still interpret it as best possible. Even a signal which is known to be imprecise or biased can still change rational Bayesian beliefs, sometimes substantially, and can still be highly informative.

The following case study based on Burgstahler (1987) illustrates how a given statistical signal can have quite contrary interpretations under different levels of Bayesian conditioning, or essentially under different levels of subjectivity.

The inference problem raised by Burgstahler (1987) is to use the published report "significant at α" to revise belief in the null hypothesis H_0 against alternative H_1. Without introducing an alternative hypothesis it is not possible to calculate the probability of H_0, because there exists only one half of the likelihood ratio.[1] The idea of Burgstahler's analysis is to help empirical researchers interpret classical (i.e. frequentist) statistical evidence in a Bayesian way, so as ultimately to assess the probability of the null hypothesis, which is theoretically disallowed in Neyman–Pearson classical statistics. Burgstahler correctly notes that the usual non-Bayesian way of calculating and interpreting significance levels does not admit any statement about the probability of H_0.[2]

There are at least three possible meanings to "significant at α". If all three are plausible, the Bayesian posterior belief is a mixture or probability-weighted average of the three corresponding posterior

[1]Neyman found the lack of an explicit alternative to be a weakness in Fisher's logic of significance tests. For a fascinating journey through the history of statistical tests, see Neyman (1950) for his reconstruction of Fisher's famous tea lady test. Then see Lindley's (1984) "Bayesian Lady Tasting Tea" for the "final episode".

[2]A significance level is a probability of the data conditional on the null hypothesis, not the null hypothesis conditional on the data.

distributions. The three possible meanings and their associated Bayesian interpretations are examined individually below.

To match standard textbook hypothesis testing models, I assume a point null hypothesis $H_0 : \theta = \theta_0$ against a composite alternative $H_1 : \theta \neq \theta_0$, and let H_0 have non-zero prior probability π_0. The remaining prior probability is distributed over $\theta \neq \theta_0$ as $f(\theta)$, so $\int f(\theta)d\theta = (1-\pi_0)$ and $f(\theta|H_1) = f(\theta)/(1-\pi_0)$. Burgstahler considered a point alternative hypothesis, but that case is subsumed in the following model.

Meaning 1: Burgstahler's result

Report "significant at α_1" might be taken to imply no more and no less than its literal meaning, namely, that the observed *p-level* is less than or equal to α. Letting S_1 denote "significant at α_1", implying $f(S_1|\theta_0) = \alpha$. The posterior odds against H_0 are

$$\begin{aligned} \frac{f(H_1|S_1)}{f(H_0|S_1)} &= \frac{f(H_1)}{f(H_0)}\frac{f(S_1|H_1)}{f(S_1|H_0)} \\ &= \frac{\int f(\theta|H_1)f(S_1|\theta, H_1)d\theta}{\pi_0\, f(S_1|\theta_0)} \\ &= \frac{\int f(\theta|H_1)f(S_1|\theta, H_1)d\theta}{\pi_0\, \alpha_1}. \end{aligned}$$

Now, as the sample size $n \to \infty$, the power of the experiment with respect to any θ in H_1,

$$f(S_1|\theta, H_1)$$

goes monotonically to one, for any fixed $\alpha_1 > 0$, and hence the limiting odds against H_0 as $n \to \infty$ are

$$\frac{\int f(\theta|H_1)d\theta}{\pi_0\, \alpha_1} = \frac{1-\pi_0}{\pi_0}\frac{1}{\alpha_1},$$

which approaches ∞ as $\alpha_1 \to 0$.

Re-expressing the odds against H_0 as a posterior probability gives a limiting probability

$$\lim_{n\to\infty} f(H_0|S_1) = \frac{\pi_0\, \alpha_1}{\pi_0\, \alpha_1 + (1-\pi_0)},$$

which approaches zero as $\alpha_1 \to 0$.

Whether shown in odds form or probability form, the conclusion is that "significant at α_1", interpreted literally, constitutes *stronger* evidence against H_0 as n increases and as fixed α_1 decreases. That is the usual intuitive way that statistical significance is interpreted.

Meaning 2: A more Bayesian interpretation

Report "significant at α_1" might be understood as code for "significant at α_1 but not significant at α_2", where $\alpha_2 < \alpha_1$. For example, "significant at 0.05" might imply significant at $\alpha_1 = 0.05$ but not at $\alpha_2 = 0.01$.[3] Letting $S_2 \subset S_1$, and writing the intersection of the two tail-areas as S_{12}, the posterior odds against H_0 are

$$\frac{f(S_{12}|H_1)}{f(S_{12}|H_0)} = \frac{f(H_1)}{f(H_0)} \frac{f(S_{12}|H_1)}{f(S_{12}|H_0)}$$
$$= \frac{\int \{f(S_1|\theta, H_1) - f(S_2|\theta, H_1)\} f(\theta|H_1)d\theta}{\pi_0 (\alpha_1 - \alpha_2)},$$

As $n \to \infty$, both $f(S_1|\theta, H_1)$ and $f(S_2|\theta, H_1)$ go to one, and hence $f(S_{12}|H_1)$ approaches zero, implying that the limiting posterior odds against H_0 are zero. Interestingly, therefore, a result known to be significant at one level but not at a lower level (greater than zero) carries *less* weight against the null hypothesis H_0 as $n \to \infty$. The limiting probability is

$$\lim_{n\to\infty} f(H_0|S_{12}) = 1.$$

Meaning 3: Lindley's paradox

Report "significant at α" might be just a conventional or rhetorically appealing way of saying $p\text{-}level = \alpha$. Bayesian interpretation of this information follows immediately from the results above, because $p\text{-}level = \alpha$ implies "significant at α_1 but not at α_2, where α_2 is infinitesimally less than α_1". Hence, remarkably,

$$\lim_{n\to\infty} f(H_0|p\text{-}level = \alpha) = 1,$$

[3] The "stars" system of reporting significance levels is of this explicit form, where for example two stars might signify "significant at 0.01 but not at 0.001". See Ohlson (2015) and Ohlson (2018).

implying therefore that *any* arbitrarily small fixed *p-level* represents stronger and stronger evidence in favor of H_0, not against it, as n increases.

That result is one of the main points in the Bayesian critique of conventional frequentist statistical tools, and is widely known as Jeffreys' or Lindleys' paradox. See Lindley (1957). Johnstone and Lindley (1995) give example calculations by which to interpret "significant at α" in the common test of a normal mean with a normal prior.

4.2 Bayesian interpretation of frequentist reports

The end result is that the report "significant at α" can induce quite opposite inferences depending on what is read into those words, and on the (usually known) sample size of the test. A subjective Bayesian interpretation would put some weight on each possible interpretation, and hence average the three posterior probabilities of H_0 (this is akin to what's called model averaging in Bayesian methodology). Choice of those weights will naturally involve background knowledge of reporting conventions and motivations, and would be innately subjective, thus making the phrase "significant at α", when reported in isolation, ambiguous in terms of its weight of evidence.

That is an important result for empirical accounting research, but also for accounting theory, where so often the interpretation of accounting disclosures depends on how they are "read" or what is "read between the lines". A direct analogy applies whenever for example the firm reports that sales are expected to increase by at least 5%. Does this mean "hopefully 5%" or "at least 5%", or perhaps "somewhere close to 5%"? In a Bayesian inference model, those different plausible interpretations of statements of the form $x \geq x_c$ can result in very different posterior beliefs.

4.3 A generic inference problem

Note that the problem of interpreting "*p-level* $\leq \alpha$" is the same general problem as arises whenever a measurement or report of some variable x comes in the form $x \leq x_c$, or $x \geq x_c$. That inference problem arises when a measurement instrument has a minimum or maximum reading of x_c and the object measured hits that stop. Similarly, reporting thresholds often leave a signal known only to belong to an interval. When the firm is classed as a "going concern", there is no express statement about how "going" it is.

In accounting theory, Dye and Hughes (2018) discussed the Bayesian interpretation of an accounting "non-report" taken as implying that the not-reported observation in question is somewhere less than a critical threshold x_c (see later Discussion). They make a correct subjectivist Bayesian point of interpreting this information in context with all of its surrounding circumstances, including the background motivations of the reporter. That general approach is common in the analytical strategic disclosure literature in accounting, and is distinctively "Bayesian" in its insistence on using all available prior and background information, even when those extra considerations are highly subjective.

5

Is Accounting Bayesian or Frequentist?

> Since uncertainty lies at the heart of accounting, there is no alternative to describing accounting information in probabilistic terms. (Dye, 2001, p. 212)

Probabilistic terms can be either frequentist (often called classical, objectivist, orthodox, non-Bayes or sampling-theoretic) or Bayesian.[1] It is hard to categorize the accounting literature as either predominately frequentist or Bayesian, as there are strands, concepts and language of both schools of thought inter-mixed across and often within individual research papers. Nor has there been any significant debate between Bayesian and non-Bayesian camps, unlike for example the longstanding debate in psychology over the validity of conventional hypothesis tests and *p*-values.[2] Accounting as a discipline has not entered this debate,

[1] What's in the name "Bayesian", asks Fienberg (2006).

[2] The dispute in psychology concerning the evidential value of significance testing began decades ago and has lately blown up into a fight about the scientific validity of much of the empirical social sciences. The multi-authored paper by Benjamin *et al.* (2018) is one of many recent fundamental rejections of conventional hypothesis testing research practices. Weak reproducibility and concern about the credibility of new empirical discoveries based on "significant at 0.05" has drawn attention in social sciences to institutionalized "*p*-hacking" and overly easy rejection of straw-man

excepting perhaps Lindsay (1995), Basu (2015), Johnstone (1997), Ohlson (2015), Ohlson (2018), and Dyckman (2016).

Implicit reliance on both approaches is common in empirical research. The theory component of an empirical paper will usually presume rational Bayesian investor behavior, and formulate hypotheses based on Bayesian information-economics models developed in the analytical accounting or finance literatures. However, when it comes to the empirical section of the paper, the methods and language used are almost always explicitly frequentist. A common juxtaposition is that Bayesianly justified hypotheses are tested using frequentist significance tests, but the results of those tests are interpreted quasi-Bayesianly in terms of an intuitive degree of evidence.

While it might seem odd that this apparent mismatch has long become the norm, there is some awareness of the issue and some possible remedies. Thirty years ago, Burgstahler (1987) raised the issue in accounting research of how to interpret frequentist hypothesis tests Bayesianly, and there is a substantial analytical literature in statistics that looks not only for inconsistencies, but also for possible translations of p-values to Bayesian beliefs, and for any common ground where the two approaches lead to qualitatively similar conclusions. See for example Berger and Sellke (1987) and the series of papers on this topic by James Berger. Interestingly, Lindley initially set out to "write" classical statistics in Bayesian terms, but then struck inherent logical inconsistencies.

One glaring contradiction occurs when conclusions from empirical testing are based on observed "significance levels" rather than confidence intervals. That is where "Lindley's paradox" reveals a real disparity in the two schools' conclusions from the same data, and has lately

hypotheses, and a general cultural readiness to overstate the weight of evidence implied by "statistically significant" results. These authors as a group recommend at least the patch solution of setting the threshold for new discoveries at p-level $<$ 0.005. They say that this suggestion and all the related critique is decades old, but that a critical mass of researchers now endorse such a shift in routine practice. In accounting, this approach would be called "conservatism", because it requires heightened evidence to accept a "desired" outcome. There is to date little "official" recognition in accounting of such unrest in the logical statistical foundations of conventional non-Bayes methods.

been the subject of much introspection across many fields, including finance.[3]

While empirical researchers in accounting have a choice in what statistical school of thought they apply, or in how they apply a "Bayes/non-Bayes compromise", there is no hesitation within the subset of accounting theory that is variously described as "information economics", "noisy rational expectations" or "strategic information provision models". That body of work is unquestionably Bayesian, because it is built explicitly on the logic and choice of rational agents.[4] Often without any conscious rejection of frequentist constructs, these models take Bayesian inference and decision-making as axiomatic.

Being "Bayesian", all such models are "immune" to the specific criticisms and counter-examples that Bayesian thinking points at frequentist statistics. It is possible, always, to be "more Bayesian", in the sense of introducing a higher level of conditioning or subjectivity into essentially the same Bayesian model, and hence different apparently equally "Bayesian" models can produce quite different and sometimes opposite results. An example based on a reconstruction of Simpson's paradox is set out later in this monograph. See also the previous example based on Burgstahler (1987).

[3]Significance tests are criticized along Bayesian lines for flawed logic, but they are also criticized for being too easy to "milk" in the sense of finding "significant results" (a principal–agent problem between researcher and funding agency). That critique is starting to emerge in accounting, e.g. Basu (2015) and Ohlson (2015) and Ohlson (2018). Parts of finance came to Bayesian method via the parameter risk literature in portfolio optimization, and there have been basic criticisms of significance tests made in that literature. See for example Shanken (1987), Lewellen and Shanken (2002) and recently Harvey (2017).

[4]Behavioral finance would see this as a weakness, not in terms of interesting and revealing modelling, but in terms of the empirical descriptive validity of the conclusions. There is however a great overlap because of the self-interest assumption that runs through Bayesian models of strategic disclosure and the like, rather than for example a more old-fashioned assumption of providing information that is most informative or in the best interests of the user (not the provider). See also Dye (2001) responding to any suggestion of modelling information users other than Bayesianly.

5.1 Two Bayesian schools in accounting

Gao (2013a) and Gao (2013b) distinguishes correctly between two overlapping strands of Bayesian accounting theory. The oldest is the single agent textbook example, where accounting is understood as effectively little more than a non-strategic "black box" from which signals of different apparent Bayesian qualities emerge. Signals are thus exogenous from the perspective of the user, who merely applies subjective Bayesian decision theory to interpret and act on them. Gonedes (1975, p. 847) called these bits of information "signals from nature".[5]

Gao (2010) holds that while treating accounting signals as exogenous is insightful in the ways shown by Demski, Feltham and others, the full expression of subjectivist Bayesian thinking in accounting requires strategic disclosure models that allow for signals being partly endogenous.[6] The user, and the contracts that the information user imposes on the sender, are designed to affect the resulting statistical information properties, and hence influence the Bayesian meaning of realized signals. Similarly, Stocken (2013) holds that when the firm and those who receive information interact strategically, normative Bayesian information criteria, including Blackwell's theoretical ranking of alternative signals, are not sufficient grounds for the firm to choose its self-interestedly optimal reporting strategy, or for regulators to choose between different accounting rules:

> . . . Feltham (1968) considered the value of changes in an information system within a setting containing a single investor using Bayes' Theorem as an organizing framework. He considered information having the attributes of relevance, timeliness, and accuracy to be desirable. This literature

[5]Hitz (2007) views this version of Bayesian accounting theory as one of two branches of the "decision usefulness school", it being the "information theory" branch and the other branch, which does not involve Bayesian inference or any logic of inference, being the "measurement school".

[6]Gonedes (1975, p. 615) anticipated this view. He suggested that accounting standards are likely in place to prevent users of information being sent Akerlof lemons.

> seems to have suggested the ingredients for the hierarchy of qualitative characteristics developed in Statement of Financial Accounting Concepts No. 2 — Qualitative Characteristics of Accounting Information — issued in 1980. This early work considered the properties of accounting information within a single decision-maker context. The revised criteria for evaluating information characterized in Statement of Financial Accounting Concepts No. 8 — Qualitative Characteristics of Useful Financial Information — released in 2010 seem to reflect this view. The financial reporting environment, however, features several decision-makers. Investors gather information from various sources, including a firm, the firm's competitors, analysts, trade journals, and government statistical releases. ... management must be cognizant of the scrutiny of policy-makers and regulators and also of investors' rights to take legal action in the event of fraudulent material misstatement or omission of required information. In short, the accounting information environment is populated with many strategic players making payoff maximizing decisions. ... Blackwell's Theorem does not hold when multiple decision-makers interact strategically. Thus, the notion that providers of capital would prefer finer information over less fine information offers little guidance to firms when deciding on their disclosure policy. In contrast, the extant literature emphasizes that a firm's optimal disclosure policy is sensitive to the features of the information environment. Consequently, in the absence of precisely characterizing the environment, the desirable properties of accounting disclosure cannot be characterized. (Stocken, 2013, pp. 199–200).

Gao's two Bayesian schools have one essential commonality — Bayesian logic. Similarly, even in the strategically convoluted environment well recognized by Stocken, there has to be a logic on which to interpret what signals ultimately arise. That logic will imply its own criteria by which signals are useful or not. In the end, whether the firm's signal design and production is influenced or not by the user, the rational user applies Bayes theorem to interpret it. Focus in this monograph is on that process of Bayesian interpretation *of what arrives.*

The strategic information setting will be reflected in Bayesian probability revisions, to the extent at least that the signal user can assemble a sufficiently complicated and realistic model. When a Bayesian user interprets a signal x so as to make an inference about V in the form of a probability distribution $f(V|x)$, that posterior distribution is actually $f(V|x, BK)$ where BK represents all of the user's complex background knowledge, including for example the terms of the contract or the accounting rules and conventions under which the signal was produced, terms which may come within the user's influence during the preceding contracting process (which is now history, until next time). For example, the BK in Bayesian inference models will incorporate perceived dependencies between the accountant's measurement and reporting practices and the state of underlying fundamentals such as costs and revenues, whereby for example the firm might be more inclined to conservatism under some conditions than others.

In Gao's realistic circumstances, where signal arrival and quality are at least partly endogenous rather than entirely "from nature", a Bayesian user must model the reporter as much as modelling the "natural causes" of payoff V itself. For example, put simplistically, the received might need to make a probability assessment $f(V|x, BK)$ where BK includes knowledge that the firm will try to report the value of x that is "as favorable as possible under the accounting standards", or "as favorable as possible while not raising too much doubt about its own credibility". Those Bayesian probability assessments are the bread and butter of strategic accounting information equilibrium models, since Bayesian logic is required in any economically rational assessment of a conditional probability regardless of how difficult it is to model and make that assessment when information providers have individual rewards apart from merely "satisficing" users.

Although Gao's and Stocken's argument is undoubtedly correct in terms of how accounting as a strategic practice should be realistically understood, there remain everyday circumstances where a user, like a fund manager, has to interpret an accounting signal like an earnings report over which she has virtually no control, and go on from that assessment to buy or sell stock at a market price over which she has virtually no control. She is then in the position envisaged by Feltham, Demski and Savage. She treats the signal as exogenous, like the weather forecast on tonight's news.

That is a typical situation for most shareholders who have no sway over what signals the firm puts out. They will however collect other information at their own discretion, most of which is also of a quality outside their control, and yet belongs in BK as does all that is recognized about its perceivably poor quality or accuracy.[7]

Gao (2010, p. 865) looks specifically at accounting conservatism, and says that its strongest justification in standard setting is that it counteracts overly-positive reporting tendencies in firms. Its other possible justification, apparent from Demski, is that a conservative signal might have "optimal bias" (i.e. maximum expected utility) from the perspective of a given user's utility function (see later illustrations of this). That would be supported by traditional Bayesian decision analysis, and would motivate that particular user to (try to) pre-ordain, via contracts, a suitably conservatively biased signal, which is ultimately Gao's point. In essence the user might influence the signal, its content and qualities, but in the end (my point) the user must still weigh and interpret what she gets.

What information she gets is to some extent the result of prevailing accounting standards. That is similarly the case of consumers of approved medicines and pharmaceuticals who are given information about the effectiveness of those products. Agencies regulating that information are aware that drug companies are motivated to provide consumers with the most favorable information possible. That reality leads regulators back to normative considerations of what information and tests are statistically useful and reliable to users. It should not be concluded therefore that normative (Bayesian) information criteria

[7]Reliance on information that is very largely outside the receiver's influence will possibly increase with greater use of robots and machine-leaning algorithms in financial analysis.

are rendered obsolete in a regulatory environment merely because the providers and users of information are partly strategic adversaries.

5.2 Markowitz, subjectivist Bayesian

It is probably fair to say that accounting information theory and empirical research sets its foundations largely in financial theory and the statistical methods of finance. The following depiction of Markowitz as a Bayesian advocate in finance will therefore be of interest.

Markowitz, famous for his development of mean–variance portfolio theory, was one of the first avowed subjectivist-Bayesians in financial economics. Some of his many convictions to Savage and the neo-Bayesian movement are as follows.

> Of course, none of us know probability distributions of security returns. But I was convinced by Leonard J. Savage, one of my great teachers at the University of Chicago, that a rational agent acting under uncertainty would act according to "probability beliefs" where no objective probabilities are known; and these probability beliefs or "subjective probabilities" combine exactly as do objective probabilities. (Markowitz, 1991, p. 469)

Markowitz went on to say that it makes no difference whether the probabilities we have in mind are "objective" or "subjective" — ultimately they are all personal beliefs and subjective.[8]

5.3 Characterization of information in accounting

Across much of the accounting literature, especially in empirical research papers, there is a conventional way of describing accounting information quality which, on the face of it, seems to be an essentially frequentist

[8]Interestingly, also, further connecting Markowitz with the Bayesians, it has been suggested by Rubinstein (2006) and Pressacco and Serafini (2007) that mean–variance portfolio theory was in fact invented by Bayesian de Finetti, and Markowitz (2006) has confirmed that possibility.

notion. Specifically, it is normal to describe information quality in just one convenient word – "precision". For example, among many[9]:

> First, we associate earnings quality with precision, in the sense that higher quality earnings are more precise with respect to an underlying value-relevant construct that earnings is intended to describe. . . . we associate earnings quality with precise (that is, low variance) information about a construct that earnings is intended to describe (Francis *et al.*, 2006)

And then further clarification of how "precision" is a statistical notion:

> We identify quality of information in the capital markets with a statistical notion, specifically, the precision of a measure with respect to a valuation relevant construct. For a given construct, higher quality information is more precise (contains less uncertainty) with respect to that construct. (Francis *et al.*, 2006, p. 8)

Just as in these quotes, our accounting vocabulary calls "better" accounting information more "precise". That characterization is formal and statistical, and adopts the term used in frequentist statistics to describe the variance of an estimator. In frequentist terms, an estimator is more precise when it has lower variance around the true parameter, or around its mean if it is biased.[10]

An issue with using just this single frequentist notion to depict information quality is that it does not encompass all of the other ad hoc frequentist statistical signal characteristics, like bias, consistency, efficiency and so on, as set out in orthodox non-Bayesian statistics textbooks. Pre-dating the modern generic use of "precision", Ijiri and

[9] Veronesi (2000, pp. 810–11) describes . . . the standard "signal equals fundamentals plus noise".

[10] It is an issue for frequentist statistics that a biased estimator might be "precise", and the term "consistency" was introduced in frequentist theory to describe an estimator that becomes less biased and more precise as the sample size increases.

Jaedicke (1966) described the "reliability" of accounting measures as a combination of their objectivity (meaning variance) and bias, and to be measured by a mean-squared error. Similarly, in other early works:

> There are two basic components of error: bias and variability . . . (Feltham, 1968, p. 694)

> Classical techniques usually assume that the distribution is constant over the historical period and the future period. The historical observations represent samples from a population and this sample information is used to derive "unbiased" and "efficient" estimates of the parameters of the distribution which describes the population. (Demski and Feltham 1972, p. 541)

It seems incongruous that accounting theory relies so much on the word "precision", with its statistical overtones of merely variance, when at the same time accounting research is rightfully occupied with earnings management, accounting re-statements, fraud and conservatism, all of which sound like explicit, albeit possibly sometimes desirable, information biases.

The problems created by adoption of the word "precision" come out in Francis *et al.* (2006) when having defined earnings quality repeatedly in terms of precision or variance, it is later appended as almost a bonus that information is more valuable if unbiased:

> . . . information is valuable (useful) to the extent it is *unbiased* (the mean of the information variable is the same as the mean of construct described by the information). (Francis *et al.*, 2006)

Much attention is devoted here to the monograph on earnings quality by Francis *et al.* (2006) for the reason that it genuinely attempts to define what is meant in the accounting literature by "precision". The difficulties that arise are an issue traceable to frequentist statistics as a school. In particular, a biased estimator might also be "precise" or low variance. How can these two competing estimator attributes be wrapped into a single descriptor?

Statistics textbooks address the issue of combining bias and precision with another frequentist term, "consistency". A consistent estimator is one that becomes less biased and more precise as the sample size increases, which in some ways is what better accounting information aspires to do. For example, a closer audit or more scrutiny by the accountant might draw out a less noisy and less "managed" earnings report. Accounting might swap the term precision for consistency, except that consistency is an "asymptotic" property of an estimator rather than one that always exists in a given degree (like precision).

If the most appropriate frequentist word were chosen to describe better accounting information, there is an argument for "consistent", in the sense that accounting information should improve when more resources are put into it. However, better still might be the term "efficient". An "efficient" estimator is efficient relative to a cost function. In simple terms, if the estimate is to be used as input to a decision where the outcome of the resulting action is measured by a cost, the most efficient estimator is the one that presents minimum expected loss (Gelman *et al.*, 2004, pp. 111–12). That notion of information quality is possibly closer to what accounting standard setters have in mind than mere "precision" or low variance. It not only allows for bias, but also allows for an estimator to be optimal and yet still be biased, so it leaves open the possibility that a "conservative" earnings report can be optimal.

The frequentist criterion of cost "efficiency" seems loosely consistent with a normative Bayesian experimental design framework, however there remains a fundamental difference between the two approaches. In the frequentist model of efficient estimators, there is no Bayes theorem. Instead, the decision based on the observed estimate is just the estimate itself. In the simplest case, the cost attached to an estimate x of true parameter θ might be just a linear loss $|x - \theta|$ or more often a quadratic loss $(x - \theta)^2$. The Bayesian decision model is more developed, and also more subjective. It finds the expected loss stemming from estimate x by examining the action $D(x)$ that would have been taken after calculating the Bayesian posterior belief $f(\theta|x)$. To a Bayesian x is not a "decision" with an objective loss, like $(x - \theta)^2$. Instead, x is merely information upon which to revise beliefs − and then decide. That is, actions have losses, estimators do not. That is an essential difference between frequentist methods, by which estimators have loss functions, and Bayesian decision analysis.[11]

5.4 Why accounting literature emphasizes "precision"

The vernacular or convenient shorthand in accounting of talking routinely about "precision", and thus implicitly overlooking bias, can be traced it seems to two longstanding traditions in accounting literature.

The first explanation is that most formal Bayesian statistical models of an accounting signal, generally for convenience and mathematical tractability, assume a statistical information model in which the signal received or transmitted is normally distributed around the true parameter and is unbiased (has noise with mean zero) and has known variance (or precision).[12] These assumptions are part of the standard introductory Bayesian model for inference about an unknown mean under the assumption of a known population variance. That model has a very intuitively appealing closed-form distribution for the unknown population mean, which has been put to use in hundreds of accounting theory papers. Its simplicity and tractability relative to the same model with an unknown population variance, and hence unknown sampling (signal) variance, is its theoretical attraction, belying the fact that in most applications in accounting, the population variance and the mean are realistically both unknowns.

By assuming an unbiased normally distributed signal of known variance, a signal's variance or precision is its sole advantage over another signal, because both sources are unbiased. Signal quality is thus captured entirely and simply by a single parameter — "precision", or the reciprocal of its variance — and that has become the accepted

[11] Schlaifer and others used the term "decision analysis" to separate the Bayesian way from the older statistical "decision theory" that traced to Neyman and Wald. See (Gelman *et al.*, 2004, pp. 543–544).

[12] There are literally hundreds of papers built on this assumption, too many to warrant citation.

way that accounting theory and empirical work commonly characterizes information quality.

Dye and Sridhar (2007) raise the potential problem of taking "quality/precision" as "one-dimensional":

> Since the present paper emphasizes the precision of accounting estimates, and precision is a measure of the quality of accounting reports, our paper is also related to the theoretical literature on information quality. While we take quality/precision to be one-dimensional, Antle and Demski (1989) note that the quality of accounting information can often be a multidimensional concept . . . (Dye and Sridhar, 2007, p. 736)

The general consensus seems to be that an express focus on precision, rather than also bias, is part and parcel of trying to sum up, in a practical sense, enough of the relevant quality of the signal to serve the purposes of accounting theory.

Bhattacharya *et al.* (2012) note that their approach focusses on precision because analytical accounting theory does:

> Because the analytical models we rely on tend to focus on the precision of information and to view cash flows as fundamental, we believe our research question calls for accounting-based earnings measures that capture the precision of earnings with respect to accounting fundamentals that are meant to capture the value-generating process of the firm, especially cash flows. (Bhattacharya *et al.*, 2012, p. 458)

There are some models that specifically examine the economic costs, or possibly benefits, of accounting bias, in which case a different Bayesian model appears. Such models, implicitly at least, allow for signal qualities that include bias and precision. An example is a model of a binary signal that can have discretionary Type I and Type II error probabilities, and hence clearly different inherent biases.[13]

[13] In using the terms "Type I" and "Type II" error probabilities, usually denoted by α and β, we risk upending Bayesians who set out to use only their own language.

Models of that binary form can be insightful because of their flexibility; see the discussion later in this monograph. Accounting theory has rightly drawn heavily on the elegance and Bayesian insightfulness of models involving a binary parameter (state) and a binary signal. References include Gox and Wagenhofer (2010), Ewart and Wagenhofer (2011) and Ewart and Wagenhofer (2017), Wagenhofer (2014), Gao (2013a) and Gao (2015), Gao and Zhang (2018), Bertomeu (2013), Bertomeu *et al.* (2017), Gigler and Hemmer (2001), Chen *et al.* (2007), Laux and Stocken (2017) and Smith (2017).

A less likely explanation of accounting information being characterized as more or less "precise" traces to the model of a manager's disclosure in Fischer and Verrecchia (2000). In that model, the direction of suspected bias in a report is unknown, which leaves the mean of the user's distribution unchanged. The unknown bias merely adds a noise term to the disclosure, and leaves it subjectively unbiased. That is like throwing a dart at a moving target. The throws are unbiased but high variance. A "random" bias merely adds variance around the same grand mean.[14]

The second plausible explanation for the literature's reliance on "information precision" is that the term "precision" is not meant to be taken literally or "statistically", but is merely a primitive term used to summarize a composite of possibly vaguer information characteristics. A typical statement, repeated in sentiment throughout the literature, goes as follows:

> A possible explanation for the different stock price response coefficients is that earnings affect investors' beliefs to a greater extent than forecasts because earnings are a more precise signal of future cash flows than forecasts. (Beyer, 2009, p. 1715)

In Bayesian terms, the probabilities α and β are "likelihoods", or probabilities of data given hypothesis (parameter).

[14]Lindley and other subjectivist-Bayesians hold that it is rarely realistic that a user would have no inkling whatsoever. In the case of a manager's unknown direction of bias, there is usually enough surrounding detail and background knowledge to make that bias more likely "up", or more likely "down", or alternatively that it is small or large.

A broadly "better" signal is thus in accounting speak "more precise", where "more precise" is an intuitive notion, and is left at that rather than being given a formal explication.

Information "precision" is thus a shared metaphor in Basu's (2013) description of habits and paradigms in accounting research. Typically, for example, an earnings measure might be understood as "more precise" if it appears less subject to earnings management. That seems to be the final and pragmatic understanding reached by Francis et al.:

> We note that if quality encompasses *any* attribute that makes information more valuable to capital market participants, then a full characterization of information quality is surely empirically intractable. Our focus on precision as the construct that underlies information quality is based on the (previously discussed) applicability of precision to many types of information and on its wide use in the accounting literature. (Francis *et al.*, 2006)

> ... the choice of auditor affects investors' perceptions of the credibility of earnings. ... earnings credibility (which we associate with quality, because more precise earnings should be more believable)... Francis *et al.* (2006)

In what follows, I argue a familiar Bayesian claim, namely that the Bayesian way of thinking supersedes all of the patchwork of ad hoc estimator criteria that frequentist statistics throws up. Bayesian logic avoids the intractable tangles that arise when trying to sum up information qualities in any single frequentist dimension (like "precision").

A Bayesian view of why accounting uses the word "precision" as an overall summary of information quality is that accounting has met a problem that is endemic in frequentist statistics. Bayesian theorists refer to the "adhockery" of frequentist methods, concepts and criteria (that description came from de Finetti; see O'Hagen, 1994, p. 20). Specifically, frequentist textbooks list several estimator characteristics, like bias, precision, consistency, and BLUE-ness, which all seem desirable per se but are not easily weighted against one another, cannot be subsumed into one well-defined measure of signal quality, and do not arise from any unifying principle.[15] Faced with such a bundle or signal characteristics, accounting took up the single word "precision" as the summation of signal quality.

As a last note on the depiction of information quality in the monograph by Francis *et al.* (2006), it stands out that despite references throughout to "decision usefulness", "value relevance" and the "value of information", no explicit Bayesian theory is discussed or referred to. That is a notable departure in terms of exposition, rather than position, from the early works in accounting information theory by Feltham, Demski and others, where the entire opus is explicitly Bayesian and information qualities are described technically–statistically using explicit Bayesian concepts.

A general subordination of a formal Bayesian decision framework is apparent in the common definition of accounting "value relevance" as merely empirical correlation between information and market outcomes (e.g. prices), rather than information influencing outcomes via a mechanism of belief formation and then action:

> The key commonality in the definitions is that an accounting amount is deemed value relevant if it has a significant association with equity market value. (Barth *et al.*, 2001, p. 79)

Bayesian theory came to play a lesser role in accounting than it did in the 1960s and 1970s. The rise of empirical "capital markets" accounting research, which adopted frequentist statistical methods (p-levels) and available software, took attention away from Bayesian theory in PhD programs. The Bayesian critique of significance tests that arose in empirical psychology in that era,[16] was not raised or possibly

[15] Wagenhofer (2011, pp. 232–233), notes the deficiency of any single parameter characterization of earnings quality: ". . . there is a large set of accounting recognition, measurement, and disclosure rules that jointly determine reporting quality. Collapsing these financial reporting choices into a single parameter does not allow generating insights into the delicate mix of instruments a regulator has available."

[16] See Morrison and Henkel (1970). Wagenmakers *et al.* (2018) offers partly psychological explanations beneath empirical researchers' continued loyalty to conventional frequentist methods and language.

widely known. It remains still that while accounting theorists ignore non-Bayesian logic and tolerate only Bayesian logic as the foundation of uncertain inference from limited data or information (see later references), most empirical research method and reporting follows the same unchanged frequentist statistical culture and philosophy as in previous decades.

5.5 Bayesian description of information quality

The alternative to frequentist adhockery offered by Bayesian theory is a single mathematical representation of signal quality — the likelihood function. A clear illustration of how the likelihood function reveals the character of a signal and its innate strengths and weaknesses is provided in Figure 5.1.

Two signals are compared. Both are sample means $\overline{x}$, of amount $\overline{x} = 50$, drawn from a population with unknown mean θ and known standard deviation $\sigma = 50$. The likelihood function of $\overline{x}$ is

$$f(\overline{x}|\theta, \sigma, n) = \sqrt{\frac{n}{2\pi\sigma^2}} \exp\left[-\frac{n}{2}\left(\frac{\overline{x}-\theta}{\sigma}\right)^2\right].$$

The signal depicted by the dotted line is an unbiased sample mean of sample size $n = 5$. The other signal is a sample of size $n = 30$.

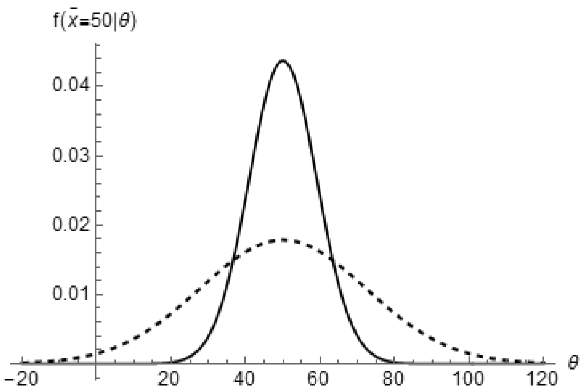


Figure 5.1: Likelihood functions $f(\overline{x} = 50|\theta)$ of two signals.

The idea of the plots is to show how even a quick look at the likelihood function of the data tells much about what the signal says and how strongly it says it. A sharply peaked likelihood function implies strong evidence for parameter values around that peak, and strong support against parameter values for which the likelihood is near zero. An unbiased signal is peaked at the true parameter value, and a larger sample size makes the peak sharper.

Note that the plot may look familiar but is not what is usually seen in a frequentist exposition on estimators. Usually on the horizontal axis would be $\overline{x}$, and the plot would show the probability distribution of $\overline{x}$ centred around some true θ. Our plot instead has θ on the horizontal axis and shows the probability of observing the realized value of $\overline{x}$ (here $\overline{x} = 50$) conditional on different plausible or admissible θ values. The two plots "look" the same but are not. The frequentist version shows a conditional distribution over a range of possible $\overline{x}$, and the Bayesian one shows a conditional distribution over a range of possible θ.

5.6 Likelihood function of earnings

Bayes theorem summarizes information x by its likelihood function of x, $f(x|\theta)$ over the class of all possible θ or states of nature. The location and sharpness of that function feed into Bayes theorem and, jointly with the prior $f(\theta)$, drive the posterior belief distribution $f(\theta|x)$.

Any Bayesian representation of accounting information and its qualities will necessarily be built on likelihood functions and the underlying Bayesian principle of sufficiency. Less formal notions of information properties, like the "asymmetric timeliness" of accountants' readiness to recognize revenues versus expenses, must be expressed in a likelihood function to have a clear Bayesian definition.

As an exercise (see below), the task of translating traditional accounting information qualities into likelihood functions imposes a conceptual rigor on how these properties are understood. Ultimately, they need to be understood in likelihood terms if their "decision relevance" is to be demonstrable within the accepted rational framework for decisions under uncertainty. That level of specification of signal characteristics is the accepted standard in disciplines such as medicine.

Accounting information, and its strengths and weaknesses, can be characterized by Bayesian likelihood functions in ways that formalize many existing accounting concepts. Given the wide importance in accounting of notions like “earnings quality”, the Bayesian way of representing general signal quality through the likelihood function seems to be the natural conceptual approach for accounting to adopt as its theoretical ideal. Desirable “accounting information properties”, including possibly conservatism, will show up in well-specified likelihood functions in ways that imply either a stronger signal or a signal that has higher expected utility to at least a subset of decision makers.

Consider the likelihood function $f(x|\mu)$ of reported earnings x, conditional on “true earnings” μ. Possible subjective likelihood functions, as might occur in the inference model of a user of reported earnings, are shown in the Figure 5.2.

Let the realized earnings be $x = high$, meaning that reported earnings are “high”, and hence the relevant likelihood function is $f(x = high|\mu)$, or really $f(x = high|\mu, K)$ where K is the often important

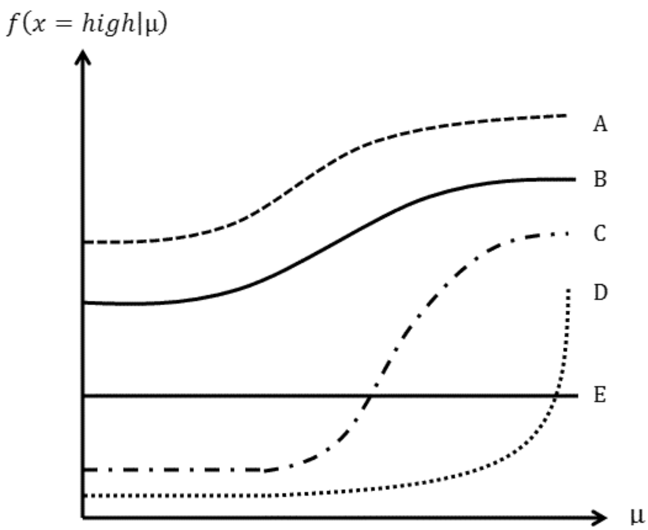


Figure 5.2: Subjective likelihood functions $f(x = high|\mu)$.

background information (K being suppressed for convenience in the notation).

Likelihood function A (i.e. signal A) represents a perception of uniformly less conservative accounting than comparable function B, because it gives higher probability to high reported earnings across the entire spectrum of feasible true earnings; cf. Basu (1997).

Function C is indicative of a type of "conditional conservatism" in that it attributes locally reduced probability to a high earnings announcement when the true earnings are low.

Function D is the most peaked. It implies that the observation $x = high$ amounts to virtual Bayesian proof of high true earnings, because that signal has virtually zero probability of occurring under low true earnings (this is an extreme form of conditional conservatism).

Function E is completely diffuse, indicating that there is something apparently so dubious or deficient about $x = high$ to the point that it carries no subjective weight whatsoever, and thus has no affect on Bayesian beliefs.

There is any number of other possible shapes and a Bayesian user is allowed (indeed required) to specify beliefs about the properties of the earnings announcement, no matter how these are perceived, because without a likelihood function there can be no revision of beliefs upon reported earnings.

This exercise in Bayesian thinking about signal qualities offers a formal way to characterize any aspect of accounting "earnings quality". Many earnings attributes are conceivable and can be pictured in similar functional shapes and locations, including shapes consistent with earnings bias, imprecision and intentional misstatement.

Articulating earnings properties by way of possible likelihood functions brings home the level of specification and flexibility that the Bayesian formalism encourages. A complete Bayesian explication of the perceived "quality" or failings of an earnings measurement regime requires specification over all feasible parameter values μ for all feasible observations x. More conveniently, if earnings can be considered in qualitative levels like "high" and "low", as in the illustration above, the task is simpler. To make Bayesian calculations on observing qualitatively "high" earnings, $x = high$, we need only consider $f(x = high|\mu)$ over the feasible set of μ values.

It is generally not hard to capture the suspected deficiencies in a signal x, even vague feelings of what might be wrong with x. For example, suppose that the user is worried that, although all seems "normal", the reported earnings figure might in fact be fraudulently inflated. This concern is formalized by placing higher personal probability $f(x = high|\mu)$ over the interval of (very) low μ. Another approach is to model $f(x = high|\mu, fraud)$ and $f(x = high|\mu, no\ fraud)$ separately, and then average posterior probabilities over the two possible worlds according to their respective subjective probabilities $f(fraud|x = high)$. This is a standard Bayesian method of allowing for "model risk" or model averaging. Specifically, $f(\mu|x = high) \propto f(\mu)f(x = high|\mu)$ where

$$f(x = high|\mu) = p(fraud|\mu)f(x = high|fraud, \mu)$$
$$+ p(no\ fraud|\mu)f(x = high|no\ fraud, \mu).$$

Bayesian logic has many nice subtleties that can be used to capture well-known accounting characteristics. For example, suppose that an auditor is known to be stringent and rarely allows the sorts of earnings management that boost earnings to a "high" level. With that auditor, a report of high earnings really means something, it rarely occurs. An auditor like this would have a likelihood function resembling D. But there is a downside, because an auditor with this same virtue has the weakness that she reports "medium" earnings most of the time when true earnings are in fact high. So in terms of the likelihood ratio, it can be that the report "medium" is as indicative of true earnings being high as of them being medium, that is,

$$\frac{f(x = \text{``}medium\text{''}|\mu = medium)}{f(x = \text{``}medium\text{''}|\mu = high)} \approx 1.$$

Hopefully the likelihood ratio is greater than one, but it may not be sufficiently greater than one to allow an accounting report of "medium" earnings to help the Bayesian user infer whether true earnings are much more probably medium than high.

The general characteristic of information under Bayes theorem is that different realizations of the same signal (e.g. the same auditor) can carry very different amounts of information.[17] An auditor can be known to be reliable when reporting a highly unfavorable result, but not when reporting a favorable result. That is, a potential Bayesian basis for some results or announcements having a greater effect on the market's beliefs and prices than others. Note that it also shows how an assumption of constant signal precision under all possible true parameter values can be highly unrealistic.

Another helpful feature of Bayesian inference is that in assessing the likelihood function $f(x|\mu)$, it is not necessary that "true earnings" μ are observable. Nor indeed is it necessary that earnings quality is observable. Ultimately, the assessor or user adopts a personal subjective likelihood function $f(x|\mu)$ when forming her Bayesian posterior belief $f(\mu|x)$, and then comes to a decision, the utility or money outcome of which is the only available indicator or "validation" of $f(x|\mu)$.

5.7 Capturing conditional conservatism

Likelihood function F in the first plot below captures a form of "conditional conservatism" in the sense that it becomes more conservative when true earnings are low. At higher earnings it is more lax, and has the same readiness to be "high" as the earnings depicted by likelihood function G.

All sorts of conditional conservatisms are possible in this Bayesian characterization of earnings quality. In the bottom plot in Figure 5.3, I is more stringent about reporting "high" when true earnings are high. However, it becomes more lax than H when true earnings are low.

Drawing likelihood functions for earnings is a good route to Bayesianism. It makes the user think very hard about the earnings signal's statistical properties and provides a mathematically rigorous expression of any number of different signal behaviors.

Note that in these examples, in the simple case where the firm reports either "high" or "low" earnings, the likelihood function of "low" is

[17] The boy who cried wolf lost credibility for his calls of "wolf" because they occurred too frequently when there was no wolf.

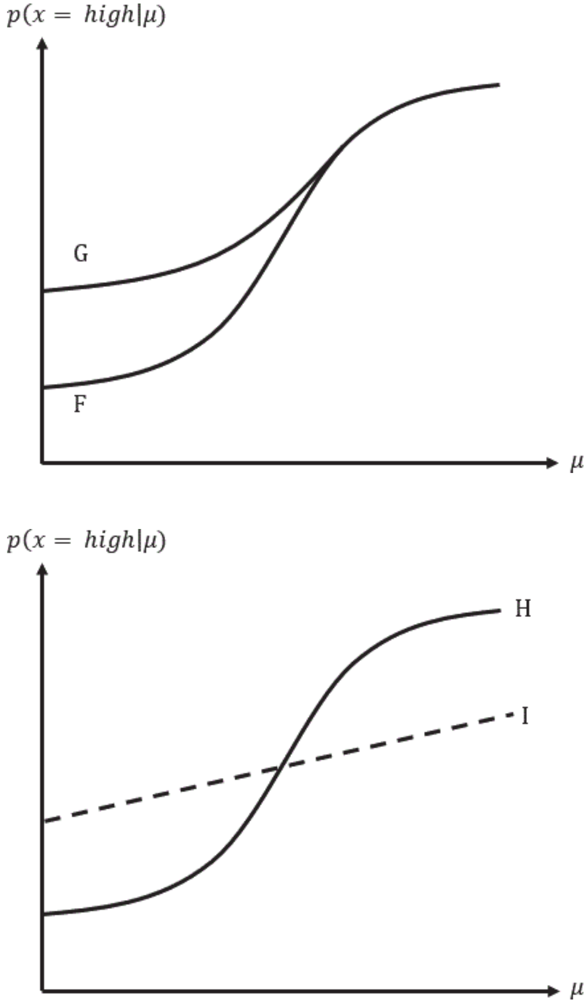


Figure 5.3: Subjective likelihood functions $f(x = high|\mu)$.

implied by the plots we have, because $f(x = low|\mu) = 1 - f(x = high|\mu)$. If alternatively there were say three categories of earnings, we would need to think again to find $f(x = low|\mu)$ and $f(x = medium|\mu)$, although once we have two of the functions, the last is implied.

This exercise in drawing likelihood functions reveals how much more there is to a general Bayesian depiction of earnings quality than the common simplification in accounting, whereby an accounting signal is modelled as an unbiased point estimate with fixed "precision", and where precision is generally not conditioned on (does not change with) the true parameter or underlying state of the firm (which is here denoted by μ). A measure like earnings has many different drivers and is not likely to stay unaffected in either its frequentist precision or bias by the true state of the firm and its underling operations. Ultimately, as a methodological stance in accounting theory, there comes a choice between subjective reality and mathematical tractability. In any field where actual decisions and outcomes hinge on this methodological choice, and knowledge is put to the test in real-world implementation, tractability will usually be readily foregone when reality is clearly better represented by a less easily computed model.

The likelihood function of a normally distributed mean or signal that is assumed unbiased and of known fixed precision is easily written down, but more general depictions of signal characteristics via subjective likelihood functions like those in Figures 5.2 and 5.3 offer a more satisfying articulation of the traditionally understood (good and bad) properties of accounting signals. They allow the assessed degree and direction of signal bias, and precision, to change continuously and simultaneously at different rates over the full domain of feasible population parameter values.

6

Decision Support Role of Accounting Information

Accounting theory has long understood that one role, often the main role, of financial reports is to assist investors and others in their decision-making. Statements to that effect usually have a strong Bayesian flavor. They allude to beliefs, judgments, decisions, uncertainty and sometimes explicitly Bayesian decision-making. The Bayesian decision-analytic standpoint coincides with the FASB conceptual framework, and a large part of both the analytical and empirical accounting literature:

> ... we note that a capital market participant who receives financial reporting information is typically using that information, along with other information, to make a judgment (e.g. about default risk, or about the divergence between an outcome and a prediction) and/or a decision (e.g. about where and how much to invest). Therefore, we take as given that the *primary* purpose of financial reporting information in the capital markets is to support certain judgments and decisions. Financial reporting quality is of interest, then, primarily because of the view that high quality financial reporting information is more decision useful than low quality information. (Francis *et al.*, 2008)

In Bayesian eyes, accounting standard setting is closely akin to "the design of experiments" in statistical theory. To be rigorous, experimental design or ex ante signal optimization must make clear the decision logic presumed of the user, because ex ante information quality, and even the mere concept of "information", is hard to define without reference to at least the main principles of a formal theory of statistical inference and decision-making. It was Feltham (1968) and Demski (1973) who imported a mathematical–statistical way of thinking about experimental design to its natural application in designing and evaluating accounting signals.

The FASB framework is silent about how decision makers reason logically. The framework leaves the decision maker to her own means and tries to define good information properties in an ecumenical wordy way, but not with any firm mathematical or statistical-logical footing. Christensen and Demski (2007) find fault with the FASB's hierarchy of somewhat ad hoc qualitative and sometimes competing information criteria. Their position is that by skirting around a logical (Bayesian) way of understanding decision-making and information qualities, the conceptual framework is essentially unscientific:

> GAAP, as promulgated by regulatory authorities, amounts to regulation of an information source. The conceptual framework, however, substitutes qualitative characteristics of the financial statement information (i.e. the primary qualities of relevance and reliability) for specification of the finer details of the decision or control problem that the user of the financial statements is facing . . . Qualitative characteristics are designed to keep the finer details at bay, to gloss over them, so to speak, by inviting them into the analysis in a reduced form. But this reduced form cannot carry all the essential economic details of the underlying resource allocation exercise. From an economic perspective, reliance on qualitative characteristics is, by its very nature, prone to error. (Christensen and Demski, 2007, pp. 352–353)

There may be some reticence in accounting theory to opening discussion about Bayesian versus frequentist foundations. That would be understandable, given the level of cultural and personal acrimony that the Bayes versus frequentist debate once brought. It is well known in Bayesian legal theory that, in the US, courts and lawyers have been highly resistant to arguments based on explicit Bayesian logic (Fenton *et al.*, 2016). However, logical inference under uncertainty is required, and in Bayesian literature there is a saying after Savage that you "cannot make a Bayesian omelette without breaking Bayesian eggs".

Some accounting papers are manifestly Bayesian without actually making any express statement of that stance. Even Demski's (1973) Bayesian paper, which cites and uses Savage's subjectivist Bayesian philosophy, does not use the word "Bayes". Another telling sign of an apparent standoff-ishness in accounting research is that there are very few citations to any of the Bayesian statistics literature proper. This pattern might go back as far as Blackwell, who Demski, Feltham, Dye and others rely on often. Blackwell's work is now very celebrated as that of a leading Bayesian, but was not originally written up by Blackwell as being "different" or "Bayesian".[1]

6.1 A formal Bayesian model

In a formal and necessarily simplified model, the objective of accounting information x is to assist investors assess a subjective probability distribution

$$f(\boldsymbol{V}|x) \equiv f(V_1, V_2, \ldots |..\ x\ ..),$$

which represents the subjective joint probability distribution of all the individual one-period payoffs V_j of the firms $j = 1, 2, \ldots, N$ in the market.

The simplest yet still insightful model is of a single Bayesian decision maker in a one-period stock market, made up of N correlated assets or firms. This is the common model, and a clear example articulated in Bayesian terms is Lambert *et al.* (2007). A mean–variance model of decision-making under uncertainty is conventional, at least as a starting point. It allows identification of the value-relevant parameters

[1]See Chaloner https://mathalliance.org/wp-content/uploads/2013/12/Bayesian-Statistics.pdf, and references to Blackwell elsewhere in this monograph.

of the belief distribution $f(\boldsymbol{V}|x)$, and therefore yields insights into how information acts via those parameters to alter individual stock prices along with the value (market cap) of the market, as well as the associated price-implied required rates of return, including the overall risk premium on the market portfolio. Johnstone (2017) derives this same model in an explicitly subjectivist decision analysis framework.

6.2 Parallels with meteorology

Meteorology or weather forecasting works with information and models in the face of natural uncertainty. It is a highly evolved statistical discipline, exemplified by its readiness to provide outputs that are crystallized as probability distributions $f(Rain|x)$, just like the distribution $f(\boldsymbol{V}|x)$ presumed of investors in Lambert *et al.* (2007).

Probabilities and probability distributions are very concrete and have measurable accuracy ex post, so fields that express forecasts so clearly and rigorously are open to acute scrutiny of their expertise. If before each day we place a probability on rain, or on the stockmarket going up, it becomes quickly evident whether we have expertise beyond what a naive forecast would say. In the stockmarket, we could simply write down, every day, $p = 0.55$ or the long-run empirical average, but that would be the naive forecast and would not show any fundamental expertise. In meteorology, that forecast is called the "climatological probability" of rain, and forecasting models are evaluated against it and against other models.

Of all fields, the work done by meteorologists can be viewed as "fundamental analysis". Accountants do not usually get so far as to produce explicit probability distributions, although there have been suggestions over the years that this would be a good way to present random or uncertain quantities like sales dollars or asset values. Instead accountants produce numbers that are merely part of the "x" in users probability assessments of $f(\boldsymbol{V}|x)$.

Interestingly, the full expression of Bayesian statistics for decision makers is that a user, when assessing her subjective probability of rain, $p = p(rain|x)$, will condition that probability on all the available information including the weather forecaster's stated assessment of the same probability. As in Demski and Feltham's understanding of Bayesian accounting theory, any signal or "number", like someone else's stated probability assessment, or an accountant's reported depreciation assessment, can be a useful input into the decision maker's final personal (subjective) probability distribution.

Parallels between accounting and meteorology arise often in this monograph and provide insights into how another decision-support profession thinks and assists users. Weather models are advanced sufficiently that forecasts emerge as probability distributions over physical outcomes. Accounting valuation models, as far as they exist, produce "accounting numbers", or point estimates of future-oriented outcomes, with names like "depreciation" and "fair value". The difference in concreteness of the two fields is obvious, as might be inevitable given that one deals in "physical" phenomena, and the other deals more with human processes (like costs, sales and market values).

The potential, nonetheless, for accounting to take insights from another Bayesian field is evident. The following quote from the Bayesian forecasting literature emphasizes the cross-disciplinary nature of the Bayesian toolset, and might whet the appetite:

> Subjective probability forecasting is now well established among meteorologists, particularly in the United States (Murphy and Winkler 1977). Weather forecasters routinely make predictions such as "the precipitation probability for Denver today is 30 percent"; they have also experimented with credible interval temperature forecasts of the form "the probability is 75 percent that today's maximum temperature in Denver will be between 63° and 67°F". The probabilities quoted refer to the forecasters' subjective "degree of belief," given their information at the time of the forecast. This information may include the "objective forecast" output from a climatological analysis, or a computer forecasting system; however, no explicit modeling process need be involved in arriving at forecast probabilities. Such probability forecasting fits neatly into the general Bayesian world-view as conceived by de Finetti

> (1975). The coherent subjectivist Bayesian can be shown to have a joint probability distribution over all conceivably observable quantities. Forecasting then is merely a matter of summarizing the conditional distribution of quantities still unobserved, given current information. In this article we shall, for definiteness, talk mainly in terms of weather forecasting, but it should be understood that the scope of the discussion is much wider, taking in all applications in which a subjectivist makes repeated probability forecasts. (Dawid, 1982, p. 605)

6.3 Bayesian fundamental analysis

The inference-plus-CAPM approach in Lambert *et al.* (2007) depicts a Bayesian framework for "fundamental analysis" truly so-called. It reveals, in one asset-pricing world, how information logically drives market beliefs and market outcomes (specifically firm stock prices and the price-implied costs of capital). A seminal paper for a payoffs understanding of CAPM, and the Bayesian effects of information and information quality according to CAPM, is Coles *et al.* (1995).

The CAPM derivation in Lambert *et al.* (2007) and Johnstone (2016) and Johnstone (2017) rests on a model of investment appraisal and decision-making by individual Bayesian investors. It presumes rational expected utility maximization under conditions where expected utility is a function of just two parameters, payoff mean and payoff (co)variance.[2] Lambert *et al.* (2007) is essentially a Bayesian pro forma for fundamental security analysis under CAPM. Investors deduce from the CAPM the particular statistical payoff or cash flow parameters that they need to consider, and apply Bayesian statistical logic to make inferences about only those value-relevant payoff parameters.

The conceptual advantage of the Lambert *et al.* (2007) approach is that it is built on cash *payoffs* rather than returns. Most related

[2] Johnstone and Lindley (2013) re-examine mean–variance decision-making from the point of view of its logical "coherence", finding ultimately that under severe restrictions, mean–variance finance methods have a direct analogy with expected utility.

finance models are written in terms of returns, for the reason that they will be tested empirically using stock market returns data. Returns expressions are not suited however to understanding the effects of belief distributions about future cash flows or payoffs. The fundamental problem, circumvented by Lambert *et al.* (2007), is that returns are the endogenous results of exogenous random cash flows, or of investors' probability beliefs about largely exogenous future cash flows. The methodological advantage of understanding Bayesian information effects using an exogenous payoffs CAPM are explained by (Coles and Loewenstein, 1988, p. 281) and Clarkson *et al.* (1996).

The market makes cash flow forecasts, in the form of probability distributions, and re-prices the firm accordingly, thus creating an observable return (i.e. a change in price divided by price).

Equivalent Bayesian models of security analysis and investment under uncertainty can of course assume different asset pricing models. For example, investors might be assumed to have log utility or power utility, under which expected utility is maximized over the same asset set at different asset prices, under the same information.

The logical constant, however, in any rational asset pricing model, is Bayesian probability belief revision.

7

Demski's (1973) Impossibility Result

Demski (1973) brought a subjectivist Bayesian end to the notion of "normative" accounting standards, by explaining how, in a Bayesian decision model following Savage (1954), the signal preferred by one decision maker can have quite different statistical error properties to that preferred by another. In effect, different users prefer different biases. Demski's opening statement is so clear and timeless that it is worth quoting at length:

> A primary goal of accounting theory is to explain which accounting alternative should be used (in some particular circumstance). Numerous attempts to develop such a theory have, of course, been offered through the years. Most of these attempts have, in turn, relied on standards, such as relevance, usefulness, objectivity, fairness, and verifiability to delineate the desired alternatives. . . . Moreover, these standards are usually viewed in terms of, or applied to, the accounting measurement process, the environment in which the measurements are taken and/or used, and perceptions regarding that environment. But any such application that is removed from individual preferences — in the slightest

> manner — creates an insurmountable difficulty. In particular, no normative theory of accounting can be constructed using any such set of standards; the standards are bound incompletely and/or incorrectly to rank the accounting alternatives — thus leading to an incorrect or undefined accounting specification. (Demski, 1973, p. 718)

The analysis below exemplifies Demski's general finding, in a common and highly insightful inference model in accounting theory. Part of its mathematical result is that there is no general decision theoretic argument for conservatism in accounting, nor for "neutrality", nor any other apparently normative discrete setting.

7.1 Example: binary accounting signals

Accounting signals in their simplest imaginable form are binary. The signal is "rounded" in its representation of data x to just $+$ or $-$, i.e. "favorable" news or "unfavorable" news in the same sense as Dye and Sridhar (2002) and Verrecchia (1990). News can be about one key item (e.g. whether the firm writes-off a capital expense or carries it as an "asset"), or about accounting earnings, or it can be an overriding accounting summation of firm prospects.

In common models involving binary states of nature, the precision or information qualities of binary signals are represented by their Type I and Type II error probabilities, usually denoted by a pair $\{\alpha, \beta\}$. Models of that form in theoretical accounting literature include Ewart and Wagenhofer (2011), Bertomeu (2013), Gao (2013a), and Gigler and Hemmer (2001) and many others.

In most statistical models, decision makers have the same broad preference for "stronger" evidence. Differences between users arise over the evidential value[1] of a given signal attribute or over how the cost of acquiring information is allotted towards improving the various competing quality attributes of that information.

[1] Remember that in Bayesian theory evidence cannot clearly "speak for itself", and its value usually depends on prior information and surrounding information and beliefs.

An advantage of the binary hypothesis testing model is that statistical signal quality is exhaustively summarized by just two signal attributes, α and β (albeit they must come as a pair, as one without the other does not allow Bayesian probability revision). A difficulty faced in Ewart and Wagenhofer (2017) and Smith (2017) is that there is in general no scalar amalgam of α and β by which to sum up signal quality. That is a methodological research issue for binary signals when compared with the other common accounting setup that assumes an *unbiased* normally distributed signal of given variance. The assumption of zero bias effectively hides a relevant parameter and gives that setup the apparent advantage that it captures signal quality in just a single parameter.

The ultimate appeal of the binary (α, β) signal model is that there is no need to artificially assume an unbiased signal. Instead, any change in (α, β) implies simultaneous changes in the "bias" and "precision" of the signal, and everything else about it too.

The Bayesian binary decision (investment) model set out below leads to a closed-form measure of signal quality that combines α and β and which gives insights into how they interact to make a signal more or less valuable to a given user. Ultimately, however, despite the assumptions made to achieve this theoretical composite of α and β, even it fails as a normative measure of signal quality. The purpose of this example is to show, by making a serious attempt to combine α and β into an objective (user-free) measure of signal quality, that it is not possible to get around Demski's impossibility result.

The investment decision

A single-person investor[2] receives a binary signal $x \in \{+, -\}$ and uses that information to invest in a market containing a single risky asset

[2]In elementary Bayesian decision theory the information user or decision maker and the experimenter, who generates the information, is a single entity. Unification of purpose allows normative insights into raw decision usefulness, which can of course inform noisy rational expectations, game-theoretic, models, where players are assumed to act normatively relative to their own self-interests (i.e. to generate the best possible information for their own decisions). The desire to self-optimize actions and outcomes in a physical way, as in maximizing expected (average) payoffs, is apparently covered by what Gao (2015) calls objective ex post statistical efficiency.

and a risk-free asset. The investment period is short enough that the risk-free rate is negligible. The risky asset is a binary asset paying at termination either S_u when "up" or S_d when "down".

The investor, who trades the underlying asset, is a price taker and can buy or sell that risky asset at its initial market price, S. To preclude arbitrage, $S_u > S > S_d > 0$. The market is regarded as large enough that the single investor's trade quantity has no impact on the market price (as explained in (Christensen and Feltham, 2003, p. 17).

An investor who buys one unit of the risky asset at unit price S effectively risks a net money loss of $(S - S_d)$ in an attempt to win a net profit of $(S_u - S)$. The market implied probability of outcome S_u is then

$$\theta = \frac{S - S_d}{S_u - S_d}, \quad (0 < \theta < 1)$$

and the investor effectively makes a bet against odds of $\theta/(1-\theta)$.[3] Trading the stock can be restated as betting against a market probability of θ, or equivalently as trading an asset (long or short) that will be worth at expiry either $V = 1$ or $V = 0$, and which is priced now at θ.[4]

Note that θ has the outward appearance of a probability but is not necessarily anyone's actual belief and is not Bayesian. It is a "probability" akin to the betting price or odds quoted by a market maker or prediction market. Market makers, like bookmakers, set prices to maximize their own expected utilities, not to indicate their personal beliefs.[5]

---

[3]Note that if there is time to expiry and we wish the opening stock price and the terminal price to be expressed in "constant dollars", the implied probability would be $\theta = \frac{S(1+r_f)-S_d}{S_u-S_d}$, which is often called the "risk-neutral" probability. Note that since r_f is the risk-free rate, $S(1+r_f)$ captures S in period-end dollars.

[4]The replica strategy of buying one unit at price θ is to hold S_d in cash and bet $(S - S_d)$ on "up" against a bookmaker whose probability of "up" is θ. That pair will produce net money amount S_u in the case of "up" and S_d in the case of "down". So by buying an arbitrary number of units of the stock, the investor effectively bets a corresponding arbitrary amount on "up".

[5]Note that the prices observed in prediction markets for 0–1 payoff assets ("Arrow Debreu securities") are widely taken as "market probabilities" and then compared with the actual relative frequencies of the events in question, so as to test the accuracy and calibration of the market as a probability forecaster. This is a technique common in the "wisdom of crowds" literature.

The mechanism and types of information by which the market arrives at θ is not relevant to the individual price-taker's investment decision (trade size). From her perspective, θ is exogenous, which of course simplifies her strategic decision problem. Specifically, it avoids the game theoretic effect that would occur in a smaller market place, whereby her trade size influences her realized trade price. Note importantly that the market price θ is not assumed to behave as a Bayesian. The only presumed Bayesian is the individual investor, who must simply decide how many units of the binary asset to buy (or sell) at market price θ.

The rational investor makes her own probability assessment, $p(V = 1|x)$ and trades against θ. The information x is taken as a private signal in the sense that it is the investor's personal "reading" of what data exists. The investor uses her personal ("private") inference model and all of her public and private background information, combined with her signal x, to form a personal belief about the probability of $V = 1$. Ideally, her inference model or method is "more accurate" than the market, since she wants to "beat the market".

The investor always expects a priori to be "more accurate" than the market, and hence to make a profit, because she views the future through her own beliefs rather than by the true probabilities. In her beliefs, whenever $p(V = 1|x)$ is not equal to market price θ, that price is "wrong" and she expects to profit, otherwise she would not trade.

Two point likelihood function

As in Gao (2013a) and Gao (2015), accounting evidence is akin to a Bayesian hypothesis test. Formally, under Bayes theorem, and based on background knowledge K,

$$p(V = 1|x, K) \propto p(V = 1|K)\, p(x|, V = 1, K).$$

For convenience in notation, K is suppressed. Similarly, there is no express recognition in the notation for the investor's model. Instead the model appears only by its effect on the investor's personal likelihood function

$$p(x|V = 1) \equiv p(x|, V = 1, K).$$

The likelihood function "is" the model, in the sense that it captures all aspects of the model of relevance to Bayes theorem. The user "chooses" her model and hence the likelihood function is a personal assessment. Its "validity" or "accuracy" is usually what separates statistically accurate forecasters or economically profitable traders from other less able competitors, and is of course the only difference between their inferences when they have the same prior and same signal (assuming that they are both Bayesian).

Investors' probability assessments are bound to be affected by the going market price θ, since θ is quoted publicly. Like everything else known to the investor at that time, the going market price θ is included in her inference as part of her background knowledge K. This is practically true of most traders' stockmarket investment decisions – the trader uses whatever information she gleans from the quoted price when deciding, based on all her information, how much to trade at that price.

An investor's probability assessment is unavoidably subjective, and does not have to be close to θ even when both investor and market are strongly influenced by one particular signal such as x. To the contrary, the same observed signal x can have different effects under different investors' models (likelihood functions) and different background knowledge.[6]

Since different investors make different inferences based on the same observed signal, it makes no difference whether signal x is public or private information. We can think of x anyway we like, it can be public accounting news, or perhaps an analyst's interpretation of that news, or it could be the user's own translation of that news, summarized Bayesianly by her personal likelihood function, $f(x|V)$.

To a Bayesian investor, the signal x is characterized by subjective error probabilities, (α, β). The signal is binary $x \in \{+, -\}$ and its perceived likelihood function is a two-point function

$$\alpha = p(x = +|V = 0)$$
$$\beta = p(x = -|V = 1).$$

[6]For example, if the firm predicts earnings of amount h, the investor might take h as a worst case or best case, or perhaps even as managers' "true" estimate, all depending on how the investor models the firms' announcements for their honesty and accuracy. This model is the Bayesian likelihood function, $f(h|V)$.

Define $\alpha = p(+|0)$ and $\beta = p(-|1)$. That labelling of α and β matches Ewart and Wagenhofer (2011) and Gao (2013a) and Gao (2015). In Bayesian statistics, the likelihoods (α, β) are subjective, thus obliging the decision maker to "put her own interpretation" on what is reported, which of course fits naturally with how accounting reports (e.g. earnings) are often, in someone's perceptions, seen as noisy, biased or otherwise unfit to be taken "literally".

Conservatism $\alpha < \beta$

A signal is defined as conservative if it has $\alpha < \beta$. This definition might be justified as follows. Suppose that the accountant runs a test with error characteristics α' and β', where $\alpha' = \beta'$, which makes this test "neutral". But suppose that when the test result is $+$, the accountant goes one further step, just to make sure. She runs the same test again, independently, and she reports $+$ if and only if the second test too results in $+$, otherwise she reports $-$. The error characteristics of the accounting procedure are now

$$\alpha = \alpha'\alpha'$$
$$\beta = \beta'(2 - \beta').$$

Since $\alpha', \beta' < 1$, the combined two-stage test has improved Type I error probability, $\alpha < \alpha'$ but higher Type II error probability $\beta > \beta'$. Also, since $\alpha' = \beta'$, $\alpha < \beta$. Thus, by "double verifying" a positive signal but not a negative signal, the accountant has built a conservative experiment, albeit at the same time harming the test's other error characteristic (its tendency to produce a false negative). The accounting design question is whether that is an "optimal bias".

Conservatism in accounting has the intuitive appeal that it might forestall false hopes, over-confidence and over-trading. The contrary intuition is that to be most accurate or economically valuable, accounting information should be "neutral" ($\alpha = \beta$) rather than mechanically biased or tilted to the negative.

A technical "Demski issue" for accounting is whether, or for what class of decision makers, conservatism makes a better "experiment" in the sense that a more conservative signal increases the ex ante (pre-signal) expected utility of information.[7]

Background on conservatism

There are broadly four modes of thinking about conservatism. First comes an array of different contracting relationships and potential legal conflicts wherein conservatism is seen to assist and arise endogenously. The second set of related arguments is built on a wider class of rational expectations equilibrium models describing the information needs and mutually respective strategic behaviors of stakeholders, who act in individually self-interested anticipation of the others' actions and objectives. The third and oldest approach is qualitative-normative. It emphasizes relatively informal non-mathematical ideas of "decision usefulness" and "information content", possibly subject to costs.[8] Arguments upholding "neutrality" rather than conservatism are motivated from this qualitative inclination against bias, in favor of accuracy or "representational faithfulness", along with a hierarchy of other desirable albeit sometimes fuzzy and apparently competing information attributes (e.g. FASB Conceptual Framework)

The fourth approach, which I illustrate here, is Demski's normative-subjectivist-Bayesian. It views conservatism, divorced from its historical and cultural appeal in accounting, as just another possible signal bias, which, like any other signal characteristic, can add to some decision makers' expected utilities and subtract from others'.[9]

---

[7] An obviously related question, little explored in accounting, is how, or whether, information with higher ex ante expected utility, or better qualities, materializes as larger ex post money outcomes. See the discussion on "economic Darwinism" in this monograph.

[8] Guay and Verrecchia (2006) took up the argument put by Holthausen and Watts (2001) that conservatism can be justified by agency or contracting benefits, regardless of whether the resulting information is "efficient" in normative ways such as catering to equity valuation. Guay and Verrecchia hold, however, that the two perspectives cannot be separated since part of the agency argument overlaps with stakeholder information needs, and hence there must be some consideration of information quality or effiiciency in contracting models.

[9] The strength of this approach is also its weakness. Specifically, the political, cognitive and behavioral process of designing, producing and reporting accounting news has subtleties beyond any single model or methodology. To read a statistical

Impossible Blackwell ranking

Blackwell ranking is understood as an objective criterion in the sense that if signal A is "finer" in Blackwell's terms than signal B, then any Bayesian decision maker, whatever her utility function, prefers signal A to signal B.[10] See Demski (1980, pp. 35–7), Christensen and Demski (2003, pp. 109–111), Christensen and Feltham (2003, pp. 94–100), Dye (1985) and Cabrales *et al.* (2013) for nice examples of how one signal can be "finer" than another.

Unfortunately a binary signal, with $0 < \alpha < 1$ and $0 < \beta < 1$, cannot generally be Blackwell ranked against another such signal with error characteristics $\{\alpha', \beta'\}$. The best that can be done is to restrict the admissible tests to those with $\alpha, \beta < 0.5$, in which case we can say that a signal with either $\{\alpha' \leq \alpha, \beta' < \beta\}$ or $\{\alpha' < \alpha, \beta' \leq \beta\}$ outranks $\{\alpha, \beta\}$. Intuitively, by reducing either α or β, while holding the other fixed, the signal has less noise. That is essentially what the Blackwell criterion requires.

Unfortunately, however, it is not possible to Blackwell rank two binary signals (each with $\alpha, \beta < 0.5$) for which one has lower α and the other has lower β. That ranking comes down to who is the user, in the way understood by Demski (1980).

Updating beliefs

Bayes theorem is written insightfully following Berger and Wolpert (1988) in terms of the relevant likelihood ratio

$$p(1|x) = \left[1 + \left(\frac{p(0)}{p(1)}\right)\left(\frac{p(x|0)}{p(x|1)}\right)\right]^{-1}.$$

Assume that the investor sets the quoted market price θ as her prior probability, that is, her assessment is $p(1) = \theta$. Her possible posterior

decision model into this process requires a great amount of abstraction (read simplification). Its justification comes only by its theoretical statistical insights into what makes information useful and what accounting motivations arise as a result.

[10] See DeGroot (1986) for background on Blackwell and his Bayesian decision theory.

probabilities are then

$$p(1|+) = \left[1 + \left(\frac{1-\theta}{\theta}\right)\left(\frac{p(+|0)}{p(+|1)}\right)\right]^{-1} = \left[1 + \left(\frac{1-\theta}{\theta}\right)\left(\frac{\alpha}{1-\beta}\right)\right]^{-1}$$

$$p(1|-) = \left[1 + \left(\frac{1-\theta}{\theta}\right)\frac{p(-|0)}{p(-|1)}\right]^{-1} = \left[1 + \left(\frac{1-\theta}{\theta}\right)\left(\frac{1-\alpha}{\beta}\right)\right]^{-1}.$$

Note that $p(1|+) > p(1)$ and $p(1|-) < p(1)$ for all $\alpha, \beta < 0.5$, implying that certainty can naturally increase or decrease, depending on what signal is observed, not merely on its error properties.

Information x affects posterior beliefs only via its likelihood ratio $\frac{p(x|0)}{p(x|1)}$, which is the minimal sufficient summary of the evidence carried by x. It is sufficient in the sense that nothing otherwise is relevant, and is minimal in the sense that it cannot be reduced without loss of relevant information.

The probability under given error characteristics (α, β) of observing $x = +$ is

$$p(+) = \theta(1-\beta) + (1-\theta)\alpha,$$

and the probability of observing $x = -$ is

$$p(-) = \theta\beta + (1-\theta)(1-\alpha).$$

There is no point in manipulating $p(+)$ or $p(-)$ merely to produce “more positives” or “more negatives”. The frequencies or probabilities with which these occur affect their evidential meaning. If a source produces more negatives (positives) by combining higher β (α) with unchanged α (β), all that is achieved is that both negative and positive signals carry less Bayesian weight. Thus, a negative signal brings a smaller downward revision in $p(V = 1|\cdot)$, and a positive signal brings a smaller upward revision in $p(V = 1|\cdot)$.

Finding the “best” pair (α, β) amounts to balancing the frequencies with which positives and negatives occur against the evidential weight that each of these carries when it does occur. Lambert (2010) notes the same inverse relationship between the propensity to report “good news” and its evidential credibility. See also Ewart and Wagenhofer (2011, p. 147).

Optimal investment

Assume that the cost to produce a signal with error probabilities (α, β) is $C(\alpha, \beta) > 0$. The wealth available for investing after having paid for x is

$$W = W_0 - C(\alpha, \beta)$$

where W_0 is the investor's initial wealth. Note that wealth at the time of choosing (α, β) is W_0 and wealth at the time of investing (after having paid for and viewed signal x) is W. This distinction matters for any utility function where the chosen investment, conditioned on x, depends on the wealth existing at the time of investment, by which point the preceding cost of obtaining x, $C(\alpha, \beta)$, is sunk.

The money amount invested by the information user is conditioned on signal x and is decided with respect to the asset's price θ. If an investor bets fraction f of her wealth on outcome $V = 1$, she buys $f\,W/\theta$ units of the asset (i.e. she buys $f\,W/\theta$ units of the "$V = 1$" contract).

By betting fraction f of wealth on $V = 1$, the possible money payoffs are

$$\begin{array}{ll} W(1 - f + f/\theta) & \text{if } V = 1 \\ W(1 - f) & \text{if } V = 0. \end{array}$$

The only common class of utility functions for which analytical results regarding optimal (α, β) are tractable is the CARA exponential utility function $U(w) = 1 - \exp[-c\,w]$ for wealth w. Numerical results based on other common utility families are, however, easily obtained, and are unchanged in all relevant aspects from those found below under exponential utility.

To find the optimal investment fraction f under exponential utility and probability $p = p_{(\alpha,\beta)}(1|\cdot)$, we maximize expected utility

$$p\,(1 - \exp[-cW(1 - f + f/\theta)]) + (1 - p)\,(1 - \exp[-cW(1 - f)]), \quad (7.1)$$

Differentiating Equation (7.1) with respect to f and solving the first order condition gives optimal betting fraction

$$f \equiv f(p) = \left(\frac{\theta}{c\,W}\right) \left\{\log\left(\frac{p}{1-p}\right) - \log\left(\frac{\theta}{1-\theta}\right)\right\}. \qquad (7.2)$$

Note that f is negative when $p < \theta$. The rational investment is then the equivalent of selling (rather than buying) $|f|$ worth of "$V = 1$ contracts" at unit price θ.

When $x = +$, the optimal fraction is $f^+ = f(p^+)$, where $p^+ = p_{(\alpha,\beta)}(1|+)$ denotes the posterior probability of $V = 1$ under signal $x = +$. Similarly when $x = -$, the optimal fraction is $f^- = f(p^-)$, where $p^- = p_{(\alpha,\beta)}(1|-)$.

From Equation (7.2), the optimal number of contracts given belief p, is

$$f\frac{W}{\theta} = \frac{1}{c}\left\{\log\left(\frac{p}{1-p}\right) - \log\left(\frac{\theta}{1-\theta}\right)\right\},$$

and is independent of investor wealth.

Investment fractions

The optimal investment depends on the log likelihood of the information x, $\log\left[\frac{p(x|1)}{p(x|0)}\right]$. No other test attribute or information characteristic affects the user's f. Combining Bayes theorem, written in log odds form,

$$\log\left(\frac{p(1|x)}{p(0|x)}\right) = \log\left(\frac{\theta}{1-\theta}\right) + \log\left(\frac{p(x|1)}{p(x|0)}\right),$$

with Equation (7.2) gives the two investment fractions

$$f = \begin{cases} f^+ \equiv f(p^+) = \left(\dfrac{\theta}{cW}\right)\log\left(\dfrac{1-\beta}{\alpha}\right) \\ \\ f^- \equiv f(p^-) = \left(\dfrac{\theta}{cW}\right)\log\left(\dfrac{\beta}{1-\alpha}\right). \end{cases}$$

The user's fraction f of wealth bet on outcome $V = 1$ upon observing a positive (negative) signal is (i) decreasing (increasing) in both α and β, (ii) increasing in the prior probability θ, (iii) decreasing in wealth (DRRA) and (iv) decreasing in the user's coefficient of absolute risk aversion c. The physical number of units of the $V = 1$ contract bought by the CARA investor at unit price $\theta = \Pr(1)$ is simply

$$\begin{cases} \dfrac{1}{c}\log\left(\dfrac{1-\beta}{\alpha}\right) & \text{if } x = + \\ \dfrac{1}{c}\log\left(\dfrac{\beta}{1-\alpha}\right) & \text{if } x = -, \end{cases}$$

implying that the user is driven by just her risk aversion c and the information brought by x, where information is measured by the change in log odds

$$\log\left(\frac{p(1|x)}{p(0|x)}\right) - \log\left(\frac{p(1)}{p(0)}\right).$$

Note that a positive test result $x = +$ can be described conventionally, in frequentist terms, as "significant at α", but the optimal investment conditioned on that description of x is indeterminate because it embodies only one side, α, of the required likelihood ratio of $x = +$, viz. $(1-\beta)/\alpha$. That is an example of how frequentist tests do not meet the needs of economically rational decision makers.

Expected Utility

The trader can either buy or sell, that is, she can bet for or against $V = 1$. The investor's unconditional expected utility after observing x and making the appropriate investment based on x is

$$EU_{(\alpha,\beta)} = p(+)EU^{+} + p(-)EU^{-}. \tag{7.3}$$

where EU^{+} is the expected utility conditional on $x = +$ and EU^{-} is the expected utility when $x = -$, that is

$$\begin{aligned} EU^{+} &= p(1|+)(1 - \exp[-(W(1 - f^{+} + f^{+}/\theta))]) \\ &\quad + p(0|+)(1 - \exp[-(W(1 - f^{+}))]), \end{aligned}$$

and

$$\begin{aligned} EU^{-} &= p(1|-)(1 - \exp[-(W(1 - f^{-} + f^{-}/\theta))]) \\ &\quad + p(0|-)(1 - \exp[-(W(1 - f^{-}))]). \end{aligned}$$

Writing all the various probabilities and associated betting fraction f in terms of the prior $\theta = p(1)$ and the error characteristics (α, β), the full expression for EU from Equation (7.3) is

$$\begin{aligned} EU_{(\alpha,\beta)} &= 1 - \exp\left[-c\left(W_0 - C(\alpha,\beta)\right)\right] \\ &\quad \times \left\{\alpha\left(\frac{1-\beta}{\alpha}\right)^{\theta} + (1-\alpha)\left(\frac{\beta}{1-\alpha}\right)^{\theta}\right\}. \end{aligned} \tag{7.4}$$

Assuming market uncertainty $0 < \theta < 1$ throughout. For any given starting wealth W_0 and fixed information cost $C(\alpha, \beta)$, expected utility is maximized at the (α, β) pair which minimizes the constant

$$\Delta = \alpha \left(\frac{1 - \beta}{\alpha} \right)^\theta + (1 - \alpha) \left(\frac{\beta}{1 - \alpha} \right)^\theta .$$

It is easily found that for $\alpha, \beta < 0.5$ and fixed θ, the derivative $\delta\Delta/\delta\alpha$ is higher when either α or β is lower. Hence, the marginal improvement in signal quality (marginal reduction in Δ) brought by lower α is greater when either α is already lower, or β is lower. Similarly, the benefit from lower β is enhanced when either β is already lower, or α is lower.

Counter-intuitively perhaps, and unlike most returns in economics, marginal "returns" from reducing α or β (or both) are therefore increasing (ignoring cost).[11] It is for findings like this that a formal Bayesian decision theory of accounting information is useful. There are mathematical insights that come from formal models that are not evident intuitively, without that structure. Demski and others called on Bayesian decision theory so as to gain insights into how accounting information can better serve investors.

Buyer or seller

The ex ante economic "decision value" of signal x is captured from the perspective of every investor with CARA utility, regardless of personal absolute risk aversion c, by the same well-defined function Δ of the two error characteristics, α and β. For any fixed signal cost, expected utility is (linear) decreasing in Δ, so Δ is the CARA measure of signal "paucity".

Demski's result holds nonetheless, for several reasons. CARA utility investors all measure signal paucity ex ante (i.e. pre-signal) by the same constant Δ, however their different levels of personal risk aversion c cause them to arrive at different expected cost–benefit assessments of any given (α, β) pair. Some would prefer to spend more, or less, for

[11] This result has a parallel in Bickel and Smith (2006) where information gathering can show increasing marginal returns.

different (α, β). That holds even when they all have the same cost $C(\alpha, \beta)$ for any given (α, β) pair. Further, the model assumes that investors are open to going long or short, based on the information received, which will not realistically be true.[12] The model also assumes that the trader starts with zero inventory and strictly cash wealth. If an investor already has a large inventory in the asset, or a related side bet or hedge, her investment on a given signal will be greatly altered.

Conservatism as an optimal bias

The first-order conditions of Equation (7.4) are not solvable for closed form optimal α and β, however by simple inspection of (7.4) it can be seen that:

(i) The unique case of $\theta = 0.5$ implies symmetry between the possible losses from buying and selling and the ex ante optimal signal is unbiased, $\alpha = \beta$.

(ii) For $\theta < 0.5$ the optimal signal pairing (α, β) has $\alpha > \beta$, and, for $\theta > 0.5$ (by symmetry), $\alpha < \beta$. Hence, conservatism $\alpha < \beta$ is unlikely to be a "desirable bias" other than when the possible loss from buying the asset, namely θ, exceeds the possible loss from short selling, namely $1 - \theta$. The opposite to conservatism, $\beta > \alpha$, might potentially be desirable when the possible money loss is higher when short selling, that is, when $\theta < 0.5$ Consideration of the relative amounts of the possible losses arises also in a model developed by Bertomeu *et al.* (2011, p. 862).

(iii) Consistent with (ii), $d\Delta/d\alpha > d\Delta/d\beta$ for all $\alpha = \beta$ when $\theta > 0.5$. Thus, if $\alpha = \beta$, the marginal expected utility gained by reducing

[12] There seems to be relatively little discussion in the literature of how conservative accounting practices might privilege stock buyers over sellers. Losses incurred in the specialist activity of short selling are likely to have been largely overlooked in the evolution of accounting practices. Note however that the FASB/IASB Conceptual Framework states explicitly that investment "decisions involve buying, selling or holding", and appears to view buyers and sellers as equal stakeholders when designing information of value relevance or decision usefulness.

α is greater than by reducing β. Similarly, $d\Delta/d\alpha < d\Delta/d\beta$ for all $\alpha = \beta$ when $\theta < 0.5$, making it better to reduce β.

These results (i)–(iii) rest on the assumption that α and β have the same cost. That is, the marginal cost of reducing either error probability is always the same.

It follows that a decision-theoretic justification for conservatism requires the equivalent in stock trading of $\theta > 0.5$. Put into words, $\theta > 0.5$ implies that the possible money loss from a one-unit buy is higher than the possible money loss from a one-unit short sale, or that "stocks go down by greater amounts than they go up".[13] That condition lacks any obvious justification, empirical or theoretical, but is the "Demski-like" decision-theoretic consideration for standard setters when designing information or "experiments" that cater respectively to buyers and sellers.

Christensen and Feltham (2003, pp. 94–100) explained how "finer details" of the decision problem, such as the investor's opportunity set, risk aversion and so on, alter what makes the best "experiment". To that list we can add the differences between investors motivations and circumstances listed above. An interesting aspect of this illustration of Demski's result is that even in the unique case where the signal error characteristics α and β can be combined into a parsimonious scalar, there is still no agreement between even that narrow class of investors with exponential utility about the best (α, β) signal pair.

Numerical example

Since there are no closed-form results for optimal (α, β), differences between optimal α and β are illustrated by numerical optimization. Signal cost is taken as

$$C(\alpha, \beta) = \left(\frac{1}{k\alpha} + \frac{1}{k\beta}\right). \quad (k > 0)$$

[13] Guay and Verrecchia (2006) note a similar asymmetry in a debt holder's view. There is no upside benefit to a debt holder when the firm becomes more successful but a large downside loss if the firm fails.

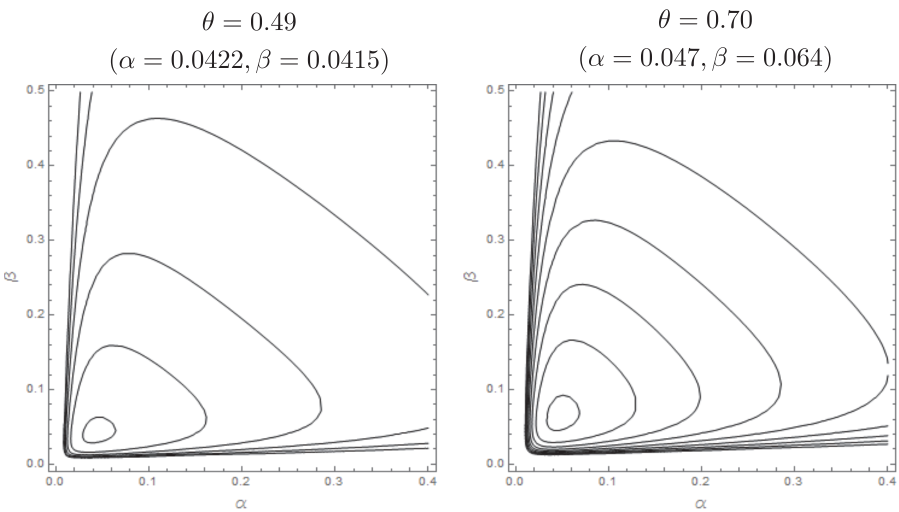


Figure 7.1: Contour plots of $EU \equiv EU_{(\alpha,\beta)}$.

To show how the optimal bias changes with θ, two sets of results are plotted. The calculations assume, for example only, that $W_0 = 10$ and, $c = 1$ and $k = 100$.

Figure 7.1 shows contour plots of the relevant EU function (7.4) and numerical optimization gives the best ex ante test characteristics (α, β) in each case. EU is plotted as a function of signal error properties α and β.

The contour plots show equi-EU lines and the highest EU is in the "eye" of the plot in both cases. Note that with $\theta = 0.49$, we find $\alpha > \beta$ but the gap between α and β is tiny because θ is so close to 0.5. The optimal excess of α over β increases as $\theta \to 0$ because the potential loss of $(1 - \theta)$ from short selling increases. The second plot shows how the optimal pairing has $\alpha < \beta$ when $\theta > 0.5$, and the gap increases with higher θ.

Optimal bias

A revelation in the Demski information economics model that seems to defy common sense and offend primitive notions of "professional independence", "impartiality" or "neutrality" in normative accounting measurement is that the rational user cum investor might, of her own choice, prefer a "biased" (e.g. conservative or aggressive) signal.

The ideal signal in the frequentist world is a measure of the "true probability", or maybe even a 0 or 1 that corresponds exactly to the eventual true outcome. Setting those conceptions of objective accuracy aside, as untestable or impractical, the Bayesian model calls for a practical approach of biasing or optimizing the user's signal so as to probably avoid the kinds of errors in beliefs (in either direction or magnitude) that would probably do most harm, on an ex ante expectations basis. According to this expression of Demski's point, the only possible "neutrality" between individual users is to provide them all with statistically worse signal error characteristics than they would choose for themselves.

Ultimately, there is no objective measure of probability accuracy. There are different possible "scoring rules" that calculate concepts of "distance" between a probability like $p(E = 1) = 0.725$ and an actual outcome like $E = 1$, but there is no one such rule (Bickel, 2007). Scoring rules all capture "distance" according to different criteria (e.g. a differential entropy score might be $\log(\frac{1}{0.725})$). Cabrales *et al.* (2013) support the log or entropy score, but there is no objectively best measure of information.[14] Some scoring rules equate to positive linear transformations of a utility function, and in essence measure the utility that the investor would have obtained had the event been $E = 1$ and the investor's belief been 0.725 (i.e. how much would she have bet and what utility would that bet have produced).[15] An economic decision maker would regard such a score function as an economic measure of the "accuracy" of the belief, and would hope for probabilities that

[14]If there were, Demski's result would not hold as the normatively best signal would be the one that offers highest expected log score or highest expected reduction in entropy.

[15]Savage (1971) interpreted the Brier score economically as a share of the firm. Murphy (1966), who used de Finetti's notion of a scoring rule in weather forecast evaluation, showed that when the "cost-loss ratio" (a common measure in meteorology) is uniformly distributed, the Brier score is a measure of the expected loss (in money) from acting on the forecast probability so as to minimize expected loss.

are "accurate" or do well by that very measure. That is the ex post perspective explained later in this monograph.

7.2 Conservatism and the user's risk aversion

The natural presumption is that a more risk-averse investor desires a more conservative brand of accounting, taking "conservative" as meaning generally less prone to overstate the firm's profits. That presumption is generally not correct (see Birchler and Bütler, 2007, pp. 49–50; Johnstone, 2011).

If there is more than one signal combination of Type I and Type II error probabilities with the same cost, there are situations where a more risk-averse decision maker prefers a signal balanced less conservatively between the two error types. Johnstone (2010) illustrates situations where information properties bringing a marginal increase in the probability of a favorable payoff has higher marginal certainty equivalent to a more risk-averse investor. That is, a more risk-averse investor can be ready to pay more for information with the potential to bring about higher confidence in the investment succeeding. Intuitively, a most risk-tolerant investor needs little convincing and has high expected utility with relatively little information, making it hard for further information to offer much marginal expected utility.

8

Does Information Reduce Uncertainty

The conventional wisdom is that accounting information can, or should resolve uncertainty. Surely anything that qualifies as "information" must make some unknown more known than it was? Or does information sometimes merely add confusion or raise new grounds for concern?

8.1 Beaver's (1968) prescription

An early prescription of the Bayesian position in accounting literature was put by Beaver (1968) in his formative paper on the evidential value or "information content" of accounting earnings announcements.[1]

Beaver held that earnings have information content only if they change market beliefs about future firm payoffs. He added to this definition the proviso that, from an instrumentalist point of view,[2] information has content only if it changes beliefs so much as to cause investors to reverse or modify their investment actions:

> ...a firm's earnings report is said to have information content if it leads to a change in investors' assessments of

[1]The Beaver paper and its relevant footnote were pointed out to me by Sudipta Basu (private communication).

[2]See Feltham (1968, p. 690) on this instrumentalist definition of information.

> the probability distribution of future returns (or prices)... Another definition of information states that not only must there be a change in expectations but the change must be sufficiently large to induce a change in the decision-maker's behavior. (Beaver, 1968, pp. 68–9)

In a footnote, Beaver insists that information does not always reduce uncertainty, and should not be defined as if it must. To the contrary, he explains in Bayesian spirit that information can bring more or less certainty about an unknown state, depending on what exactly that information says:

> As a final parenthetical comment, note that reduction of uncertainty was not one of the definitions chosen. It should be apparent that in a dynamic situation (i.e. where probability distribution assessments are changing over time), a decision maker might be more uncertain about a given event after receiving a message about the event than he was before he received the message. (Beaver, 1968, pp. 69)

8.2 Bayesian basics

Beaver's Bayesian position is correct, as the following simple example shows. If the uncertain quantity is say θ, which might be the period-end state of the firm, then x is "relevant" or "informative" with respect to θ if $p(\theta|x) \neq p(\theta)$. "Information" is any signal or indication (e.g. what is not said) that leads the investor to change beliefs about θ.[3]

Consider the unknown of whether a firm is a "going concern", described as G, or not. Suppose that the investor's prior probability is $p(G) = 0.8$. Now consider a signal x that can take one of two values, say $x = x_1$ or $x = x_2$. This signal has known "likelihoods" or error probabilities $p(x_1|G) = 0.6$ and $p(x_1|\textit{not } G) = 0.8$. By observing $x = x_1$,

[3]Strictly, information x is informative with respect to pre-existing information ϕ if $p(\theta|\phi \cap x) \neq p(\theta|\phi)$.

the investor's posterior probability becomes

$$\begin{aligned} p(G|x_1) &= \frac{p(G)p(x_1|G)}{p(x_1)} \\ &= (0.8 \times 0.6)/(0.8 \times 0.6 + 0.2 \times 0.8) = 0.67. \end{aligned}$$

The investor is now less certain that the firm is a going concern, since $p(G|x_1) < p(G) = 0.8$, i.e. $p(G|x_1)$ is closer to 0.5, where uncertainty is maximum.

In another expression of the same Bayesian thinking, the observation x_1 is more likely if the firm is *not* a going concern, than if it is. This ratio, known as the likelihood ratio, is

$$\frac{p(x_1|G)}{p(x_1|not\ G)} = 0.6/0.8 = 0.75,$$

which is less than one, implying that the evidence favors or points to the firm not being a going concern. In simple terms, the evidence in x_1 is counter to what was initially believed, and hence makes that belief less strong. Conversely, if the observed signal had been the confirmatory signal x_2, then the posterior is

$$p(G|x_2) = p(G)p(x_2|G)/p(x_2) = 0.8(0.4)/0.36 = 0.89,$$

thus making the user more certain of the firm being a going concern.

An insightful way to think about these calculations is to remember that today's probability is today's expectation of tomorrow's probability. Imagine that signal x has not yet been observed. Rather, suppose that the firm will report tomorrow, and will report either $x = x_1$ or $x = x_2$. The probability today of observing x_1 tomorrow is thus

$$\begin{aligned} p(x_1) &= p(G)p(x_1|G) + p(not\ G)p(x_1|not\ G) \\ &= 0.8 \times 0.6 + 0.2 \times 0.8 = 0.64. \end{aligned}$$

Today's expected value of tomorrow's probability is therefore

$$\begin{aligned} &p(x_1)p(G|x_1) + p(x_2)p(G|x_2) \\ &\quad = 0.64 \times 0.67 + 0.36 \times 0.89 = 0.8, \end{aligned}$$

which is today's probability, $p(G)$, but today the effect on certainty of tomorrow's information is unpredictable.

Information must sometimes increase the probability in question and sometimes reduce it. That must be so, because if we know that tomorrow's probability will necessarily be lower, we will have already arrived at a lower probability.[4]

8.3 Contrary views in accounting

The position put by Beaver brought no argument at the time, and is unquestionably the Bayesian position, but it is hard to resist the natural intuition that information should — by its being "informative" — reduce uncertainty. Thinking along those lines, accounting theorists have often posited that accounting information disclosures, like earnings announcements, can reduce stock market investors' assessments of risk and thereby reduce the market's imposed cost of capital:

> ... greater disclosure reduces estimation risk arising from investors' estimates of the parameters of an asset's return or payoff distribution. That is, greater uncertainty exists regarding the "true" parameters when information is low. (Botosan, 1997, p. 324)

> ... we show (not surprisingly) that higher quality information reduces the assessed variance of a firm's cash flow. (Lambert *et al.*, 2007, p. 387)

> Private information about systematic factors affects risk premiums by resolving uncertainty... (Hughes *et al.*, 2007, p. 706)

> ...more disclosure reduces the uncertainty about firm value, which in turn reduces the potential information advantage

[4]It is interesting therefore that we can sometimes know today that the cost of capital conditioned on tomorrow's signal will certainly be lower, or certainly be higher, than the cost of capital conditioned only on what we know today. See Johnstone (2015) for proof.

> that an informed trader might have. (Leuz and Wysocki, 2008, p. 7)

> Releasing more information and, in particular, more public information through financial reports and other public disclosures by firms reduces the uncertainty about the size and the timing of future cash flows and, therefore, also the risk premium. ...if the forthcoming public report will be more informative, then more uncertainty will be resolved once the report is released. (Christensen *et al.*, 2010, pp. 817–818)

There are many analogous statements in the contemporary accounting literature, and the position taken so succinctly by Christensen *et al.* (2010) has been endorsed by financial regulators:

> Numerous academic studies have concluded that more information in the marketplace lowers the cost of capital. Upon reflection ...academic studies are not really necessary to reach this conclusion — it is intuitive. More information always equates to less uncertainty, and it is clear that people pay more for certainty. Less uncertainty results in less risk and a consequent lower premium being demanded. In the context of financial information, the end result is that better disclosure results in a lower cost of capital. (Foster, 2003, p. 1)

8.4 Bayesian roots in finance

The roots of how accounting lost sight of Beaver's direct instructions lie in the remarkably advanced Bayesian theory of portfolio investment that arose in finance in the 1970s.[5] That highly instructive literature is often cited cursorily in accounting empirical papers as providing the theoretical base for claims that more data/information should reduce uncertainty and the risk premium imposed by investors (e.g. Botosan

[5] See Kalymon (1971), Klein and Bawa (1976) and Klein and Bawa (1977), Barry (1978) and Barry and Brown (1985).

and Plumlee, 2002; Bhattacharya *et al.*, 2012, p. 455; Core *et al.*, 2015, Leuz and Wysocki, 2016). A review of this estimation risk literature related to accounting information is provided by Artiach and Clarkson (2011).

Rather than simply "plugging" into portfolio optimization models the empirical frequentist point estimates of each firm's mean return and the covariance matrix of returns — as if those finite sample estimates were the "true" returns parameters — Bayesian portfolio theory adopted the subjectivist method of forming prior distributions for the unknown returns parameters and updating those distributions from historical empirical data. The technical result that recurs over much of the Bayesian "estimation risk" literature in finance is that by admitting formally that the parameters of the joint returns or payoff distribution are uncertain, the probability distribution of next period's return is revealed to have two sources of uncertainty: (i) even after collecting data, there remains posterior uncertainty about the distribution mean, (ii) there is innate variance in the population around its true mean, so even knowing that true mean would not bring certainty.[6]

The general effect of the Bayesian "parameter uncertainty" literature set out in the seminal papers cited in accounting is that uncertainty is understood as being:

(i) higher, merely by being made Bayesian, due to the formal recognition of parameter uncertainty,

(ii) decreasing in the amount of data or returns sample size n (samples are assumed unbiased).

The effect (i) of "going Bayesian" disappears only when the sample size $N \to \infty$, since with that much evidence the empirical parameter estimates are also the certain Bayesian beliefs (making the big assumption that the model is correct). However, more appealingly, the related result (ii) implies that uncertainty is reduced with better/more information. It was by (ii) that accounting claimed that better financial information should reduce investor uncertainty.

[6]In more realistic models, the population returns variance is also treated as unknown, adding further uncertainty.

Background

It will help to recap how the above results (i) and (ii) arose in the literature. The following is a simplified summary of Kalymon (1971) who introduced explicit Bayesian mean–variance portfolio theory. In Kalymon's model, the uncertain variables of interest are stock returns.

Assume that one-period returns $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)$ of n firms in the market are joint normal $\boldsymbol{r}|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with an $n \times 1$ vector $\boldsymbol{\mu}$ of means and an $n \times n$ covariance matrix $\boldsymbol{\Sigma}$. To introduce parameter uncertainty, imagine somewhat artificially that $\boldsymbol{\Sigma}$ is known but $\boldsymbol{\mu}$ is unknown. The uncertainty about $\boldsymbol{\mu}$ is represented by a joint normal prior distribution $\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$. In addition to prior information, the market observes an iid random sample $\boldsymbol{y}$ of N observations on firm returns $\boldsymbol{r}$

$$\boldsymbol{y} = \{(r_1, r_2, \ldots, r_n)_1, (r_1, r_2, \ldots, r_n)_2, \ldots, (r_1, r_2, \ldots, r_n)_N\}.$$

By a standard result presented in Bayesian textbooks, the posterior distribution $\boldsymbol{\mu}|\boldsymbol{y}, \boldsymbol{\Sigma} \sim N(\boldsymbol{\mu}_N, \boldsymbol{\Lambda}_N)$ is joint normal with mean vector

$$\boldsymbol{\mu}_N = \frac{\frac{1}{\boldsymbol{\Lambda}_0}\boldsymbol{\mu}_0 + \frac{N}{\boldsymbol{\Sigma}}\overline{\boldsymbol{y}}}{\frac{1}{\boldsymbol{\Lambda}_0} + \frac{N}{\boldsymbol{\Sigma}}}$$

and known precision

$$\frac{1}{\boldsymbol{\Lambda}_N} = \frac{1}{\boldsymbol{\Lambda}_0} + \frac{N}{\boldsymbol{\Sigma}}.$$

Hence, (i) the posterior mean of unknown $\boldsymbol{\mu}$ is a weighted average of the prior mean and the sample mean, where the respective weights are the prior precision matrix $1/\boldsymbol{\Lambda}_0$ and the sample precision matrix $N/\boldsymbol{\Sigma}$, and (ii) the posterior precision of the belief distribution for $\boldsymbol{\mu}$ is the sum of the prior and sampling precisions.

Having found the posterior distribution of uncertain parameter $\boldsymbol{\mu}$, it remains to form a probability distribution $f(\widetilde{\boldsymbol{r}}|\boldsymbol{y}, \boldsymbol{\Sigma})$ for the next random observation $\widetilde{\boldsymbol{r}}$. This is the posterior predictive distribution and is found by

$$f(\widetilde{\boldsymbol{r}}|\boldsymbol{y}, \boldsymbol{\Sigma}) = \int f(\widetilde{\boldsymbol{r}}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) f(\boldsymbol{\mu}|\boldsymbol{y}, \boldsymbol{\Sigma}) d\boldsymbol{\mu},$$

following the general Bayesian approach of integrating out nuisance parameters.

By another standard Bayesian result, this distribution $\tilde{\boldsymbol{r}}|\boldsymbol{y}, \boldsymbol{\Sigma} \sim N(\boldsymbol{\mu}_N, \boldsymbol{\Sigma} + \boldsymbol{\Lambda}_N)$ is joint normal with mean vector $\boldsymbol{\mu}_N$ and covariance matrix $\boldsymbol{\Sigma} + \boldsymbol{\Lambda}_N$.

Influence on accounting

The main finding above is that while the population covariance matrix of $\boldsymbol{r}$ is known (by assumption) to be $\boldsymbol{\Sigma}$, the covariance of the predictive distribution of $\tilde{\boldsymbol{r}}$ is $\boldsymbol{\Sigma} + \boldsymbol{\Lambda}_N$. Hence, the predictive covariance is greater than the conditional (on $\boldsymbol{\mu}$) covariance (assuming positive prior and sample covariances). This added variation occurs for the reason that when predicting the next observation $\tilde{\boldsymbol{r}}$ of $\boldsymbol{r}$ we don't know the mean $\boldsymbol{\mu}$ of $\boldsymbol{r}$. Rather, we know that $\boldsymbol{r}$ has a true covariance matrix $\boldsymbol{\Sigma}$, but we don't know what $\boldsymbol{\mu}$ is. Hence, the wider or more diffuse the posterior belief distribution $f(\boldsymbol{\mu}|\boldsymbol{y})$, the wider the predictive distribution of $\tilde{\boldsymbol{r}}$, and thus the greater its predictive variance or covariance.

This result and a series of parallel results in the estimation risk literature are taken to imply that admitting Bayesian recognition of parameter uncertainty makes perceived returns (co)variances higher, and hence adds to the perceived overall market return variance and the market cost of capital. However, as the sample size $N \to \infty$, the empirical parameter estimates converge on the "true" returns parameters; i.e. the Bayesian posterior distributions become points at those values, so the Bayesian method and the frequentist "plug in" method come to agree.

More generally, the Bayesian implication, picked up in accounting research, is that a higher sample size N brings lower predictive returns variances and covariances. That intuitive finding proved irresistible to accounting theory and empiricists. It implies in a rigorous mathematical way that information reduces uncertainty or risk — and that more information reduces risk more, thus raising the possibility that better accounting can monotonically drive down investors' perceived risk and thus the firm's cost of capital.

The problem, however, is that models like the one just rehearsed are too narrow in their assumptions to yield any general law of how more information affects Bayesian uncertainty.

The first problem is that the result shown depends on the returns population covariance matrix $\boldsymbol{\Sigma}$ being known (with just $\boldsymbol{\mu}$ unknown). Once both parameters are admitted as unknown, the model becomes far more complicated (as noted by Harvey and Zhou, 1990, p. 226) and it becomes theoretically possible to learn from the empirical data that the stock covariances are probably wider than was expected before seeing that data.

Coles *et al.* (1995, p. 348) raise this oversimplification:

> .... the prior literature generally relies on the artificial and unrealistic assumption that, while the vector of mean returns or payoffs must be estimated the associated covariance matrix is known.

The second problem is that the model is really only a stationary "urn" model. There is no guarantee under nonstationarity or regime shifts that certainty will ever be reached, no matter how much data is collected (Barry and Winkler, 1976; Barry, 1978; Lewellen and Shanken, 2002). Indeed, it is possible that the most recent (i.e. most relevant) empirical observations might indicate that a jump or shift has occurred in the mean or the natural volatility of the returns or payoff process, in which case a Bayesian observer may become far less certain about the future value of the firm than she was beforehand.

Assumptions of stationarity are often disparaged as unrealistic in the estimation risk literature (e.g. Barry and Winkler, 1976; Barry, 1978; Avramov, 2002; Avramov and Zhou, 2010). This reality is captured in the Bayesian model of investor learning in Du and Huddart (2017) who find that, within their model, Bayesian learning in markets does not have a predictable path:

> Because the accounting signals are imperfect indicators of the underlying state of the firm and the state can change from one period to the next, the stock price does not stabilize or settle down. (Du and Huddart, 2017, p. 1)

Further detracting from the early finance Bayesian literature as a general foundation for our understanding of Bayesian inference in accounting, it must be accepted that the Bayesian literature is itself narrow in that it rests on models designed primarily for their mathematical tractability rather than descriptive validity. That methodological defect was emphasized by Winkler (1973, p. 402) and Barry and Winkler (1976) who noted that standard closed-form models involve certain tractable forms of prior and likelihood distributions, rather than capturing practical concerns such as asymmetry, fat tails, regime shifts and other real-world distributional properties.

In essence, the optimistic view that emerged in accounting of information naturally resolving some amount of uncertainty traces to a mathematically tractable but not realistic Bayesian model. A general Bayesian formulation of the problem must allow uncertainty to rise and fall in the way, for example, that perceived volatility rises and falls in derivatives markets, or that growers become more or less sure of a good crop as the weather changes.

8.5 The general Bayesian law

Bayesian portfolio theory models apply to hypothetical stockmarket returns where data are treated as if being drawn daily "from an urn", and further data can be drawn as required. In the usual frequentist way, samples are assumed to be random and iid from the same population.

The inference problem in accounting is much less well structured than the urn model assumed in finance. The investor might observe just one new signal, like this year's earnings announcement, "drawn" from an ill-specified probability distribution $f(V|\cdot)$ over a future cash flow V, and then set out to revise $f(V|\cdot)$ on that one item of information and with whatever comes with it (like the notes in financial reports).

In such a general inference problem, where all probability distributions are subjective and of no simple parametric form, the only Bayesian law that necessarily holds is the law of total variance, and its equivalent for covariance. See Johnstone (2015) and Johnstone (2016) for introduction to these two "distribution-free" laws.

Taking the latter, it is immediately evident that the covariance of asset payoff V_j and asset payoff V_k conditional on signal S, $\text{cov}(V_j, V_k|S)$, can be higher than the unconditional or prior covariance, $\text{cov}(V_j, V_k)$.

$$\text{cov}\,(V_j, V_k) = E\big[\text{cov}\,(V_j, V_k|S)\,\big] + \text{cov}\big(E\,[V_j|S]\,, E\,[V_k|S]\,\big).$$

Since the covariance of the conditional expectations $\text{cov}(E[V_j|S],$ $E[V_k|S])$ can be negative while the prior covariance $\text{cov}(V_j, V_k)$ is positive, it follows from this identity that the expected conditional covariance over all the possible signal realizations $S \in \{s_1, s_2, \ldots\}$, $E\big[\text{cov}(V_j, V_k|S)\,\big]$, need not be lower than the prior covariance.

Hence, as a general rule, assuming nothing but probability theory:

(i) the CAPM measure of risk or uncertainty, $\text{cov}(V_j, V_k|\cdot)$, need not shift towards zero upon receipt of imperfect information S; and

(ii) prior to observing signal S, the investor's expectation may be that the covariance will increase upon revelation of the signal outcome. This occurs where the sample space (i.e. the set $\{s_1, s_2, \ldots\}$ of all feasible sample observations of signal S) is known before S is realized, and is such that $\text{cov}\big(E\,[V_j|S]\,, E\,[V_k|S]\,\big)$ is negative, as would occur if possible realizations of S have somewhat contrary implications for the means of V_j and V_k.

Intuition about uncertainty

Intuition and the law of conditional covariance are clearly in agreement. Imagine that data exhibits a much stronger covariance between two variables than the prior covariance assessment. Data or accounting information must obviously have potential to bring an increase in the observer's subjective assessment of the covariance.

New information can always reveal that two variables have a stronger common driver than was previously understood, and hence rational models of statistical inference must sometimes produce a higher subjective posterior covariance between two random variables.

The same of course goes for the variance, where data or fundamental analysis must have the facility under Bayesian probability to indicate that the quantity in question is more variable or volatile than was previously appreciated. This is no surprise to portfolio managers whose subjective estimates of firm beta or returns covariance often increase with new information, rather than approaching zero. Accounting information undoubtedly plays a role in these forward-looking reassessments. Accounting information should by "doing its job" reveal greater uncertainty when say the firm invests in an inherently less predictable assets or ventures.

Reductio ad absurdum

The reductio ad absurdum is that if further information takes us monotonically towards certainty, it would follow that as soon as our probability of some true/false proposition gets above 0.5, it would automatically be put at one, because we would know with certainty that more information is going to take us monotonically in that direction.

8.6 Rogers *et al.* (2009)

A simpler form of the Kalymon (1971) model is widely leant upon in accounting literature. It involves inference about an unknown normal mean in a population with known variance (e.g. among a great many papers relying on this model, see Dye and Sridhar, 2007, pp. 737–738). It is sometimes called the normal–normal model because of its use of a conjugate normal prior and normal likelihood function. Its specific assumptions, particularly a known population variance, have the implications that the sampling variance is known (for any given n) and the posterior distribution of the unknown parameter must always have a lower variance than the prior distribution. The user is therefore "obliged" to be more certain after seeing any data. Rogers *et al.* (2009) noted this result as really a constraint on beliefs:

> . . . assuming the underlying distribution is held constant, by providing investors "with more balls from the urn" earnings guidance increases the rate at which investors learn about underlying profitability, lowering uncertainty. Other theory models generate similar predictions. . . . Consequently, any disclosure with precision greater than zero (with any information content) reduces uncertainty about firm value. This class of model thus leads to the conclusion that increased

> disclosure cannot increase investor uncertainty and will generally reduce uncertainty. (Rogers *et al.*, 2009, p. 92)

Thinking of a wider class of models, Rogers *et al.* (2009) pre-empted the recent paper by Dye and Hughes (2018) by describing how information can exacerbate market uncertainty by indicating a possible distribution or regime change, i.e. a change from "one urn to a different urn" with different parameter values, or even different parameters:

> An alternative possibility is that the forecast announcement, to the extent it is a surprise to investors (either because the forecast itself is a surprise or because the news it conveys is a surprise) creates uncertainty . . . larger surprises may create greater uncertainty in the sense that investors will be unsure about whether the surprise is an extreme outcome from an unchanged underlying earnings distribution or whether it signals a shift in the underlying distribution itself (a "regime" shift). (Rogers *et al.*, 2009, p. 92)

Uncertainty does not easily resolve. An accounting earnings report fully resolves uncertainty about that period's accounting earnings, but not about the firm's long-term fundamentals, which is the key unknown and ultimately why the current earnings report is of interest:

> Once earnings are announced, uncertainty about current period profitability is resolved. Consequently, the only uncertainty that remains is that created by disclosure about underlying (long-run) firm profitability. (Rogers *et al.*, 2009, p. 94)

8.7 Dye and Hughes (2018)

Recent work by Dye and Hughes (2018) extends formal Bayesian strategic accounting information theory to allow for a type of news or signal that can make the market either more or less certain. The unknown variable of interest is a future cash flow C and the news that creates uncertainty is that the firm issues "no report". That null report leaves the observer uncertain about whether there is no news existing inside the firm or the news indeed exists and is bad enough that the firm does not opt to report it.

If in fact news exists within the firm, it exists as an observed estimate x of C. That estimate is inherently informative in the Dye and Hughes (2018) model because it is specified as correlated and bivariate normal with the random C.

The completely general law of uncertainty in probability theory is the law of total variance. Under this law, a random variable like C can be understood as an "average draw" from all of the possible distributions of C. In general, C can arise out of a different distribution whenever some underlying state S changes, and, under each of these state distributions, C has a different mean and different variance.

Without knowing which state obtains, the prior (unconditional) variance of C, is given by the sum of (i) the average conditional variance, $E[\text{var}(C|S)]$, and (ii) the variance of conditional means $\text{var}(E[V|S])$. That is, by the law of total variance

$$\text{var}\,(C) = E[\text{var}(C|S)] + \text{var}(E[C|S]), \tag{8.1}$$

which implies that investors "expect" ex ante (before knowing S) that their perceptions of the variance of C will be lower after (or, if) S becomes known. Statistically, we reason that $E[\text{var}(C|S)] < \text{var}(C)$ because by construction $\text{var}(E[C|S]) > 0$. Yet crucially, for some states $S = s$, $\text{var}(C|s) > \text{var}(C)$, implying that knowing the state of the firm can increase the market's perceived variance of C.

It follows from Equation (8.1) that any hint that changes the investor's belief about which S obtains can add to, or subtract from, certainty (O'Hagen, 1994, p. 86). That is, the statistical law on which the Dye and Hughes (2018) model is based. It opens up the possibility within the model that uncertainty might increase with even very good information, contrary to the "conventional" understanding in accounting research:

> This contrasts starkly with the conventional statistical result that were the manager to disclose her private information, then this *always* causes investors' perceptions of both the

> variance of the disclosing firm's CF [cash flow] and the covariance between the disclosing firm's CF to shrink toward zero relative to investors' prior beliefs. (Dye and Hughes, 2018)

In the Dye and Hughes (2018) model, there are just two possible states or conditional distributions. When there is no information inside the firm, C is a draw from $f(C)$, which is the prior distribution of C applicable when no news exists. When there is information *and* the firm fails to report, C is a draw from the conditional distribution $f(C|x < x_c)$ where $x = x_c$ is the cutoff below which the firm omits to report x even when it exists. An equilibrium solution for x_c allows observers to infer enough about x_c that non-reporting at $x < x_c$ is informative. The unconditional distribution of C given a null report is the mixture distribution of $f(C)$ and $f(C|x < x_c)$, and its variance is given by Equation (8.1).

By applying standard results for a bivariate or jointly normal distribution, Dye and Hughes (2018) derive probability distributions for C conditional on either knowing x or knowing only that $x < x_c$. These are conditional probability distributions, and so are automatically Bayesian posterior distributions.

A general model

The following model can be interpreted as a generalization of the inference model in Dye and Hughes (2018). Their model was built on the assumption of a bivariate normal distribution between the payoff of interest and the firm's signal. My model extends the usual Bayesian "normal–normal" model based on a normally distributed payoff and an unbiased normally distributed signal of known variance. Its essential features are:

(i) the average effect of information or news is to reduce uncertainty. "Over time" or with more news, certainty tends to increase, albeit not monotonically and not necessarily in any hurry.

(ii) "bad news", including evidence of the firm being more probably in a "bad" state, can be highly precise and yet might still add to investor uncertainty.

In the Dye and Hughes (2018) model, the worst possible news is "no news". The event of "no news" is worrying because it adds to the probability that x exists inside the firm but is too unfavorable to report. In the more general model below, all news is represented by a reported point observation x, actually a sample mean $\overline{x}$, of arbitrary (possibly high) precision. Bad news is any reported or inferred $\overline{x}$ that shifts probability mass to the left.[7]

The main addition in my version of the normal–normal model is that it allows for there being two possible regimes or states of the firm, rather than a single assumed "population". The investor uses the observed sample mean to infer which regime or population is in effect. Note that the possibility of two regimes might be "physical" or merely subjective. As usual in Bayesian modelling, physical reality is never revealed for certain, all that is known is the user's subjective beliefs.

By introducing the possibility of two regimes, the Bayesian model becomes more realistic, because it allows the observed sample mean to sometimes leave the user less certain about the next payoff than before making that observation.

The model proceeds as follows. Suppose that the firm has two possible states S, "Normal" ($S = \mathrm{N}$) and "Distress" ($S = \mathrm{D}$). Distress can be highly improbable (later we assume a probability of 5%) but remains a possibility and is therefore admitted into the market's prior beliefs.

Let the firm's uncertain future cash flow be X. Market prior beliefs hold that if the firm is in its Normal state, (i) $X \sim N(\theta, \sigma^2)$, (ii) σ^2 is known, (iii) $\theta \sim N(\mu_1, \sigma_1^2)$ with specified values μ_1 and σ_1. Note that (iii) is the prior distribution of θ, given that the firm is in Normal state. This is a standard Bayesian model of an unknown mean θ and known variance σ^2, and has the characteristic that information is not only expected to reduce uncertainty about θ, it always does.

My assumption (ii) of known σ^2 guarantees that the sampling variance σ^2/n is known (rather than an unknown quantity) and is necessary to obtain a closed-form solution for the posterior distribution.

[7] There are realistic contexts suited to both models. Often in business, "silence" calls for the Bayesian interval interpretation explored by Dye and Hughes (2018).

The Dye and Hughes model makes this same assumption of a known sampling precision (and known zero bias). In my model, the known sampling precision changes with sample size n, whereas in the Dye and Hughes model it is fixed.

The remaining part of my model allows for Distress, in which case the firm's cash flow has the same prior distribution as above, $X \sim N(\theta, \sigma^2)$, except that $\theta \sim N(\mu_{D1}, \sigma^2_{D1})$, where μ_{D1} and σ^2_{D1} are specified, and naturally $\mu_{D1} < \mu_1$.

Taken as a whole, the market's prior beliefs are that cash flow X is drawn from a mixture distribution across the firm's two possible states, one state of which can be highly improbable ex ante (i.e. prior to any sampling or observation).

Before arriving at posterior beliefs about the next cash flow X, the market observes a signal, $\overline{x}$. That signal has the character of a statistical sample or sample mean $\overline{x} = \sum(x_1 + x_2 + \cdots x_n)/n$,[8] with arbitrary sample size n, and hence arbitrary sampling precision. By admitting $n > 1$, the model's conclusions can be adjusted to allow for information of different qualities or precisions.

By a standard result for normal distributions, the likelihood function of $\overline{x}$ given θ is

$$p(\overline{x}|\theta) = \sqrt{\frac{n}{2\pi\sigma^2}} \exp\left[-\frac{n}{2}\left(\frac{\overline{x}-\theta}{\sigma}\right)^2\right].$$

The posterior distribution of θ conditional on the firm being in Normal state is, by standard Bayesian results, $\theta \sim N(\mu_2, \sigma_2^2)$, where

$$\frac{1}{\sigma_2^2} = \frac{1}{\sigma_1^2} + \frac{n}{\sigma^2}$$

and

$$\mu_2 = \frac{\frac{1}{\sigma_1^2}\mu_1 + \frac{n}{\sigma^2}\overline{x}}{\frac{1}{\sigma_1^2} + \frac{n}{\sigma^2}}.$$

[8]For clarity observed x's are lower case and random unknown X's are upper case.

By the same equations, the posterior distribution of θ conditional on the firm being in Distress is $\theta \sim N(\mu_{D2}, \sigma^2_{D2})$, where

$$\frac{1}{\sigma^2_{D2}} = \frac{1}{\sigma^2_{D1}} + \frac{n}{\sigma^2}$$

and

$$\mu_{D2} = \frac{\frac{1}{\sigma^2_{D1}}\mu_{D1} + \frac{n}{\sigma^2}\overline{x}}{\frac{1}{\sigma^2_{D1}} + \frac{n}{\sigma^2}}.$$

If we know what state the firm is in, uncertainty is reduced by any new sample evidence, because both $\sigma^2_2 < \sigma^2_1$ and $\sigma^2_{D2} < \sigma^2_{D1}$. It remains uncertain, however, even after observing $\overline{x}$, whether the firm is in its Normal state or a state of Distress. That remaining uncertainty must be allowed for and can sometimes have the effect that the amount of the next cash flow X becomes more uncertain after the evidence than it was prior to the evidence.

The observation $\overline{x}$ gives some indication of which state the firm is in, because lower values of $\overline{x}$ are relatively more probable under Distress. More specifically, the likelihood ratio,

$$\frac{p(\overline{x}|N)}{p(\overline{x}|D)},$$

generally decreases with (sufficiently) lower $\overline{x}$.

To find the likelihood ratio, we need to obtain

$$p(\overline{x}|N) = \int_{-\infty}^{\infty} p(\theta|N)p(\overline{x}|\theta)d\theta,$$

and

$$p(\overline{x}|D) = \int_{-\infty}^{\infty} p(\theta|D)p(\overline{x}|\theta)d\theta.$$

Note that these expressions assume from the model that $\overline{x}$ is independent of S once given θ; that is,

$$p(\overline{x}|\theta, N) = p(\overline{x}|\theta, D) = p(\overline{x}|\theta).$$

We know already that under state N, $\theta|\overline{x} \sim N(\mu_2, \sigma^2_2)$, and under state D, $\theta|\overline{x} \sim N(\mu_{D2}, \sigma^2_{D2})$, so, by the usual formula for the probability density of a normal distribution[9]

$$p(\theta|\overline{x}, N) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{1}{2}\left(\frac{\theta-\mu_2}{\sigma_2}\right)^2\right] \tag{8.2}$$

and

$$p(\theta|\overline{x}, D) = \frac{1}{\sqrt{2\pi\sigma_{D2}^2}} \exp\left[-\frac{1}{2}\left(\frac{\theta-\mu_{D2}}{\sigma_{D2}}\right)^2\right]. \tag{8.3}$$

We now know that the predictive distribution of the next cash flow X conditional on the firm being in state N is

$$\begin{aligned} p(X|N) &= \int_{-\infty}^{\infty} p(\theta|\overline{x}, N) p(X|\theta) d\theta \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{1}{2}\left(\frac{\theta-\mu_2}{\sigma_2}\right)^2\right] \\ &\quad \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{X-\theta}{\sigma}\right)^2\right] d\theta \\ &= \frac{1}{\sqrt{2\pi(\sigma_2^2+\sigma^2)}} \exp\left[-\frac{1}{2}\frac{(X-\mu_2)^2}{(\sigma_2^2+\sigma^2)}\right], \end{aligned}$$

which is a normal distribution with mean μ_2 and variance $(\sigma_2^2 + \sigma^2)$.

In this calculation, we use the sample observation $\overline{x}$ to revise the probability distribution of θ, or more specifically the distribution of θ given N. Once we have that new distribution $p(\theta|\overline{x}, N)$ for θ, we have new probabilities for each possible θ under N, and we use them to find the revised predictive distribution of the next payoff X given N.

By identical calculations, the posterior predictive distribution of the next cash payoff X given D is

$$p(X|D) = \frac{1}{\sqrt{2\pi(\sigma_{D2}^2+\sigma^2)}} \exp\left[-\frac{1}{2}\frac{(X-\mu_{D2})^2}{(\sigma_{D2}^2+\sigma^2)}\right],$$

which is normal with mean μ_{D2} and variance $(\sigma_{D2}^2 + \sigma^2)$.

[9] Note that $\frac{1}{\sqrt{2\pi\sigma_2^2}}$ and $\frac{1}{\sqrt{2\pi\sigma_{D2}^2}}$ are normalizing constants designed to make the integrals of Equations (8.2) and (8.3) over the real line equal to one.

The unconditional predictive distribution of X is a probability-weighted mixture distribution of the two conditional predictive distributions. The mixture weights are the posterior probabilities of the two states, namely, $p(N|\overline{x})$ and $p(D|\overline{x})$.

Given $p(\overline{x}|N)$ and $p(\overline{x}|D)$ above, we find the probabilities of the two possible states, conditional on the sample observation $\overline{x}$. Letting the prior probability of Normal be $p(N) \equiv 1 - p(D)$, the posterior probability of Normal is

$$p(N|\overline{x}) = \frac{p(N)p(\overline{x}|N)}{p(N)p(\overline{x}|N) + p(D)p(\overline{x}|D)}.$$

We now have all the ingredients to find the unconditional posterior predictive variance of the next cash flow X.

The parameters of this mixture distribution are, by the law of complete probability,

$$\mu_{m|\overline{x}}(X) = p(N|\overline{x})\mu_2 + p(D|\overline{x})\mu_{D2},$$

and, by the law of total variance,

$$\sigma^2_{m|\overline{x}}(X) = E[\text{var}(X|S)|\overline{x}] + \text{var}(E[X|S]|\overline{x}), \tag{8.4}$$

where $E[\text{var}(X|S)|\overline{x}]$ is the posterior average of the two predictive variances across the two states, that is,

$$\begin{aligned} E[\text{var}(X|S)|\overline{x}] &= p(N|\overline{x})\left(\sigma_2^2 + \sigma^2\right) + p(D|\overline{x})\left(\sigma_{D2}^2 + \sigma^2\right) \\ &= \sigma^2 + \left(p(N|\overline{x})\sigma_2^2 + p(D|\overline{x})\sigma_{D2}^2\right), \end{aligned}$$

and $\text{var}(E[\theta|S]|\overline{x})$ is the variance of the two posterior means, that is,

$$\text{var}(E[X|S]|\overline{x}) = p(N|\overline{x})\left(\mu_2 - \mu_m\right)^2 + p(D|\overline{x})\left(\mu_{D2} - \mu_m\right)^2.$$

Note that the posterior predictive variance of θ, denoted by $\sigma^2_{m|\overline{x}}(\theta)$, is

$$\begin{aligned} \sigma^2_{m|\overline{x}}(\theta) &= E[\text{var}(\theta|S)|\overline{x}] + \text{var}(E[\theta|S]|\overline{x}) \\ &= \left(p(N|\overline{x})\sigma_2^2 + p(D|\overline{x})\sigma_{D2}^2\right) \\ &\quad + \left(p(N|\overline{x})\left(\mu_2 - \mu_m\right)^2 + p(D|\overline{x})\left(\mu_{D2} - \mu_m\right)^2\right), \end{aligned}$$

which allows us to rewrite Equation (8.4) neatly as

$$\sigma^2_{m|\overline{x}}(X) = \sigma^2 + \sigma^2_{m|\overline{x}}(\theta). \tag{8.5}$$

Observe from Equation (8.5) that there are two sources of uncertainty affecting our final uncertainty about cash flow X. First, even if we knew θ for certain, X still has known variance σ^2, because the model holds that $X \sim N(\theta, \sigma^2)$. Second, despite observing $\overline{x}$ we don't know θ. In fact, despite potentially narrowing down what state the firm is in, we don't know for certain the state or distribution from which θ is drawn. Rather, θ is drawn from a mixture distribution and its posterior predictive variance is $\sigma^2_m(\theta)$. Conveniently, the sum of these two uncertainties is captured correctly by the sum of the two variances, shown in Equation (8.5).

For explanation of these two sources of variance see Winkler (2003, p. 181), who notes that the predictive variance "takes into account both uncertainty about θ and uncertainty about X given θ".

Note that although the mixture distributions of θ and X are mixtures of normals and so are not normal, their predictive variances are found nonetheless by the law of total variance, which is a general law and distribution-free, see Gelman *et al.* (2004). Importantly, also, note that n does not appear explicitly in Equation (8.5), for two reasons. First, this is the predictive variance of just a single observation X, and, second, n has been incorporated into the assessed variance of the mixture distribution of θ, $\sigma^2_{m|x}(\theta)$.

To understand the possible effects of information on investor certainty, we can now compare the posterior predictive variance (8.5) with the prior predictive variance

$$\sigma^2_m(X) = \sigma^2 + \sigma^2_m(\theta). \tag{8.6}$$

which is found the same way as Equation (8.5) but without the influence of the new observation $\overline{x}$. Specifically, $\sigma^2_m(\theta)$ is the prior predictive variance of θ, which is the variance of the mixture distribution of θ across the two states. That is,

$$\sigma^2_m(\theta) = E[\text{var}(\theta|S)] + \text{var}([E[\theta|S]),$$

where

$$E[\text{var}(\theta|S)] = p(N)\sigma^2_1 + p(D)\sigma^2_{D1}$$

and

$$\text{var}(E[\theta|S]) = p(N)(\mu_1 - \mu_m)^2 + p(D)(\mu_{D1} - \mu_m)^2.$$

and

$$\mu_m = p(N)\mu_1 + p(D)\mu_{D1}.$$

It is obvious that Equations (8.5) and (8.6) have a common ingredient in σ^2. They differ therefore only to the extent that $\sigma^2_{m|\overline{x}}(\theta)$ differs from $\sigma^2_m(\theta)$.

Interestingly, it can occur that $\sigma^2_{m|\overline{x}}(\theta) > \sigma^2_m(\theta)$ implying that the sample observation $\overline{x}$ can add to investor uncertainty about θ and the next cash flow X. That can occur if $\overline{x}$ adds sufficiently to uncertainty about which state N or D obtains.

Intuitively, suppose that the distributions of X are far apart under the two different possible states. Information that leaves the user's beliefs mid-stream between the two states means that she sees a very wide range of feasible X outcomes, even if the conditional (upon $\overline{x}$) variances within each of the states are in her assessment low.

A clearer depiction of how information $\overline{x}$ affects investor uncertainty about X is obtained by looking at numerical examples. Figure 8.1 plots the posterior predictive variance for the next X against the observed

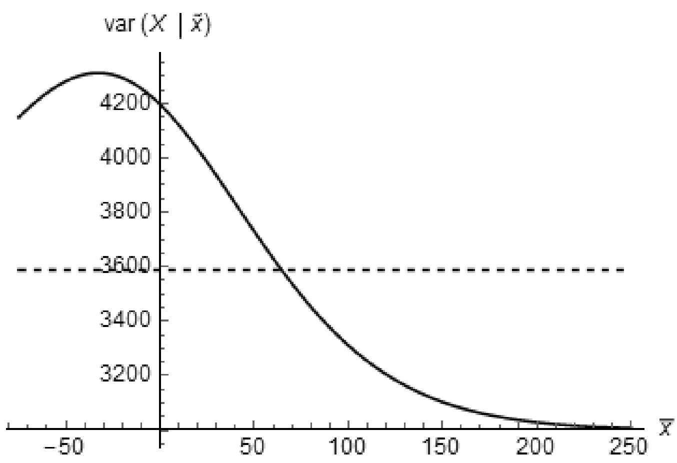


Figure 8.1: Posterior predictive variance $\text{var}(X|\overline{x}) \equiv \sigma^2_{m|\overline{x}}(X)$ of cash flow X, given observed signal $\overline{x}$ with $p(N) = 0.95$, $\sigma = 50$, $\mu_1 = 150$, $\sigma_1 = 25$, $\mu_{D1} = 50$, $\sigma_{D1} = 20$, $n = 1$.

sample mean $\overline{x}$ under some assumed input values. The horizontal line shows the prior predictive variance of X, calculated from Equation (8.6). For the sake of illustration, we start with prior probability $p(\mathrm{D}) = 0.05$, which is roughly in line with the empirical probability of US bankruptcies.[10]

Note that in Figure 8.1, uncertainty reacts to the observed $\overline{x}$ in an S-shape and there is a range of $\overline{x}$ for which uncertainty about X actually increases, that is,

$$\mathrm{var}(X|\overline{x}) \equiv \sigma^2_{m|\overline{x}}(X) > \mathrm{var}(X) \equiv \sigma^2_m(X).$$

That happens when the observed mean $\overline{x}$ is "lowish", dragging the probability of Distress higher from its prior of 0.05 and hence adding sufficiently to uncertainty about which state the firm is in.

Figure 8.2 presents the same example as Figure 8.1 but the sample size is increased from $n = 1$ to $n = 5$. That increase in sampling precision implies better information and has the effect of reducing each of the

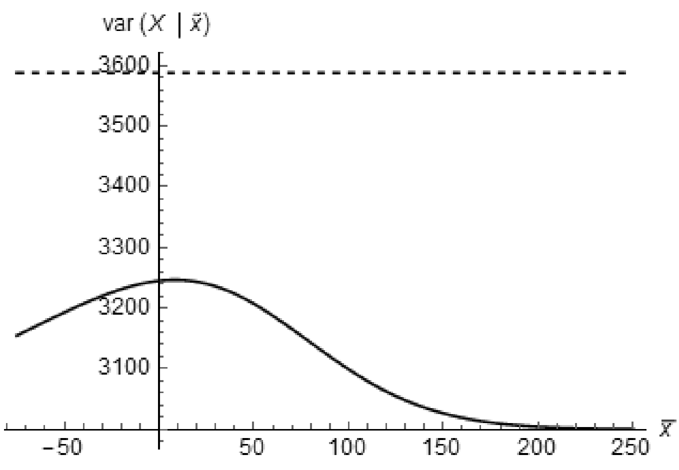


Figure 8.2: Posterior predictive variance $\mathrm{var}(X|\overline{x}) \equiv \sigma^2_{m|\overline{x}}(X)$ of cash flow X, given observed signal $\overline{x}$ with $p(N) = 0.95$, $\sigma = 50$, $\mu_1 = 150$, $\sigma_1 = 25$, $\mu_{D1} = 50$, $\sigma_{D1} = 20$, $n = 5$.

[10] A firm in Distress need not end up bankrupt, and conversely a firm in Normal state can end up bankrupt. Bankruptcy would occur with "high" probability if the next cash flow X is sufficiently low (say near zero or negative), because that observation would imply that the firm is probably in Distress state.

conditional variances of X sufficiently that the posterior predictive variance $\mathrm{var}(X|\overline{x})$ is always lower than the prior predictive variance $\mathrm{var}(X)$ for any observed value $\overline{x}$.

8.8 Why a Predictive Distribution?

Predictive distributions were devised in Bayesian statistics by Zellner and Chetty (1965) and were brought into the information uncertainty literature in finance by Barry (1978) and Brown (1979).

There has been little application of Bayesian predictive distributions in accounting theory, however Dye (1990, p. 6) correctly applied the main idea by distinguishing between two separate risk premia, one for the known variance of a payoff around a given "true" parameter value, and the other for uncertainty about the actual value of that unknown parameter. The latter can be viewed as "parameter risk".[11]

Predictive distributions apply usefully when the task is to predict — in the form of a probability distribution — the next outcome of a statistical process, such as the very next cash flow C or say the monthly stock return or the firm's sales units or costs.

There is a difference between inference about a parameter, like a population or process mean φ, and inference about the next observation that comes out of that stochastic process. The intrinsic value of a firm might be better considered as a parameter φ, because φ captures the firm's fundamentals, like a population parameter. The next cash flow is a single random outcome of a process characterized by φ, and may vary widely around φ. That extra variation is captured in the predictive distribution. The next outcome is conceptually a "draw" from the predictive distribution.

As an example, suppose that the firm is a Bernoulli distribution producing random $Y \in \{0, 1\}$ with parameter φ, where φ is unknown $(0 < \varphi < 1)$. Our prior distribution for φ is discrete with $\Pr(\varphi = 0.4) = \Pr(\varphi = 0.8) = 0.5$, so $E(\varphi) = 0.6$ and $\mathrm{var}(\varphi) = 0.04$. We can use this prior knowledge to work out the mean and the variance of the next observation Y. Its mean is obviously $E[Y] = E[E[Y|\varphi]] = 0.5(0.4) +$

[11] There is similarly a Bayesian posterior predictive distribution, calculated on results from DeGroot, in Dye and Sridhar (2007, p. 738).

$0.5(0.8) = 0.6$. And given that $E[\text{var}(Y|\varphi)] = 0.5[(.4)(.6) + (0.8)(0.2)]$, its variance is

$$\begin{aligned}\text{var}\,(Y) &= E[\text{var}(Y|\varphi)] + \text{var}(E[Y|\varphi]) \\ &= 0.2 + 0.04 \\ &> \text{var}(\varphi).\end{aligned}$$

The main point of this discussion is the need to distinguish between uncertainty about a population parameter like φ and the uncertainty that exists about random outcome Y even if φ were known (which usually it is not).

That Bayesian distinction is fundamental to the concept of cash flow uncertainty. The model outlined above, adding to the Dye and Hughes discussion, has the advantage that it treats the next cash flow as driven by a "random" (i.e. uncertain) parameter. The predictive distribution of Y accounts for both the uncertainty surrounding parameter φ and the uncertainty about Y given φ, therefore absorbing all of the underlying uncertainties that determine how precisely we can forecast Y.

A methodological issue for analytical accounting research is that predictive distributions, being mixtures, are not normal even when the distributions being mixed are normal. That creates a problem for the usual mean–variance equation for the equilibrium price of an uncertain future cash flow, as used in Lambert *et al.* (2007) and many other accounting theory papers. That equation presumes exponential utility and a normally distributed payoff. An alternative that will not change the equation substantially is to use the comparable mean–variance asset price equation derived under quadratic utility, which does not need a normal payoff. See Johnstone (2016) for that derivation.[12]

8.9 Limits to certainty

Examples of information and analysis that add to doubts rather than resolving them are as inevitable in financial reporting as in other fields.

[12] An important methodological point is that the asset price of a mixture distribution is not found by a mixture (weighted average) of the conditional asset prices.

Observed monthly sales volumes or costs or cash receipts may sometimes cast great doubt over the sustainability of a firm or product. These figures are not inaccurate or imprecise. Provided that measurement is credible, even the most common types of accounting data can shake investor confidence, while nonetheless leaving investors more accurately informed about fundamentals and less likely to make unprofitable investment decisions.

Activity-based-costing (ABC) advocates have long made this kind of argument. They hold that a more refined costing system might expose previously unseen doubts over a firm's or product's economic sustainability. Any similar enhancement that leads to more timely loss recognition is likely to dent investor confidence. Firms may wish not to divulge information of this ilk, which is why nondisclosure is often viewed as a negative signal (Verrecchia, 1983; Verrecchia, 2001).

Accounting information might reveal the full exposure of the firm to derivatives or foreign exchange rates, which could suggest jointly that (i) the probability distribution of future profits has a peak in the far left tail that was not recognized and (ii) the firm's beta or dependence on market-wide factors is higher than previously understood. At any time in its reporting sequence, the subject firm can take on inherently more risky activities. Even its existing assets or income streams might then be at higher risk. Relevant accounting information should not hide or ignore that risk, or purport to resolve it. Put another way, the ability of accounting information, or accountants, to predict cash flows is naturally limited, thus limiting the facility of accounting to lessen the firm's cost of capital.

These examples are sufficient to suggest that even very "good" accounting should not by its nature always assist the firm by lowering its cost of capital. To the contrary, better disclosure or measurement systems can produce signals that lead an analyst to simultaneously downgrade expected earnings and raise the required return on equity, thus bringing about a big price fall.

This point is well made by Leuz and Wysocki (2008) in relation to accounting restatements, which are Bayesianly just another transmission of accounting information. Similarly, Rogers *et al.* (2009) explain how some earnings forecast releases, particularly those bringing unexpected negative news, can "create uncertainty" rather than resolve it. Accounting "restatements" or corrections to previous reports are seen as concerning and hence contributors to market uncertainty. Hribar and Jenkins (2004) find that:

> ... overall information uncertainty is increased following a restatement, causing investors to require a higher rate of return (Hribar and Jenkins, 2004, p. 339)

If a restatement of an accounting result can add to uncertainty, so presumably should any form of adverse accounting news, or adverse news about the company delivered by the managers through the accounting statements.[13]

Realistically, the accountant does not have the capacity to produce signals that by necessity make uncertain business outcomes (e.g. cash flows including costs and revenues) more predictable. That would be akin to a stock analyst who always makes users more certain of whether the firm will succeed or not.

Sometimes the most technically advanced indicator leaves an expert analyst less certain of the underlying state of nature or of future outcomes. This is not to say of course that such information is not worth paying for ex ante. Rather, it merely acknowledges that a test or signal with high expected information value, and hence clearly worth its cost ex ante, will not always, when realized, bring about greater certainty ex post.

In some scientific fields, it is possible to set up a test or signal with known theoretical error characteristics (or a sharply peaked likelihood function). In others, such as medicine, test error properties are well known from the accumulation of empirical evidence. In principle, accounting theory and practice faces the same signal design task as other fields, however there is generally less theory and laboratory technique by which to work. This inherent difficulty is reflected in decades of academic and professional dispute about accounting standards and

[13] In principle good news will also add to uncertainty, if it contradicts a strong belief that the firm is for example insolvent or failing. In general, any news that contradicts strong prior beliefs adds to uncertainty.

fundamental questions such as how to value assets in the balance sheet (historical cost, current cost, "fair value" etc.). Similarly, there are many unresolved issues in the empirical literature regarding how to explicate and measure "earnings quality" (Dechow and Schrand, 2004; Dechow *et al.*, 2010; DeFond, 2010).

In medicine, tests exist that tell the technician whether cells are infected or not, generally with very low error rates, but there is not the same technology available to accountants when the unknown quantity is whether a firm is a "going concern" or not, or whether a product or a new costing system will succeed. For one thing, the mechanisms modelled in medicine are, it would seem at least, much more stationary and "urn like" than changeable accounting and financial market contexts.

Many economic events "are" or appear on present knowledge to be "as random" as spins of a wheel. In those contexts, information of any practically achievable quality should usually leave the investor feeling unable to beat the odds. Obvious examples might include weather or political events on which the firm's cash flows hinge. Many economic outcomes have a stochastic or "pure chance" element. They can depend on technological breakthroughs, demand shifts, regulation changes, competitors actions and many micro-variables such as the variable skills, efforts and insights of managers and employees.[14] These are contexts where the main danger is over-certainty and overstatement of the analyst's or forecaster's capacities. The practical difficulties of accurate economic forecasting and probability assessment are well known and long documented. They trace to innate uncertainty about future cash flows and are just as troublesome for accounting valuation.

To allow for intrinsically highly random or uncertain events, Bayesian inference models like Dye and Hughes (2018), that do not constrain the user to become more certain after observing more data, but instead permit uncertainty to sometimes increase, and perhaps never decrease much even with more information, are theoretically essential to how we

[14] Accidental outcomes like the Iridium satellite collision are raised in finance textbooks as things that are pure chance and uncorrelated with known variables.

frame the normative ideals and practical limits of accounting information. A philosophical stance (see later) is that investors are often best served by clearer revelation of a firm's riskiness.

8.10 Lewellen and Shanken (2002)

The Bayesian logic of inference-then-decision is exemplified in a paper on learning in markets by Lewellen and Shanken (2002). They take an expressly anti-frequentist approach, for the reason that rational investors will lose money if they base their investments on assumed parameter values rather than probability distributions over parameter values, because their beliefs about parameter values will inevitably be wrong and their investments will be unhedged against parameter risk.[15] The following is a short summary of the Lewellen and Shanken (2002) depiction of Bayesian inference in a stock market.

Investors are modelled as having two characteristics. First, they have the rationality to be Bayesian. Secondly, they have a finance logic in mind, particularly an asset-pricing model. That is essential to Bayesian decision-making, because investment decisions hinge on probability assessments over future asset prices, requiring the investor to anticipate how the market mechanism will react to new information. Thus, successful investment by a Bayesian requires a wider class of theoretical acumen than merely Bayes theorem (this monograph is mostly confined to the subtleties of Bayesian thinking, of itself).

Following Markowitz and others, investors are assumed to act on their subjective probabilities, not on the exogenous unknown "true" probabilities. Parameters are modelled as uncertain "random variables", because they are never known with certainty. Even after much data, Bayesian investors never fully learn a "true" distribution, especially in a non-stationary environment.

By merely facing up to parameter uncertainty, investors reveal for themselves greater uncertainty. In a stationary environment, investors can in some models learn about parameter values over time, but the

[15] The NBER *Working Paper 7699* that preceded this publication was decidedly more critical of the frequentist approach and the Bayesian mistakes that it brought to the finance literature.

real economy and firms evolve over time, with unobservable shocks to parameters, meaning that parameter uncertainty will not disappear.

The investor plays a game of "blind man's buff", because in the empirical dividend stream she sees only a noisy representation of a changing underlying payoff or returns process. In the investor's vision, the underlying payoff parameters seem to fluctuate randomly over time, even when they do not. She sees little bits of empirical evidence indicative of the underlying parameter values, but even as data accumulates she may never veer close to the true distribution of returns in her assessment of a continually revised but still lagging subjective distribution.

The actual returns process might shift one way in its mean or its variance, while, at the same moment, based on last period's data, the investor might update her beliefs in exactly the opposite direction. Investors can have a less true picture of the current underlying payoff or returns process after even a long observation period.

8.11 Neururer *et al.* (2016)

Neururer *et al.* (2016) summed up three types of Bayesian thinking in accounting about the effect of information on uncertainty:

> This paper investigates alternative Bayesian models of learning to explain changes in uncertainty surrounding earnings announcements. Specifically, we examine three alternative scenarios reflecting how the magnitude of a performance signal affects investors' posterior variances. The first scenario we investigate is that additional signals of firm performance reduce investor expectations of posterior variances - that is, lead to declines in uncertainty (Lewellen and Shanken, 2002, e.g.). Under these models, the magnitude of the signal plays no role in the extent to which uncertainty is resolved. This is predicated on the notion that a firm has a fixed distribution of outcomes and that the release of the signal helps to reveal this distribution, thereby reducing uncertainty about future

> firm value. We label such models as "*Constant Uncertainty Resolution.*"
>
> The second scenario is that the magnitude of the signal plays a role in the uncertainty resolution surrounding a signal's release (Rogers *et al.*, 2009, e.g.). We label this as "*Uncertainty Resolution Conditioned on Signal Size.*" Under these alternative Bayesian models, while the release of a signal helps to provide information about the distribution of outcomes (and thus leads to reduced uncertainty), the extent to which the signal deviates from expectations affects the amount of uncertainty resolution. However, note that the models under this scenario allow for attenuation — but not net increases — in uncertainty due to signal size.
>
> The third scenario we investigate is that sufficiently large signals cause a net increase in uncertainty, or a "regime shift" (Pastor and Veronesi, 2009). We label this as "*Bayesian Learning with Increased Posterior Uncertainty.*" Critically, such models of Bayesian learning allow for signals, which deviate sufficiently from expected values, to lead to increased uncertainty. In other words, these models allow that some signals are large enough to increase investors' posterior variance. (Neururer *et al.*, 2016, p. 401)

Ultimately the authors find evidence supporting the last and most obviously realistic of the three underlying models. If we are interested in the probability that the firm's costs will be high next month or year, we will of course be influenced not only by the sample size of previous monthly costs, but also by the dollar amounts of those costs. And if last month, costs were much higher than previously, then previous confidence in continued low costs will be generally reduced. Any Bayesian model of the effects of accounting earnings, or similar disclosures, that ignores these realities is bound to mislead in principle and can only exist for its simplicity or mathematical convenience rather than methodological validity. Again, we are led back to the caveat in Beaver (1968) that information must be "allowed" to heighten rational users' uncertainty.

8.12 Veronesi (1999)

Veronesi (1999) developed a dynamic stochastic model of belief revision and market re-pricing in a market where the underlying regime changes between two states and decision makers infer the probabilities of the states from past dividends. This model is revealing but should not be considered to be "Bayesian" because the homogeneous investors' probability that the market is in a given state evolves by an assumed diffusion process or "law of motion" rather than by Bayesian updating. There is no connection explained via Bayes theorem between dividends and users' inference. Specifically, there is no likelihood function. Instead, users are assumed to "estimate" the probability of the state, rather than "reason" it by Bayes theorem.

Veronesi's model can be considered as a stylized time series of how a Bayesian probability might evolve dynamically with information, but it does not delve into how investors would reason when forming their beliefs. For example, by what likelihood function would they form beliefs about next period's dividend from last period's dividend, or from say a run of successive dividend increases? Ottaviani and Sorensen (2015) show that market aggregates, unlike possibly their individual traders, are not naturally coherent Bayesian updators, in the sense that equilibrium prices do not evolve as if "the market" has a prior belief shown in the last price and a corresponding Bayesian posterior belief shown in the next price.

9

How Information Combines

Under Bayes theorem, information y interacts with information x in ways that are not always predictable. It can occur for example that $f(A|x) > f(A)$ and $f(A|y) > f(A)$ but $f(A|x, y) < f(A)$, implying that the "meaning" of x is quite different "of itself" than in combination with y, and vice versa.

Similarly, a signal can be weak of itself but very strong, in the same direction, when combined with another signal. It does not matter whether x or y is received first, the sequence of updating beliefs will arrive at the same end result $f(A|x, y)$ either way. The only difference is that the intermediate probability, either $f(A|x)$ or $f(A|y)$, will be different, which in a way shows how the belief path contains random ups and downs, leaving the probability held at any moment a matter of "luck", since the arrival of observations is usually "random".

Synergies and interactions between information sources give a Bayesian basis to the proverb, "a little knowledge is a dangerous thing". The apparent meaning of a given signal can be reversed when another condition or signal is added to the information pot. See Penman (2010) on how accounting can add utility by being more independent of other financial information. Similarly, Francis *et al.* (2006) note that users must account for "statistical associations" between earnings quality and other items of market information:

> ... interactions among components of capital market information undoubtedly exist and are potentially important for understanding financial reporting quality. (Francis *et al.*, 2006)

The interaction between information sources is a major issue for accounting standard setting. An analogy exists with portfolio theory, in the sense that a signal is more valuable cet.par. when it is uncorrelated with the other signals. Put another way, like stocks in a portfolio its apparent individual value depends on what it is combined with. In the literature on combining probability forecasts from different sources, there are formal arguments for why it can assist to add forecasts that are less accurate of themselves if they contribute by adding independent and possibly often contrary opinion. See for example Grushka-Cockayne *et al.* (2017) and Winkler *et al.* (2019). Underlying this argument is a notion like Penman's that it helps to introduce different prior opinion, expertise, training, methods, experience, biases or vested interests, and other sources of individuality, into the mix.

9.1 Combining two risky signals

The following is a re-interpretation of the calculations in Pearl (1999), which were used to illustrate Simpson's paradox. My re-interpretation is in terms of how the meaning of a given signal sometimes switches to the opposite when its likelihood function is conditioned on another more or less correlated signal. On Simpson's paradox in its general form, see Lindley (2014) and Kadane (2011).

Suppose as above that $V_j \in \{0, 1\}$ is binary with prior probability $p(1) = 0.45$. Consider two risky binary signals from different sources. The first signal $\tau \in \{+, -\}$ has error probabilities, $p(+|0) = 15/44$ and $p(-|1) = 11/36$, in which case it follows that $p(+) = 0.5$, $p(1|+) = 0.625$ and $p(1|-) = 0.275$. The second signal $\kappa \in \{F, U\}$ has error probabilities $p(F|0) = 5/11$ and $p(U|1) = 4/9$, in which case $p(F) = 0.5$, $p(1|F) = 0.5$ and $p(1|U) = 0.4$.

On the face of these assessments, a positive signal from the first source and a favorable signal from the second source would seem to each add evidence supportive of $V = 1$, and thus apparently corroborate one another when they occur together.

That intuition does not allow, however, for the risk of correlated errors. To see how Bayesian allowance for this risk affects beliefs, imagine that the perceived error probabilities of signal κ when conditioned on signal τ are as shown in the two contingency tables in Table 9.1.

Given the joint probabilities in Table 9.1, the conditional error probabilities of signal κ are $p(F|1,+) = 18/25$, $p(F|1,-) = 2/11$, $p(U|0,+) = 12/15$, and $p(U|0,-) = 8/29$. The resulting posterior probabilities under different levels of conditioning are then as shown in Table 9.2 (see the Appendix for calculations).

The paradoxical result apparent in Table 9.2 is that signal κ has one evidential interpretation when received by itself and the opposite interpretation when taken in conjunction with signal τ. Specifically, $\kappa = F$ by itself adds evidence favoring $V = 1$, yet contributes evidence against $V = 1$ when it occurs jointly with τ, no matter whether $\tau = +$ or

Table 9.1: Joint probabilities $p(V, \kappa|\tau)$.

When $\tau = +$		
	$V = 1$	$V = 0$
$\kappa = F$	9/20	3/10
$\kappa = U$	7/40	3/40

When $\tau = -$		
	$V = 1$	$V = 0$
$\kappa = F$	1/20	1/5
$\kappa = U$	9/40	21/40

Table 9.2: Conditional probabilities of $V = 1$.

$p(1)$	0.45
$p(1\|+)$	0.625
$p(1\|F)$	0.5
$p(1\|+, F)$	0.6
$p(1\|+, U)$	0.7

$p(1)$	0.45
$p(1\|-)$	0.275
$p(1\|U)$	0.4
$p(1\|-, F)$	0.2
$p(1\|-, U)$	0.3

$\tau = -$. That is, $p(1|+, F) < p(1|+)$ and $p(1|-, F) < p(1|-)$. Similarly, $\kappa = U$ by itself presents evidence against $V = 1$, but adds evidence favoring $V = 1$ when joint with τ, no matter whether $\tau = +$ or $\tau = -$.

It follows without need for further calculations that the reversals of evidential direction, which can occur with each new level of conditioning, flow through to the costs of capital or risk premia of individual securities.

A good way to understand this connection is to go back to the usual context in which Simpson's paradox is explained in statistics. The numerical example above was translated into an "information risk" context from its original medical context. In the medical analogy, V is a binary for survival or not ($V = 1$ indicates survival), $\tau \in \{+, -\}$ indicates whether the subject is male or female ("+" is male, "−" is female) and $\kappa \in \{F, U\}$ signifies whether the subject was given the drug or not ("F" means drug, "U" means no drug). In this context, the respective probabilities show that the drug had a positive effect on the average survival rate over the population as a whole, but had a negative effect for males as a subset and also a negative effect for females as a subset. So if an insurance company were to see these results, the life insurance premium would go in opposite directions depending on whether the subject was viewed as just a person who was given the drug or whether a two-way partitioning was used, in which case the drug would be seen as having a negative effect on survival probability.

This is indicative of how a signal can appear to say or mean one thing of itself, yet when the same signal is observed jointly with a correlated confounding factor, such as an experimental condition or another source of information, its meaning can reverse. Or, similarly, it can be found to be uninformative or to be far stronger than first thought.[1] In this sense, we can never be sure that a signal means what it apparently "says". The Bayesian approach to this problem, which is a defence but not a panacea, is to impound in the likelihood function as many of the conditions associated with any given signal, affecting its Bayesian meaning or implications, as are known or perceived.

[1]Related theory, showing how information can combine counter-intuitively, exists in the Bayesian literature on combining individual probability forecasts of the same events (e.g. Winkler, 1989).

Appendix

Here are the calculations for $p(F|1,+)$.

$$p(F|1,+) = \frac{p(F|1)p(+|1,F)}{p(+|1)} = \frac{(5/9)(0.9)}{25/36} = 0.72,$$

where

$$p(+|1,F) = \frac{p(+)p(1,F|+)}{p(1,F)} = \frac{p(+)p(1,F|+)}{p(+)p(1,F|+) + p(-)p(1,F|-)} = \frac{0.5(0.45)}{0.5(0.45) + 0.5(0.05)} = 0.9.$$

It follows then that

$$p(1|+,F) = \frac{p(1|+)p(F|1,+)}{p(F|+)} = \frac{0.625(0.72)}{0.625(0.72) + 0.375(0.8)} = 0.6.$$

The corresponding conditional likelihoods $p(F|1,-)$, $p(F|0,+)$ and $p(F|0,-)$ are found the same way.

10

Ex Ante Effect of Greater Risk/Uncertainty

It is market folklore that "investors abhor uncertainty", but that is too simple. What if uncertainties were all resolved? All investments would be risk-free and no investor could earn more than the risk-free interest rate. If an entirely risk-free investment were desirable, investors could do that now, and there might be no stockmarket. So the stockmarket, and its risky stocks, must be adding to investors' expected utility, which suggests in turn that accounting is well served to consider how it can contribute not to greater certainty per se, or to always lowering the cost of capital, but to the greater task of maximizing investors' expected utility, both ex ante and ex post.

It is essential to differentiate between ex ante utility and ex post expected utility. Ex ante expectations rest on ex ante probability assessments, and may never be realized or even close to realized. The decision maker forms beliefs Bayesianly on all information that seems worth gathering (based on ex ante expectations). Actions are taken on those subjective beliefs according to the rule of maximizing ex ante expected utility. The decision maker knows that actions may or may not produce the utility "expected", they are merely the best actions, given current probability beliefs. The utility actually produced is partly due to the "accuracy" of her beliefs, partly due to how bold or risk tolerant she is in that choice of action, and partly due to "luck" since any random bet, "sure thing" or "long shot", might or might not payoff.

Ex post expected utility, meaning the decision maker's actual average realized utility, is considered in later parts of this monograph. It comes under the heading "economic Darwinism", which is the study of the innate statistical advantages that make financial decision makers "fit" to survive and prosper over the long term of repeated albeit changing investment decisions. See for example Blume and Easley (2006).

10.1 Risk adds to ex ante expected utility

Bertomeu (2013), Cheynel (2013) and Gao (2010) raise the possibility that an increase in market risk or the cost of capital can be favorable to investor welfare. The following analysis leads to a conclusion that more risk, when correctly "priced", adds to the prospective investor's ex ante expected utility.

If there are two stockmarkets offering the same mean payoff and differing only in their level of certainty, which market of itself offers higher ex ante expected utility to a potential new investor who buys the rational portfolio of the risky market and holds the remainder of her wealth in a risk-free asset?

A simple way to investigate the consequences of greater ex ante certainty is to work with the usual mean–variance CAPM equation, following Lambert *et al.* (2007), for the price of the stock market

$$P_M = \frac{1}{R_f}\,[E[V_M] - c\,\mathrm{var}(V_M)], \quad R_f = (1 + r_f) \tag{10.1}$$

where V_M is the aggregate market payoff, and c is the premium imposed by the market for the risk (variance) of V_M (and r_f is the risk-free rate).

The stockmarket can be considered as a single unit of a single risky asset that pays random amount V_M at period-end, and trades at unit price P_M at period-start. Its ex ante market price is P_M (which is the sum of the individual prices of the risky assets in the market). See Fama and Miller (1972) for this "payoffs" expression of CAPM.

We wish to take the viewpoint of a prospective investor, who can choose to buy into the risky market. Think of her as having the same risk aversion as "the market", noting that the market chooses to hold amount P_M of its initial wealth W_0 in a risky asset that pays uncertain cash payoff V_M. From here I use these amounts to represent the rational portfolio weights of our single investor.

If V_M is ex ante more certain, meaning lower $\text{var}(V_M)$, its price P_M is higher, provided that the expected payoff $E[V_M]$ is held constant. Higher P_M implies that the investor holds more of initial wealth W_0 in the risky asset, leaving a smaller amount $(W_0 - P_M)$ in risk-free cash.

That larger holding in the risky asset is expected to earn a lower rate of return, since $E[R_M] = E[V_M]/P_M$ is lower when P_M is higher, given that $E[V_M]$ is constant. So what result does that combination of higher P_M invested at a lower expected rate of return have on the prospective investor's overall expected utility?

The answer to this question under CAPM is that in a market with greater certainty, i.e. lower $\text{var}(V_M)$, the ex ante expected utility obtained by taking a CAPM rational portfolio of risky and risk-free assets is lower, which sits well with the basic intuition that we do not want all assets to be risk-free.

The following simple proof, under exponential utility $U(W) = 1 - \exp(-cW)$, assumes that V_M has a normal distribution, $V_M \sim N(E[V_M], \text{var}(V_M))$.

The terminal realized cash from investors' overall portfolio is

$$W_1 = V_M + (W_0 - P)R_f,$$

where again $R_f = (1+r_f)$ is the risk-free return factor, which is assumed to be constant. Define the mean and variance of terminal wealth W_1 as respectively μ and σ^2.

Hence,

$$\mu = \mu(W_1) = E[V_M] + (W_0 - P_M)R_f,$$

and

$$\begin{aligned}\sigma^2 = \sigma^2(W_1) &= \text{var}\left[V_M + (W_0 - P_M)R_f\right] \\ &= \text{var}(V_M),\end{aligned}$$

since P_M is a constant given constants $E[V_M]$ and $\text{var}(V_M)$.

Assuming exponential utility $U(W) = 1 - \exp(-cW)$ for cash wealth W, the equilibrium market price P_M consistent with the usual CAPM is shown in Equation (10.1).

Substituting for P_M in the equation above for μ gives

$$\begin{aligned} \mu &= E\left[V_M + \left(W_0 - \frac{1}{R_f}\left[E[V_M] - c\,\mathrm{var}\,(V_M)\right]\right)R_f\right] \\ &= W_0 R_f + c\,\mathrm{var}(V_M), \end{aligned}$$

revealing that the investor's expected aggregate payoff W_1 is dependent only on $\mathrm{var}(V_M)$. The expected payoff of the risky asset $E[V_M]$ is netted out in the price P_M paid to buy that asset, so makes no difference to the final cash wealth W_1.

By a standard result for exponential utility $1 - \exp(-cW)$, the expected utility of normally distributed terminal wealth $W_1 \sim N(\mu, \sigma^2)$ can be written as

$$EU = 1 - \exp\left[-c\left(\mu - \frac{c}{2}\sigma^2\right)\right].$$

Substituting for μ and σ^2 gives

$$\begin{aligned} EU &= 1 - \exp\left[-c\left(W_0 + c\,\mathrm{var}\,(V_M) - \frac{c}{2}\mathrm{var}\,(V_M)\right)\right] \\ &= 1 - \exp\left[-c\left(W_0 R_f + \frac{c}{2}\,\mathrm{var}\,(V_M)\right)\right]. \end{aligned} \tag{10.2}$$

This result says that the expected utility from an optimal weighted portfolio across the risky and risk-free asset depends on the variance of the risky payoff, and increases with that variance. Hence, the market with higher ex ante payoff variance, $\mathrm{var}(V_M)$, affords prospective investors a higher ex ante (forward-looking) expected utility.

Investors in that market have a lower amount P_M of risky investment, but that dollar amount P_M earns a higher ex ante expected percentage return. That pairing of lower P_M and higher $E[R_M]$, alongside a higher residual amount $(W_0 - P_M)$ earning risk-free interest, yields a total random cash payoff W_1 with higher ex ante expected utility.

If there were no risky asset, the investor's expected utility would be only what is obtained by investing all initial wealth W_0 in the risk-free asset, that is,

$$\begin{aligned} EU &= 1 - \exp\left[-cW_0R_f\right] \\ &< 1 - \exp\left[-c\left(W_0R_f + \frac{c}{2}\operatorname{var}(V_M)\right)\right] \end{aligned}$$

And, conversely, if there exists a risky asset, its price P_M would not exceed zero unless its existence added to expected utility.

Numerical example

The following example illustrates the findings above, and replicates them by calculating expected utility directly by integrating over a normal probability density, rather than by using the mean–variance shortcut expression of expected utility.

The normally distributed random market payoff is V_M and its ex ante parameters are $E[V]$ and $\operatorname{var}(V)$. Its probability density is then

$$f(V_M) = \frac{1}{\sqrt{2\pi\operatorname{var}(V_M)}}\exp\left(-\frac{(V_M - E[V_M])^2}{2\operatorname{var}(V_M)}\right).$$

The utility obtained from the overall portfolio is

$$U(V_M) = 1 - \exp\left[-c\left(V_M + (W_0 - P_M)R_f\right)\right],$$

where as above $P_M = \frac{1}{R_f}\left(E[V_M] - c\operatorname{var}[V_M]\right)$.

Hence, the market expected utility is

$$\begin{aligned} EU &= \int_{-\infty}^{\infty} U(V_M)\, f(V_M)\, dV \\ &= 1 - \exp\left[-c\left(W_0R_f + \frac{c}{2}\operatorname{var}(V_M)\right)\right], \end{aligned}$$

which is exactly as above. Again, the risky market asset adds to expected utility as long as $\operatorname{var}(V_M) > 0$. To be sensible, $\operatorname{var}(V_M)$ is constrained relative to $E[V_M]$ so as to ensure that the price of the market P_M, calculated according to Equation (10.1), is positive.

Expected utility (10.2) is plotted in Figure 10.1 against the standard deviation of V_M, shown in the plot as $\sqrt{\operatorname{var}[V_M]}$. The calculations assume $W_0 = 100$, $R_f = 1.1$ and $c = 0.05$. There is no need to assume any

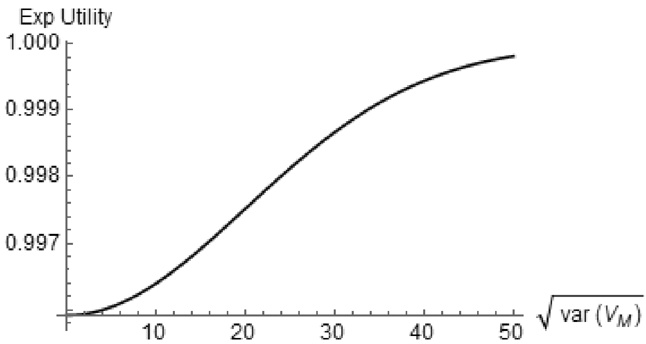


Figure 10.1: Plot of EU obtained from optimal portfolio of risky and risk-free assets as a function of $\sqrt{\text{var}[V_M]}$.

particular value for the mean market payoff, as it makes no difference. It disappeared in the algebra, and is accounted for in P_M.

Note that for the case of $\text{var}(V_M) = 0$, there is in effect no risky asset available and the investment portfolio is entirely risk-free, leaving its expected utility at

$$1 - \exp[-c\,W_0 R_f] = 1 - \exp[-c\,100(1.10)] = 0.996,$$

as shown in Figure 10.1 at $\sqrt{\text{var}[V_M]} = 0$.

The main point is that higher perceived ex ante risk makes for a higher expected utility investment portfolio, provided that investment is understood as a portfolio of the market and the risk-free asset. Expected utility approaches one (satiation) asymptotically as the risk (variance) of the market payoff gets larger. It is the utility expected from the combined investment that determines ex ante welfare.

10.2 Implications for Bayesian decision analysis

The intuition reached is that rather than trying to eliminate risk or uncertainty, ex ante, the role of information, including accounting information, is to assist investors to assess risk and take advantage of it. Given fixed $E[V_M]$, the prospective investor who is best placed ex ante is the one who can invest in a CAPM-priced market that has a higher payoff variance. Rather than hoping to eliminate risk or uncertainty, and thus also opportunity, the ideal is to assess (i.e. "price") existing risk accurately and thereby gain utility from investing in it.

The investor benefits from more risk under CAPM because the price of the risky market is actually less than the risky asset's certainty equivalent. The correct (in utility theory) certainty equivalent of the market payoff, given the utility function $U(W) = 1 - \exp(-cW)$, is[1]

$$CE_M = \frac{1}{R_f}\left(E[V_M] - \frac{c}{2}\,\mathrm{var}[V_M]\right),$$

but the CAPM equilibrium market price is

$$P_M = \frac{1}{R_f}\,(E[V_M] - c\,\mathrm{var}\,[V_M]).$$

So the market pricing mechanism trims the certainty equivalent by

$$\frac{1}{R_f}\left(\frac{c}{2}\,\mathrm{var}[V_M]\right),$$

thereby making investment in the market a positive expected utility action. Note that if the investor paid the full certainty equivalent of the asset, the marginal expected utility would be, by definition, zero.

To understand why the CAPM market price P_M is lower than the certainty equivalent CE_M, remember that (under CAPM equilibrium) capital which could otherwise be left invested at the risk-free rate is diverted to buy the risky asset only up to the point of zero marginal expected utility on the last $\$\delta \to 0$ invested.[2] All of the previous $(P_M - \delta)$ invested was invested with positive marginal expected utility.

The matter of what results are actually realized from actual (i.e. error prone) ex ante expectations is the subject of the next chapter in this monograph. Maximum ex ante expected utility is cold comfort if not ever realized. An ex ante highly risky market payoff presents a prospective investor with high ex ante expected utility under CAPM, but the investor's ultimate objective is not higher ex ante utility but higher ex post utility.

---

[1]e.g. see Christensen and Feltham (2003, p. 54). The numerator (risk-adjusted payoff) in the payoffs form of CAPM is colloquially called a "certainty equivalent" but that is an abuse of the term by the strict utility theory definition.

[2]Johnstone (2017) derives the CAPM explicitly by this criterion.

10.3 Volatility pumping

In portfolio optimization in finance, the strategy called volatility pumping takes advantage of the variance (volatility) in the risky asset, and show how without variance (risk) there is no opportunity. Volatility pumping, explained by Luenberger (1998, pp. 421–438), is an optimization strategy based on the assumption of log utility and given fixed volatility. The investor balances a portfolio of the risky market and the risk-free asset so as to maximize expected capital growth (which objective coincides mathematically with maximizing expected log utility; see the next chapter in this monograph). The CAPM investment strategy does the same except that it maximizes a mean–variance expected utility function, such as exists under the combination of exponential utility and normal distributions. The effect of that utility function is that investment is less risky but lower in its ex ante expected capital growth than the log utility strategy.

11

Ex Post Decision Outcomes

> I have not here discussed what the basic probability distributions are supposed to come from. In whose mind are they ex ante? Is there any ex post validation of them? (Samuelson, 1965, p. 48)

> Realized profits, not maximum profits, are the mark of success and viability . . . those who realize profits are the survivors; those who suffer loses disappear. (Alchain, 1950)

The "value of information" is understood in archetypal Bayesian models, like the oil wildcatter (e.g. Raiffa, 1968), as an ex ante assessment. But information or experiments with high ex ante expectations are known to be only a means to an end. The proof of the pudding or eventual test of the user's information, model, posterior beliefs and decision rule is her realized utility. Ultimately Bayesian decision analysis must include methods of review, including for example the formal probability "scoring rules" that were invented by de Finetti, Savage, Lindley, Good and others specifically for that purpose.

The end, or ex post perspective, is how signals actually perform for users in terms of "making money" or realized capital growth. That ex post view of information quality gives rise to questions about how "good", and in what sense of "good", information has to be for a user with a given decision rule (or utility function) to accumulate investment profits?

Can information that leaves relatively little certainty be sufficient? Less likely perhaps, can information that brings lower certainty, and hence smaller risky investments, yield higher capital growth? In essence, what does a rational decision maker need of information to "make money"?[1]

The stream of accounting theory beginning with Demski and based on Bayesian statistical decision theory does not focus on what makes information and information systems good ex post. Ex ante optimization is the main focus, and ex ante expectations are implicitly presumed to be borne out on average.

One exception is Feltham, who defined "ex post relevance" by whether the realized signal changed the user's decision, but did not consider the outcome of that changed decision:

> ...if a signal changed the decision then the information provided by that signal was relevant. (Feltham, 1968, p. 691)

The ex post perspective in Bayesian statistics goes to the next step of asking whether users' revised beliefs and decisions work out well in accord with the user's objectives.

The large Bayesian literature on probability "scoring rules" sets out a theory of ex post evaluation of probability assessments. See DeGroot and Fienberg (1982) and DeGroot and Fienberg (1983), Lindley (1982a) and Lindley (1982b), Jose *et al.* (2008) and Johnstone (2011). The literature on economic Darwinism, discussed below, looks at how users

[1] Ex post, utility and money are the same thing in the sense that utility is increasing monotonically in money. Hence, ex post results can be measured in money. That does not mean however that the ex ante objective can be written as maximizing expected money. That would imply risk neutrality. It is only ex post that the two perspectives coincide.

translate information to beliefs, to actions, to profits (or losses). The objective of this literature is to identify the ex post characteristics of beliefs that help investors to "beat the market" or "make money".

Empirical accounting research does something that is equally "ex post" by examining the observed stock price reactions and apparent changes in the cost of capital ensuing after changes in accounting methods and earnings quality. These studies are in essence an ex post evaluation based on outcomes, rather than ex ante expectations. That empirical research approach is, however, more focussed on market-level averages than the specific decision outcomes of specific accounting information users.

11.1 Practical investment

When supporting a mean–variance decision theory rather than a more general expected utility theory, Tobin (1969) held that a business decision maker will hardly be amused by the prescription that:

> ... he should consult his utility and his subjective probabilities and then maximize. (Tobin, 1969, p.14)

Contrary to Tobin's disregard, there is much theoretical and empirical study devoted to portfolio selection by optimization of expected utility functions, both directly and by their expression through higher moments.[2] A hallmark of this work is its attempt to make expected utility methods practical in modern portfolio management. The statistical mathematics of utility maximization of portfolios go back to the famous works of Latane, Shannon, Kelly and Thorpe, and more recently Ziemba, who assumed that the natural investment objective is to achieve an agreeable compromise between capital growth and volatility.[3] In effect, that objective function amounts to making acceptably few large losses along a path of average compound growth — which is essentially what a suitably risk-averse utility function is meant to achieve. Any utility

[2] See for example MacLean *et al.* (2005) and Cremers *et al.* (2005).

[3] The realized Sharpe ratio of an investment fund captures how well the fund actually performed in its pursuit of growth against volatility.

function that is increasing and concave implies the same general ex ante desire to "make money" while avoiding risks that are too large or too probable for what growth they offer.

The mathematics of capital growth (Kelly, 1956) became the basis for the statistical theory of economic "Darwinism", developed by Alchain (1950), Blume and Easley (2006) and Sandroni (2000), and in a series of papers co-authored by Maclean and Ziemba and others (see references). A retrospective is provided by Hakansson and Ziemba (1995).

There is disagreement among Bayesians, because the "growth optimal" strategy, which favors growth at the cost of higher volatility, and is commonly known as "Kelly betting", calls for investors to adopt log utility, whereas the fully "personal" subjectivist Bayesian position allows the investor to "have" or choose her own utility function, whatever its risk appetite.

In the practical end, there is no issue. Theorists who have explored growth optimal investment know in great detail its strengths and weaknesses. They make allowance for more risk-averse investors by identifying desirable tradeoffs between growth and "safety" achieved by practical decision rules like "half-Kelly". MacLean *et al.* (2005) and Kadane (2011) show general conditions for correspondence between fractional-Kelly investment rules and power utility functions of differing risk aversions. For a subjectivist Bayesian critique of this approach, see Kadane (2011).

A fundamental insight emphasized by Ziemba in the growth-safety tradeoff literature, but applicable to all investment, is that errors in investors' subjective probabilities that lead to over-betting are more costly than errors that lead to under-betting. In other words, subjective over-confidence, reflected in the direction and amount of the bet, is the most costly mistake or bias in investment. The implication for accounting standard setters from this literature is that investors do not benefit, and may in fact be bankrupted, by inaccurate probability beliefs (e.g. over-certainty).

Ziemba's argument contains practical notions of under- and over-betting that do not explicitly involve any particular utility function or decision rule. All that comes into calculation is the fraction of wealth ρ invested in the risky asset, irrespective of how ρ is chosen. The investor's expected exponential return is explored as a function of ρ, without needing to consider how ρ was chosen by the investor, or what utility function or decision rule drove it.[4]

11.2 Economic Darwinism

Take the most elementary investment context. Suppose that the decision maker invests (bets) some proportion of her wealth on the proposition that a firm will succeed. The firm pays $V = \$1$ per share if it succeeds and $V = 0$ otherwise.[5] The current share price is (say) $q = 0.50$. Let the "true" probability that the firm succeeds be $\pi = p(V = 1) = 0.55$,[6] and imagine that the decision maker uses proportion $\rho \equiv \rho(p)$ of her capital to buy shares in the firm, where p is her subjective belief in the firm's success.

The proportion $\rho(p)$ is fixed by the investor's utility function and is generally higher for a less risk-averse individual.[7] The user's investment is driven by her probability belief p and hence by the information she receives and the meaning she places on it. Implicitly, better accounting information makes for a more accurate assessment p, which (as we see below) does not call for greater resolution or p nearer certainty (0 or 1). Instead greater certainty and more accurate beliefs are often conflicting ideals.

It is easily found by simple algebra that her realized or "physical" expected compound interest (exponential) return factor from the

[4]Ziemba holds that any ρ or "over-betting" beyond that for which the expected exponential return is maximum is "irrational" because it produces the unfortunate pairing of greater volatility and lower expected capital growth. Subjectivist decision analysts do not respect this maximum. They hold that the utility function of a risk-averse rational investor might permit bigger risks than those accepted by a growth optimal (log wealth) investor.

[5]This asset exists and is traded on major exchanges including the Chicago Board Options Exchange (CBOE) and NYSE American.

[6]Alternatively, π can be defined as the actual proportion of outcomes $V = 1$ in the same actual set of bets. See Lewellen and Shanken (2002) for distinction between the physical or true probability distribution and a subjective distribution believed under available information by the market or by the single investor. In accounting, Francis *et al.* (2006) refer to the "true distribution" of earnings.

[7]See Johnstone (2011) for specific fractions under different HARA utility functions.

investment $\rho \equiv \rho(p)$ is

$$\exp\left(\pi \log\left[(1-\rho) + \frac{\rho}{q}\right] + (1-\pi)\log[1-\rho]\right).$$

For example, if she bets $\rho = 10\%$ of her wealth, she will return on average a factor of approximately 1.005, implying a compound return of approximately 0.5% per trial, which is the highest expected capital growth physically feasible under $\pi = 0.55$ and $q = 0.5$.

Figure 11.1 shows the expected return factor (per trial) as a function of the proportion ρ of wealth invested. Note that if the investor risks more than a 20% proportion of her wealth, her expected return factor falls below one, and she therefore loses wealth on geometric compound average. The size of this expected loss increases rapidly with greater over-betting, implying a rapid stochastic path to ruin (note how quickly the average return factor falls away as ρ increases). Over-betting and over-confidence have essentially the same appearance in terms of realized outcomes (mathematical theorists in gambling advise investors to purposely "under-bet" relative to their beliefs, so as to allow for potential over-confidence in those beliefs; see for example MacLean *et al.*, 1992, MacLean and Ziemba, 1999, MacLean *et al.*, 2004).

This simple example reveals that even when the investor has sufficiently good information to bet "in the right direction" (long or short),

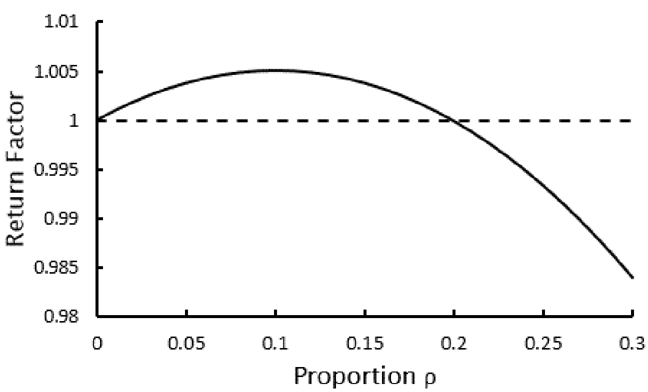


Figure 11.1: Expected return factor versus fraction invested.

she can still over-invest or over-bet and lose heavily. In other words, it is possible to "bet too much on under-priced stocks".

She "bets" in the right direction, by going "long" on the asset, since the market price of 0.5 is less than the probability π of the firm succeeding (paying \$1 per share), so she has "an edge" that can be exploited. The problem however is that she errs by investing too heavily in that direction. Physically, her resulting losses are not outweighed by her gains, and she loses money on average, despite being onto a good thing. Remarkably, an information advantage or "edge" is converted to a financial loss, despite the fact that the investor buys an asset at a market price lower than its intrinsic or true value.[8]

The potential of the investor's decision rule to convert an information advantage to a money loss is rarely discussed. That "mistake" occurs quite rationally in the example above under any risk-averse utility function for which $\rho > 0.1987 \approx 20\%$. Rational decision theory, i.e. subjective expected utility theory, allows the user to have such low-risk aversion, so even homo economicus can lose by maximizing the expectation of a sufficiently risk-tolerant utility function, despite being given sufficiently good information to win.[9]

11.3 Bayesian Darwinian selection

Discussion of economic Darwinism in accounting started with Verrecchia (1979) and came out of questions about how an investors' belief distributions compared with the "true" distribution.

For any investor, accurate probability assessment is a footing on which to make money, suggesting reason of itself to "be Bayesian". In arguing for why it makes sense to model the agents in strategic

[8]When Thorp realized that in blackjack he had an edge in some situations against the casino, he introduced Kelly's (1956) log utility decision rule as a way to maximize expected capital growth from that information advantage. Modified versions of the Kelly rule convert an advantage to slightly slower average capital growth but with much less volatility. That is essentially the subject of the "Ziemba" papers in finance and management science.

[9]If a rational investor with risk-averse utility believes probability 0.55 against the market's 0.5, her investment fraction approaches 100% as her utility function approaches linearity (risk neutrality).

information disclosure models as Bayesians, rather than to adopt any behavioral finance depiction of decision makers and inference makers, Dye (2001) stated the Bayesian's case as follows:

> In a competitive market setting, the presence of non-Bayesians does not directly influence any Bayesian investor's actions, since in a competitive setting, no single investor believes that he, or anyone else, has an impact on market prices. That is, competitive settings are like single person decision settings in which everyone perceives himself as acting in isolation of everyone else. Now in any single person decision setting, the best information processor always wins! Thus, Bayesians must generate higher operating profits than non-Bayesians in competitive markets. (Dye, 2001, p. 213)

Similarly, Verrecchia (2001) puts the Darwinian hypothesis as follows:

> . . . Bayesian investors make statistically correct portfolio rebalancing decisions (on average) in the presence of disclosure, whereas heuristic investors make inferior portfolio rebalancing decisions. Consequently, over time Bayesian behavior should outperform heuristic behaviors, and, for this reason, presumably drive heuristic behaviors from the market. (Verrecchia, 2001, p. 126)

Underlying both Darwinian statements is the natural assumption that the rational Bayesian will not negate his probability advantage by over-betting or by applying an overly bold decision rule. That is perhaps a highly realistic assumption where survival is in play, because an investor who has the cognitive ability to judge the correct direction in which to bet (long or short) will presumably adapt equally adeptly if she is over-betting, or at least she will if she can separate that mistake ex post from any thought that she must be getting the probabilities badly wrong, and provided as well that she survives for a sufficient while to adjust.

The interesting mathematical fact of economic Darwinism is that a non-Bayesian might beat a Bayesian merely by having a better "chemistry" between her probability assessment skill and her trading rule. Indeed, rank ignorance can beat an over-betting Bayesian, as occurs in the example calculations above, where an informed trader (call him a Bayesian), with an edge, bets on a binary event against a market probability of 0.5, yet succeeds only in losing overall by betting too much. A market probability of 0.5 is the "ignorance probability" in a binary event, and is often sufficient to beat a better informed but incautious investor.[10]

Dye's and Verrecchia's presumption of the market "selecting for Bayesians" surely holds true when learning is allowed. In a stationary game, the Bayesian's probability edge over any non-Bayesian will tend to increase trial-by-trial, albeit not always monotonically. A final point is that in any competition between different Bayesians all with the same decision rule, their respective sources of information, and their decisions about what information to obtain and how much to pay, should be the decider. On this point, (Dye, 2001, p. 215) explains how less well-informed traders can survive in a market:

> . . . perhaps their decision not to become better-informed was optimal: the extra trading profits they have foregone by not becoming better-informed might have been offset by the information acquisition costs they have not incurred. (Dye, 2001, p. 215)

That explanation anticipates an important part of the modern finance literature on the "limits to arbitrage", which reveals practical aspects of why less well-informed investors, or investors too "lazy" or "miserly" to obtain information, or not good at informed analysis, might well survive and even prosper.

In the end, being both Bayesian and making money still requires an element of "luck", and that might in some bets be the luck of

[10] Blume and Easley (2006) describe the interaction effect of an investor's probability assessment ability and decision rule as her "entropy". I would suggest that a better term is "chemistry", for how the two react with each other to make the end result.

having less information. Certainty must rationally rise and fall with new information, at least in the short run, and hence the final piece of information might be the one that caused the Bayesian to bet in the direction and amount that she did. If she lost the bet, it was that extra bit of information that effectively "lost it".

An interesting moral of this story is that the investor who receives the best "Blackwell-ranked" signal, because of its one extra degree of "fineness", is bound to be sometimes the trader who bets in the ex post wrong direction. That is because the probability assessment $p(V|x_1, x_2, \ldots, x_{n-1}, x_n)$ can be higher or lower than $p(V|x_1, x_2, \ldots, x_{n-1})$, thus leading to not only a different size of bet, but a bet potentially in the opposite direction, as would occur when the market probability π lies in the interval between $p(V|.., x_{n-1})$ and $p(V|.., x_{n-1}, x_n)$.

11.4 Good probability assessments

There are statements throughout accounting literature suggesting that accounting information is meant to assist decision makers to form and revise subjective probabilities. For that reason, the literature in Bayesian statistics on what makes a "good probability" assessment is of direct and fundamental interest. See for example de Finetti (1937/1964) and DeGroot and Fienberg (1982) and DeGroot and Fienberg (1983). The issue of good probability assessment, and how to assist that task, has been studied theoretically and empirically in meteorology, and the messages coming from that highly developed decision-theoretic literature apply almost verbatim in accounting.

A very quick summary, clarified and illustrated empirically by Gneiting *et al.* (2007) and Gneiting and Raftery (2007), is that the best probability assessment is the "sharpest" (most resolved) subject to calibration. Suppose, for example, that the task is to assess the probability that firms will go bankrupt (before set time T). Good probabilities are (i) calibrated in the sense that $f(Bankrupt|p) = p$ for all p, where f represents a relative frequency, and (ii) sharp, in the sense that they are nearer 0 or 1.

Note that calibration is a joint property of probabilities and realized events, whereas sharpness is a property of the probabilities of themselves. By emphasizing ex ante information value rather than ex post outcomes, accounting theory gives precedence to sharpness and implicitly ignores calibration. Empirical accounting research allows for both, because it looks back at the apparent correlation between information qualities and ex post outcomes.

In economic contexts, it is possible to imagine that ex post investment success comes from a blend of calibration and sharpness. Probability scoring rules, as introduced to accounting by Scott (1979), merge the two probability assessment characteristics into a single score, which can be tailored to match a given user's utility function.

This picture of what makes a good probability assessment can be integrated into accounting theory. As in meteorology, the objective is to generate information that facilitates good probability assessments and assists good probability assessors.

"Good" in this context is a measurable concept related to ex post outcomes. Although Bayesian subjective probability theory rejects much of frequentist theory, one unifying element is the Bayesian use of observed frequencies as an indicator of probability accuracy. See (Gelman *et al.*, 2004, pp. 111–112) on ex post frequency evaluations of Bayesian inferences.

11.5 Implications for accounting information

Rather than needing always more resolution, users benefit from any information that reveals how unreliable and hence uncertain the firm's cash flow is — or is not. Revelation, rather than resolution, of uncertainty might seem like the antithesis of accounting expertise, and will not satisfy firm management's usual desire for greater shareholder confidence. On the other hand, misplaced investor over-certainty is a sure way to economic loss.

There is no suggestion that the accountant should influence the way that users convert information or beliefs p into investment portfolio weights, $\rho(p)$. The proportion of wealth risked on a given probability belief p, or on given information, is understood as at the user's prerogative, and is merely assumed to be rational by the user's own personal utility function. Rather than counselling the investor, the FASB-envisaged role of the accountant is presumably to facilitate "accurate" or better founded beliefs p, even when that service requires a blow to investor confidence.

A "going concern warning" or a large asset write-down or drop in reported profit might of itself serve that purpose. Similarly, notes to accounts that state explicitly that there is reason for concern about the viability of any part of the firm's business or assets, or that there is insufficient evidence to be confident of their success, might often assist users to form more propitious probability beliefs about the firm's future cash flows. Every "Enron" is an instance of how doubts and evidence of financial difficulties could have been valuable if raised in the accounts, rather than dispelled, concealed or not apparent.

The most fundamental message of the economic Darwinism literature in management science and economics is that decision makers are obviously better served by information that incites beliefs that prove accurate ex post, rather than by information that is perceived valuable ex ante. Any information that changes ex ante beliefs seems valuable ex ante, but need not prove valuable ex post. The ideal is to achieve both, but if that is not possible, high ex ante certainty can prove very costly ex post.

12

Information Uncertainty

The term "information uncertainty" does not arise in Bayesian theory and sounds like a contradiction in terms. To a Bayesian theorist like Savage or Lindley, information is not a type or description of uncertainty, but is the antidote and only possible way out of uncertainty (albeit that is not guaranteed).

Outside Bayesian theory, in both the accounting and finance literatures, the terms "information risk" and "information uncertainty" are used largely interchangeably, but neither has an agreed or formal definition:

> One impediment to empirical work in this area has been the absence of any guidance as to what constitutes "information risk" and when, or whether, it should be priced. (Lambert, 2010, p. 5)

Loosely, "information risk" and "information uncertainty" describe perceptions about given signal ω, with respect to given unknown V, that preclude conclusive inference from ω to V (e.g. a broker says "Buy" but it is perceived that brokers nearly always say "Buy", so there is little added confidence about a stock price increase).

Definitions of information uncertainty (or risk) in accounting literature reveal numerous facets of accounting information that researchers recognize as defects and potentially detrimental to the stock price and cost of capital. Descriptive non-technical definitions of information uncertainty include the following:

(i) "the likelihood that firm-specific information is of poor quality" (Francis *et al.*, 2005),

(ii) "ambiguity with respect to the implications of new information for a firm's value" (Zhang, 2006),

(iii) "'value ambiguity,' or the degree to which the firm's value can reasonably be estimated by the most knowledgeable investors" (Jiang *et al.*, 2005)

(iv) "imperfect information causing forecasts to be risky" (Healy and Palepu, 2001),

(v) noisy or incomplete information as when earnings quality is low (Yee, 2006),

(vi) earnings manipulation (Strobl, 2013),

(vii) "the [in]ability of investors to ascertain the valuation parameters underlying a particular asset" (Riedl and Serafeim, 2011).

The most common technical definition of information risk appeals to the concept of estimator precision (sampling variance) from classical frequentist statistics. Risky or less than perfect information is characterized as being "imprecise", "high variance", "unreliable"[1] or "noisy".[2] One clear statement is as follows:

[1]Dechow and Schrand (2004, p. 8) state that a "reliable" number is one that does not involve much judgment and is easily verified by another accountant (e.g. cash at bank), implying that it has low subjective variance.

[2]Botosan (1997), Healy and Palepu (2001), Botosan and Plumlee (2002), Beyer *et al.* (2010), Cheng *et al.* (2011) and Core *et al.* (2015) link the definition of information risk as precision to the parameter estimation risk literature in finance.

> We identify "quality" of information in the capital markets with a statistical notion, specifically the precision of a measure with respect to a valuation relevant construct. For a given construct, higher quality information is more precise (contains less uncertainty) with respect to that construct. (Francis *et al.*, 2006)

Similarly, see for example Veronesi (2000), Francis *et al.* (2005) and Francis *et al.* (2008), Yee (2006), Cheng *et al.* (2011), Lambert *et al.* (2007), Lambert and Verrecchia (2010), Kravet and Shevlin (2010), Tang (2011), Zhang (2006), Rajgopal and Venkatachalam (2011), Verdi (2012), Bhattacharya *et al.* (2012) and Armstrong *et al.* (2016).

Another accepted way of thinking is to describe, or proxy for, information risk by its mirror inverse called "earnings quality" or "accrual quality". Francis *et al.* (2005) suggest that for empirical work "accrual quality proxies for information risk". Core *et al.* (2008), Kim and Qi (2010) and Bhattacharya *et al.* (2012) adopt this same notion. Completing the circle, information quality is characterized by Verrecchia (1990), Ng (2011), Veronesi (2000) and Lambert *et al.* (2007) in terms of its low variance (high precision).

> By earnings quality, we mean the precision of the earnings signal emanating from the firm's financial reporting system. (Francis *et al.*, 2008, p. 54)

A Bayesian characterization of evidence is sufficiently general to incorporate all of the information properties that have been mentioned, including notions of "signal quality", "noise", "reliability", "precision", "ambiguity" and the like. Although these terms have mainly frequentist/objectivist overtones, and in some cases long established frequentist meanings, the various definitions of information risk listed above, and common in accounting, make clear sense Bayesianly and can be unified and generalized using the pivotal Bayesian construct; to wit, the likelihood function.

Emphasizing its elegance and the role of the likelihood function, Bayes theorem is written compactly as

$$f(V|\omega, \Omega_0) \propto f(V|\Omega_0)f(\omega|V \cap \Omega_0),$$

where ω is new information and Ω_0 is pre-existing (prior) information or background knowledge. The qualities of ω with regard to V, to the extent that they are perceived by the decision maker, must be described within the subjective likelihood function, $f(\omega|V \cap \Omega_0)$. Any doubts or risks or unknowns that harm ω as an indicator of V must be incorporated formally at this point, otherwise they are misrepresented, with whatever generally costly effect that has on the *ex post* accuracy of the user's posterior beliefs.

Likelihood functions are subjective, like all belief distributions. Explicit admission of subjectivity is a strength of the Bayesian model, because it allows the decision maker to incorporate any background knowledge that matters to the perceived quality, and hence evidential meaning, of signal ω. For example, suppose that the signal is "+" and the two alternative hypotheses are "*up*" and "*down*". If less faith is attached to this signal, the two likelihoods $p(+|up)$ and $p(+|down)$ are set closer to equal. By making them exactly equal, the signal + is depicted as noise and irrelevant to beliefs regarding which alternative is the more probable.

A conceptual attraction of Bayesian statistical logic is that all aspects of perceived information quality are encapsulated in one place — the likelihood function (see the earlier drawings of hypothetical likelihood functions for "high" earnings).

Frequentist, orthodox or classical methods, by comparison, depict the quality of an estimate $\widehat{x}$ of an unknown parameter θ as either multidimensional, or more arbitrarily as a composite of its various signal attributes, including its bias, variance and other moments, and its other (possibly asymptotic) behaviors under repeated sampling. This aspect of classical statistics causes problems in accounting research where, for tractability in models, the quality of information is depicted by a single parameter (Dye and Sridhar, 2007). That parameter is almost always the precision (variance) of the estimate, thus either ignoring bias or presuming zero bias, notwithstanding the vast literature on accounting conservatism and earnings management and manipulation.

There is no objective way to order or weight the relative importance of an estimate's bias, variance, consistency and other singular signal attributes.[3] Note however that, by bypassing this issue, Bayes theorem does not imply that bias, precision and other classical signal error attributes are unimportant. Rather, Bayesian logic requires that these are all considered and given expression within the subjective likelihood function $f(\omega|V,\cdot)$. Their relative "weighting" occurs naturally and implicitly by the way that each such perceived estimator attribute has some greater or lesser impact on the location and shape of the likelihood function $f(\omega|V,\cdot)$.

For example, suppose that ω is a reported measurement of V. It may be known or perhaps merely suspected that ω is an imprecise observation on V, in which case $f(\omega|V,\Omega_0)$ is relatively diffuse. Importantly, the likelihood distribution $f(\omega|V,\Omega_0)$ can be shifted and reshaped to different degrees for different realizations of ω, and may in principle take different distributional forms for different signal realizations. Highly informative signals ω have characteristically peaked likelihood functions, and less informative (e.g. more imprecise) signals have relatively flat or more diffuse likelihood functions.

It is important to note that a peaked likelihood function, while being Bayesianly informative, is not always representative of a good, honest or reliable information source. Its information value is nonetheless obvious, and may in fact be due to its perceived dishonesty (if the firm reports an earnings result that the analyst knows to be false, the likelihood function for that report will be positioned and shaped to incorporate the perceived dishonesty).

In a binary model where $V = 1$ (good) or $V = 0$ (bad), a signal "+" with high Type I error probability $f(-|V = 1) = 0.7$ and high Type II error probability $f(+|V = 0) = 0.9$ (say) has a highly informative likelihood $f(+|V = 1)/F(+|V = 0) = 0.3/0.9 = 1/3$, and gives rise to a Bayesian conclusion $f(V = 1|+)$ that is strongly in the opposite direction to the reported signal (with prior $f(V = 1) = 0.5$, the posterior is $f(V = 1|+) = 0.25$). This signal is informative because it leads to a strong inference about V, but is a very inaccurate indicator if translated naively or literally rather than Bayesianly.

[3]See Gelman *et al.* (2004) for a Bayesian critique of any ordering or hierarchy of classical estimator attributes.

Given that the likelihood function is the Bayesian summation of everything known or merely questionable about information ω, I summarize informal and intuitive existing definitions (see above) with the following explicitly Bayesian definition of information risk or information uncertainty.

12.1 Bayesian definition of information uncertainty

With background knowledge Ω_0, the risk (uncertainty) associated with information ω, regarding unknown state or value V, is any perception or doubt concerning ω that affects its subjective likelihood function $f(\omega|V, \Omega_0)$ such that Bayesian conclusions $f(V|\omega, \Omega_0)$ conditioned on ω, about V, are contrary to ω or generally less influenced by ω.

This definition has the distinctly Bayesian aspect that information uncertainty is understood as a perception or subjective belief, rather than as a physical property of signal ω, thus giving an explicit role to all relevant background knowledge Ω_0 (e.g. the awareness that, however reliable ω seems with regard to $\boldsymbol{V}$, the most subjectively reliable looking signal can have that appearance merely because of what little we know).

In this way a Bayesian can allow for errors in ω that arise out of conditions that are not identifiable before the fact. Bayesianism requires that any feasible error, even if given near zero probability *ex ante*, be admitted and accounted for when subjectively assessing the shape and location of $f(\omega|\boldsymbol{V}, \Omega_0)$.[4] Note again that existing "objective" knowledge about ω, including its empirical likelihood function (observed error frequencies) observed over many repetitions, can have major influence on the subjective assessed $f(\omega|\boldsymbol{V}, \Omega_0)$.

[4] It is for this "Black Swan" type reason that Bayesian portfolio theory has gained wide industry acceptance after the financial crisis. The eminent Bayesian theorist Dennis Lindley invented "Cromwell's Principle" (Dawid, 1982), on the basis that Cromwell told the Church of Scotland to at least countenance a chance that it could be wrong.

12.2 Bayesian treatment of information uncertainty

The Bayesian position is that all uncertainty, no matter what its "type" or source, is assimilated within the posterior joint probability distribution $f(V_1, V_2, \ldots, V_n|\Omega)$, where existing information Ω includes information about information (e.g. Ω might include the sample observation x and of course the sample size n, and also collateral information regarding where x was obtained, under what experimental controls, and how it might be misleading).

In principle, Ω contains all existing knowledge, including anything that may help the decision maker interpret newly observed information, including any evidence that a particular class of signals is highly imprecise or unreliable, or the opposite. The net effect is that the Bayesian posterior distribution $f(V_1, V_2, \ldots, V_n|\Omega)$ makes allowance for not only the particular signals that have been observed, but also for information and beliefs about their reliability or about the mechanisms (e.g. instruments, experts or models) by which they were observed, and via which they could possibly fail.

There can be no clear separation of "fundamental" uncertainty, meaning uncertainty "arising from nature", from uncertainty due only to perceived imperfections in the signals by which natural processes are observed (Kalymon, 1971). Ultimately, all that is possible is that we incorporate what we observe or understand, including what we perceive about the strength or reliability of that knowledge or data, into a summary distribution absorbing all elements of our uncertainty at once.

Suppose that a signal ω arises randomly from one of a number of possible sources indexed by θ, each with its own perceived error characteristics, making its information qualities random (as in Subramanyam, 1996; Wagenhofer, 2011; Ewart and Wagenhofer, 2011). The total available information $\Omega = \Omega_0 \cap \omega$ includes signal ω, along with the prior (pre-existing) information Ω_0. Note that $\Omega = \Omega_0 \cap \omega$ is the intersection of prior information Ω_0 and new information ω, meaning that ω can negate or reverse the meaning of part or all of Ω_0, and *vice versa.*

Background information, or the conjunction Ω of background information and signal ω,[5] might suggest that ω was generated by a given source, parameterized by θ, with probability $f(\theta|\boldsymbol{V},\Omega)$. This distribution, representing the probability that signal ω came from source θ, is conditioned on $\boldsymbol{V}$ to allow for the possibility that the underlying state of $\boldsymbol{V}$ affects the source or error properties of the signal (e.g. high earnings might encourage less manipulation of reported earnings).

The required posterior distribution is then

$$f(\boldsymbol{V}|\Omega) = f(\boldsymbol{V}|\Omega_0 \cap \omega) = \frac{f(\boldsymbol{V}|\Omega_0) f(\omega|\boldsymbol{V},\Omega_0)}{f(\omega|\Omega_0)},$$

where

$$f(\omega|\boldsymbol{V},\Omega_0) = \int_\theta f(\theta|\boldsymbol{V},\Omega_0) f(\omega|\boldsymbol{V},\Omega_0,\theta) d\theta,$$

and

$$f(\omega|\Omega_0) = \int_\theta f(\boldsymbol{V}|\Omega_0) f(\omega|\boldsymbol{V},\Omega_0) d\theta.$$

In these calculations, θ is a "nuisance parameter" and is integrated out in the standard way that Bayesian inference eliminates nuisance variables. Note also that the information risk or perceived "error properties" of each possible observer state θ are embedded within likelihood function $f(\omega|\boldsymbol{V},\Omega_0,\theta)$. It can be seen therefore that the Bayesian posterior distribution incorporates information risk, provided of course that the decision maker conditions this distribution (as in the example above) on all that is known or inferred about that risk. That conditioning is the obligation of the decision maker, not of Bayes theorem.

It is impossible to use information coherently without allowing for its perceived "information risk". If a weather forecaster reports "rain", then we cannot find the probability of a sunny day without first assessing the likelihoods p(forecast "rain"|sunny) and p(forecast "rain"|rain), which jointly depict the perceived information risk of the forecast. Whether these two "error characteristics" capture all the surrounding information risk as well as possible is a matter for the

[5]The signal realization combined with background knowledge can often tell much about its own unknown source, e.g. an apparently overly enthusiastic or bold forecast can raise immediate doubts about the motives or expertise of its own source.

decision maker's subjective assessment. For example, if the decision maker believes that $p(\text{"rain"}|\text{sunny}, \theta) \neq p(\text{"rain"}|\text{sunny})$, meaning that the forecast's perceived error characteristics change depending on a condition or parameter θ, then the possibly uncertain state of θ must be taken into account when assessing today's likelihood of forecast "rain", $p(\text{"rain"}|\text{sunny}, \cdot)$. That would require a further level of integration to average over all possible values of θ, allowing for their subjective probabilities conditioned on all current information.

12.3 Model risk as information risk

If a payoff parameter has a given posterior distribution under Model A and another under Model B, the Bayesian technique is to integrate out model uncertainty by weighting each distribution by the probability of the associated model conditioned on the most up to date available evidence (including sample data). That technique was used in my generalization of the Dye and Hughes (2018) model.

The same technique is used in the context of possible regime shift in the returns process, where each distinct regime is described by a different model. It goes without saying that if new data suggests that the firm has shifted its operations to a higher volatility regime, then that evidence will sometimes add to the predictive market returns variance.

To a Bayesian, model risk is just a generalized form of parameter risk. Not only are the parameters of the assumed model uncertain, so is the model itself uncertain. Bayesian allowance for model risk must obviously affect beliefs and the cost of capital, with virtually no exception. The Bayesian treatment of model risk is in principle as follows. Starting with a probability distribution for $\boldsymbol{V}$ under assumed model $m = A$, $f_{m=A}(V|\Omega_0)$, information ω comes to light. Under prior information Ω_0, ω casts doubt on the validity of Model A. Other conceivable models are $m = \{B, C, \ldots\}$. The new predictive distribution for $\boldsymbol{V}$ is then the weighted average distribution

$$\int_m f_m(\boldsymbol{V}|\Omega_0 \cap \omega) f(m|\Omega_0 \cap \omega)\, dm.$$

This is the same process of integrating out a nuisance variable as described above. Competing models with positive probability of being true are “kept in play” (Kadane, 2011). This approach is a hedge against the event of a “very different” and low prior probability model being the “true model”. The different models will put different predictive variances on the variable of interest, say a payoff, so when we average across models, rather than simply picking one model, the predictive variance can rise or fall. That again reveals how uncertainty is hard to tame, and predict.

13

Conditioning Beliefs and the Cost of Capital

Bayesian inference is really no more than conditioning, which is why Bayesian inference, before it was called Bayesian, was often described as conditional probability.

Probabilities are revised or conditioned on new information, and ultimately depend on how far they are conditioned. That is saying merely that the probability of some event like XYZ Corp. going bankrupt changes as we add further information or conditions, so, for example:

$$p(\textit{Bankrupt}|\textit{small firm, gold mining}) \neq p(\textit{Bankrupt}|\textit{gold mining}),$$

noting that the addition of each new condition can increase or decrease the probability of bankruptcy. Risk assessment is thus a stratification process of finding relevant conditions and incorporating their effects as acutely as possible. That is the underlying rationale of experimental control, and is acutely evident in the calculations surrounding Simpson's paradox.

The risk assessment and cost of capital applied to firms in the stock market is akin to the different risk premia charged to different individuals by an insurance company. Usually, when new information comes to light, insurance companies increase the premium charged to some, and reduce it to others. Based on the insurance analogy, it seems unlikely that better information, that allows better discrimination between firms in the market, can reduce the cost of capital for all firms. A more natural presumption is that the market can benefit from better financial disclosure in the same way as the life insurance industry uses better medical tests to identify relevant subsets of the population. A relevant subset is a subset for which the conditional (i.e. Bayesian) probability of the event in question is unique to that subset, or differs from the unconditional probability. See the earlier discussion on the Bayesian concept of subjective "exchangeability" between individuals or samples.

By partitioning the population into statistically relevant subsets, some individuals are seen as materially higher or lower than average risk, and are priced accordingly. This is akin to the "lemons" example of Healy and Palepu (2001) where better information allows the market to discriminate in price between good ideas and bad ideas.

It turns out, however, and is obvious in hindsight, that better discrimination between firms in the market can either increase of decrease the cost of capital, not only for individual firms or subsets of firms but also on average across the whole market.

The following summarized example is from Johnstone (2015). The assumption is that investors have risk-averse quadratic utility $U(x) = x - \frac{b}{2}x^2$ for money x ($b > 0$).

There are just two risky assets, $j = 1, 2$, and each has uncertain payoff $V_j \in \{0, 1\}$. A representative risk-averse investor with quadratic utility[1] allocates wealth W in weights w_1, w_2, and $(1 - w_1 - w_2)$ respectively between the two risky assets and the risk free asset. Starting with normalized wealth of $W = 1$, the investor selects risky asset weights $w_1 = w_1^*$ and $w_2 = w_2^*$, subject to given asset prices P_1 and P_2, with the objective of maximizing expected utility,

$$E\left[(w_1R_1 + w_2R_2 + (1 - w_1 - w_2)R_f) - \frac{b}{2}(w_1R_1 + w_2R_2 + (1 - w_1 - w_2)R_f)^2\right],$$

[1] So as to stay with mean–variance, since the asset payoffs are Bernoulli rather than normal.

where $R_1 = V_1/P_1$, $R_2 = V_2/P_2$ and $R_f = (1 + r_f)$ are the associated asset return factors. In the calculations below, it is assumed that $R_f = 1.10$ and $b = 1/3$.

Letting $R_{port} \equiv (w_1 R_1 + w_1 R_1 + (1 - w_1 - w_2) R_f)$ represent the investor's price-weighted portfolio return, and taking advantage of the identity

$$\text{cov}(R_i - R_f, R_{port}) = E[(R_i - R_f)\, R_{port}] - E[R_i - R_f]E[R_{port}],$$

the first-order condition simplifies to

$$E\left[R_i\right] = R_f + \frac{b\,\text{cov}(R_i, R_{port})}{1 - bE\left[R_{port}\right]}, \tag{13.1}$$

from which we can find equilibrium asset prices P_1 and P_2.

13.1 Numerical example

Asset prices are calculated from Equation (13.1) under four different sets of information, $\Phi_1, \Phi_2, \Phi_3, \Phi_4$. The purpose of these calculations is to exhibit how new information can drive the cost of capital up or down, both for individual firms and for the market as a whole.

Information Ω_1. Under Ω_1, the two assets are perceived as independent and $\Pr(V_1 = 1) = \Pr(V_2 = 1) = 0.6$, giving $P_1 = P_2 = 0.409$. The price-implied expected returns are $E\left[R_1\right] = E[V_1]/P_1 = 0.6/0.409 = 1.47$, $E\left[R_2\right] = E[R_1] = 1.47$. The corresponding expected return on the "market portfolio" containing only the two risky assets is $E[R_M] = E\left[V_1 + V_2\right]/(P_1 + P_2) = 1.47$, or 47%, so the market risk premium is $1.47 - 1.10 = 37\%$.

Note that these prices were calculated directly from the assumed quadratic utility risk premium parameter $b = 1/3$, and are easily shown to be consistent with conventional CAPM. Calculations are as follows:

$$\begin{aligned}
E[V_M] &= E[V_1 + V_2] = 1.2 \\
P_M &= P_1 + P_2 = 0.818 \\
\text{cov}(V_1, V_M) &= \text{var}(V_1) + \text{cov}(V_1, V_2) \\
&= 0.24 + 0 = 0.24
\end{aligned}$$

$$\text{var}(V_M) = \text{var}(V_1) + \text{var}(V_2) + 2\text{cov}(V_1, V_2)$$
$$= 0.24 + 0.24 + 0 = 0.48$$

and hence

$$E[R_1] = \frac{E[V_1]R_f}{E[V_1] - \frac{\text{cov}(V_1,V_M)}{\text{var}(V_M)}(E[V_M] - P_M R_f)} = 1.47.$$

Similarly, $E[R_2] = 1.47$. The two asset prices $P_1 = 0.409$ and $P_2 = 0.409$ are thus as per CAPM.

Information Φ_2. Under information Φ_2, the two assets are identical but dependent. There is an underlying economic condition which can be either G (Good) or B (Bad), where $p(G) = p(B) = 0.5$. The relevant probabilities are the same for both risky assets. Specifically, $p(V_i = 1|G) = 0.85$ and $p(V_i = 1|B) = 0.35$ ($i = 1, 2$). The unconditional probability is thus, by the law of complete probability

$$p(V_i = 1) = p(G)\ p(V_i = 1|G) + p(B)\ p(V_i = 1|B) = 0.6, \quad (i = 1, 2)$$

consistent with Ω_1. The difference between Ω_1 and Ω_2 is that the two risky assets are perceived by the investor, under Ω_2, as dependent with positive covariance.

Information Ω_3. Information Ω_3 is a further refinement of Ω_2. The new conditional probabilities are

$$p(V_1 = 1|G) = 0.9 \qquad p(V_2 = 1|G) = 0.8$$
$$p(V_1 = 1|B) = 0.2 \qquad p(V_2 = 1|B) = 0.5.$$

These "more conditioned" probabilities remain consistent with the less conditioned probabilities under Ω_2, and hence also Ω_1. For a random risky asset j, $p(V_j = 1|G) = 0.85$ and $p(V_j = 1|B) = 0.35$, which equal the conditional probabilities for both assets under Ω_2. Thus, the "average" (i.e. unconditional) probability $p(V = 1) = 0.6$ is unchanged from Ω_1. But there are now two levels of conditioning. The two conditions are (i) whether the economy is G or B, and (ii) whether the asset is of type $j = 1$ or type $j = 2$.

Table 13.1: Asset prices and returns given information Φ.

	Information Set Ω			
	Ω_1	Ω_2	Ω_3	Ω_4
$E[V_1\|\Omega]$	0.6	0.6	0.55	0.55
$E[V_2\|\Omega]$	0.6	0.6	0.65	0.65
$\text{cov}(V_1, V_2\|\Omega)$	0	0.0625	0.0525	−0.0225
$\text{cov}(V_1, V_M\|\Omega)$	0.24	0.325	0.3	0.225
$\text{cov}(V_2, V_M\|\Omega)$	0.24	0.325	0.28	0.205
$P_1\|\Omega$	0.409	0.362	0.320	0.375
$P_2\|\Omega$	0.409	0.362	0.423	0.477
$E[R_1\|\Omega]$	1.467	1.657	1.718	1.467
$E[R_2\|\Omega]$	1.467	1.657	1.537	1.363
$E[R_M\|\Omega]$	1.467	1.657	1.615	1.408

Information Ω_4. Information set Ω_4 comes from a refinement of Ω_1. Information sets Ω_4 and Ω_3 are both feasible refinements of Ω_1. The conditional probabilities under Ω_4 are:

$$p(V_1 = 1|G) = 0.7 \qquad p(V_2 = 1|G) = 0.5$$
$$p(V_1 = 1|B) = 0.4 \qquad p(V_2 = 1|B) = 0.8.$$

The CAPM (quadratic utility with $b = 1/3$) equilibrium prices, and implied results, are listed in Table 13.1.

13.2 Interpretation

The possible effects on firms' risk premia of conditioning beliefs on changes to the information set are unpredictable. When new information — brought perhaps by application of new accounting standards — allows the market to better discriminate between individual firms, some firms can be left with a (much) higher cost of capital.

Contrary to intuition perhaps, the overall market risk premium $E[R_M] - R_f$ can increase with better information. In the example, it is higher under both Φ_2 and Φ_3 than under more rudimentary information

Ω_1. That tells us that a paucity of information will sometimes cause a rational market to undercharge, not overcharge, for risk capital. That is analogous to the insurance company not recognizing how truly risky a customer is, and charging too little for insurance.

Equally, new or better information can sometimes afford firms a lower cost of capital. See how $E[R_M] - R_f$ is lower under Ω_4 than under Ω_1. Interestingly, and again showing how easy it is to overemphasize the cost of capital, existing stock holders in a firm charged a lower required cost of capital will not be pleased if the better quality information brought both a lower cost of capital and simultaneously a lower share price, as would occur of course if that information caused a sufficiently large drop in the firm's perceived mean payoff $E[V_j]$.[2]

The accounting objective for asset pricing is to provide information that assists "correct pricing of risk", rather than necessarily a lower price of risk. The insurance analogy holds. Some firms, like some individuals who have life insurance, would rationally be charged a higher-risk premium if there were better information made available about them. Better information allows better calibration of risks, and thus better calibration of the risk premium charged to individual risks.

The ultimate benefit of information that proves ex post to have led to more accurate beliefs is higher average realized utility, in the way underpinning the Alchain's theory of economic Darwinism. Note however that in practice the accuracy of ex ante beliefs may never become clear, because beliefs are often about intangibles (like whether the firm's management is good and smart). All that is observable is the growth and volatility of the investor's portfolio, which reflects the quality of her beliefs only indirectly.

[2] According to Lambert *et al.* (2007) and also Fama (1977), as refound and explained in Johnstone (2017), the arrival of new information ω will induce a higher cost of capital if Fama's ratio $\text{cov}(V, V_M)/E[V]$ increases.

14

Reliance on the Normal-Normal Model

> In some simple models, the posterior variance will always be less than the prior variance, as in the simple normal model with a normal prior. . . ., but this will not always be true. There will typically be values of x for which $\mathrm{var}(\theta|x) > \mathrm{var}(\theta)$. . . (O'Hagen, 1994, p. 86)

The theoretical accounting literature could be accused of not giving Bayesian theory its full effect. In a great majority of published papers, the variable of interest $\theta \sim N(\mu, \sigma^2)$ is assumed normal with unknown mean μ but with known variance σ^2. The signal $\overline{x}$ (or just x if $n = 1$) is the mean of a random sample $(x_1, x_2, \ldots, x_n)$, and is thus understood as an unbiased noisy measure of μ of *known* fixed variance or precision, $E[(x-\mu)^2] = \sigma^2/n$. Its precision n/σ^2 is known only because σ^2 is assumed to be known. Any sample size $n \geq 1$ leads to a Bayesian normal posterior distribution $f(\mu|\overline{x})$ with lower variance over μ than the prior variance over μ (Winkler, 2003, p. 150).

The simplifying assumption of known variance σ^2 makes for a tractable model at the expense of descriptive validity.[1] More realistic Bayesian models allow σ^2 to be uncertain, like μ. They typically involve dependent rather than independent parameters, and are amenable to only numerical rather than closed-form results, making them out of favor in the analytical research culture.

Models that show that information reduces uncertainty give themselves an inbuilt advantage by assuming away any uncertainty about the payoff variance. Bayesian computations for more realistic models, in which the payoff variance is unknown and is itself the subject of a prior distribution, have in recent times been made practically applicable by advances in the numerical methods used to simulate joint and marginal posterior distributions over combinations of unknown parameters. This advance in numerical methods is a large part of why Bayesian methods have gained so much more takeup in applied statistics disciplines. See Robert and Casella (2004).

14.1 Intuitive counter-example

Christensen and Feltham (2003, p.78) explain that in the normal–normal model, with an unknown mean and known variance, the location of the posterior distribution for the unknown mean is affected by the observed sample mean (or signal) but the variance of that distribution is not affected by what's in that sample, even when the sample observations exhibit much higher or lower variability than was expected:

> The latter is only affected by the covariance and variance characteristics of the signal, not the specific signal. This feature simplifies analyses that are based on normal distributions. (Christensen and Feltham, 2003, p. 78)

The mathematical convenience of this model is its selling point, but its essence is that the information user is prohibited from learning about

[1] Rather than always an unknown mean and known variance, the model could equally assume a known mean and unknown variance, accepting that this is also generally an unrealistic setting.

an unknown variance from what the information “says” about that unknown variance. That weakness is shown by the following intuitive example.

Suppose that the market is interested in the firm’s quarterly net cash flows, for the purposes of valuing the firm by assessing its expected future cash flow and its risk. Consider two possible samples of past quarterly results:

Sample 1 ($n = 4$, $\overline{x} = 40$)	$\{10, 100, 20, 30\}$
Sample 2 ($n = 4$, $\overline{x} = 40$)	$\{35, 40, 45, 40\}$

Letting θ represent the next cash flow or quantity of interest, the simple model used in accounting says $\theta \sim N(\mu, \sigma^2)$ with unknown mean $\mu \sim N(m, s^2)$ but with already known variance σ^2. By this model, Samples 1 and 2, which have the same sample mean $\overline{x} = 40$, and the same sample size $n = 4$, carry exactly the same information.

Both samples have the same information precision (since they both have $n = 4$). Their equal effect is that the revised distribution for μ is shifted right or left by the sample mean $\overline{x} = 40$, and is a weighted average of the prior mean m and the sample mean $\overline{x}$, with $\overline{x}$ weighted effectively by σ^2/n relative to s^2.

There is no inference about σ^2 because it is assumed to be already known, so there is nothing to learn. Sample 1 seems to indicate that monthly profit and hence future profit is high variance or high risk, but that indication of higher σ^2 is ignored. The in-sample variance, which is much higher in Sample 1, plays no role in the calculations. The only aspects of the sample that count are $\overline{x}$ and n. In effect, Sample 1 is interpreted as carrying exactly the same evidence about μ and about the amount and risk of the future cash payoff as Sample 2.

Taking this example further, consider a larger (more precise) sample:

Sample 3 ($n = 8$, $\overline{x} = 40$)	$\{10, 50, 10, 110, 10, 20, 30, 80\}$

On the standard model, Sample 3, with its obviously high sample variation, would lead to a lower-risk assessment of the firm, because the higher sample size and sample precision ($n = 8$) lead to a tighter posterior distribution for μ, albeit with the same mean (since $\overline{x} = 40$ is the same for all three samples). Thus, despite the extreme uncertainty or variation that is so apparent in Sample 3, that data set is interpreted as giving more assurance about the future payoff than Sample 2, which obviously exhibits much lower in-sample variance.

That unrealistic intuition or model is endemic in accounting theory. Even in the recent paper by Dye and Hughes (2018), which shows via the law of total variance how information can add to perceived risk, there is a mathematical thread running through the model by which certainty under set conditions always increases with new information (my earlier reconstruction of the Dye and Hughes model has the same characteristic). This "conventional statistical result" (as Dye and Hughes describe it) is hardly conventional outside accounting-related information theory.[2] The assumption at fault in the usual accounting Bayesian model is its known population variance, σ^2 (or, equivalently, known sampling variance σ^2/n).

A generally more realistic approach would have both cash flow parameters, mean μ and variance σ^2, as unknowns, and would attach prior distributions to them both, or a joint prior, and then work through to numerical results via simulating posterior predictive distributions for the cash payoff.

Numerically derived posterior distributions are generally not favored by accounting theorists, but they are commonplace in Bayesian applied science, in fields like meteorology, where verifiably accurate forecasts are given first priority over theoretical elegance and tractability. Closed-form posterior predictive distributions have become less valuable as numerical Bayesian methods and the required computing power have developed.

14.2 Appeal to the normal–normal model in accounting

The workhorse model, with its normal distributions, known population variance (and hence known sampling variance) along with an presumed unbiased signal, is lent upon in so much theoretical and empirical

[2] In meteorology, there is no convention that suggests that better information will make the possibility of rain today more certain. Better analysis should move that probability but not in a predictable direction.

accounting research that it deserves its own name. It is sometimes referred to colloquially as the "normal–normal" model,[3] but that does not sum up its main advantages and disadvantages. It comes early in the Bayesian textbook for good reason. It is easy to apply and has a most appealing implication, namely, that more data is not only desirable, as seems obviously sensible, but that learning is monotonic in n. It is natural to conclude, therefore, as in Lambert *et al.* (2007) and many others, that any and all new information reduces investor uncertainty.

> Intuitively, disclosure reduces uncertainty, . . . (Smith, 2017)

The notion of information always bringing some amount of resolution fits loosely with our psychological understanding of what makes "information".[4] But formally, in terms of the general laws of probability, that monotonic relationship between more disclosure and more certainty is highly model dependent. It is contradicted by most other Bayesian structures, and also as a general law by the Bayesian law of total variance (as explained already).

Even other elementary Bayesian models show how information can widen the distribution over the unknown parameter. For example, in the beta-binomial model, which is the simple standard Bayesian model when the unknown parameter is a population proportion ρ rather than a mean μ, the Bayesian posterior of that proportion can have a larger variance than the prior distribution (Winkler, 2003, p. 141).

Lambert *et al.* (2007, p. 398) cite the Bayesian theorist DeGroot, who with Blackwell is a familiar Bayesian name in accounting literature,[5] and claim that the simplified formulation of information as an unbiased "noisy" estimator is in their words "consistent with the way that

[3]For example Neururer *et al.* (2016, p. 401). The population is assumed normal and its unknown mean is given a normal prior distribution, hence normal–normal.

[4]In accounting, it is altogether natural to think of the announced earnings revealing aspects of performance and the state of the firm that were unknown. But the Bayesian point is that earnings might often reveal risks and add to questions over future payoffs that were not previously noticed or given much weight.

[5]DeGroot was a revered Bayesian theorist, but he was one of several of his era, and it seems a habit in the literature that he became the one always cited.

information is modelled in virtually all conventional statistical inference problems". They mention as well that:

> It is also consistent with virtually all papers in the noisy rational expectations literature in accounting and finance. (Lambert *et al.*, 2007, p. 398)

The first of these claims seems more fitting of conventional frequentist statistics than Bayesian, because there are standard exceptions in Bayesian inference. The seminal oil wildcatter problem, dating to the introduction of Bayesian decision methods in the Harvard Business School MBA, is an obvious exception, since its error probabilities can be best biased either way (to maximize the expected utility of the geological test). Because Bayesian inference and logic culminates in decision-making, where the user's loss function is often asymmetric, a biased signal is often a good tradeoff even when set against a loss of statistical "precision" or expected information. Demski's general model of the expected utility of information allows for a biased and low precision signal to be optimal.

Fixed sampling precision

A simplified model where the true parameter is θ and the signal is equal to $\theta + \varepsilon$, with noisy ε of fixed distribution $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, is an obvious starting point for modelling, but has become, as Lambert *et al.* (2007) correctly say, an essential repeated part of analytical research.

A model of information as an "estimator", and always unbiased, and of known fixed precision, is inadequate to capture the potentially desirable Bayesian attributes of information. Typically, for example, if the information is a normal sample mean $\overline{x}$, then sampling precision $\sigma/\sqrt{n}$ will change with the population variance σ^2. That is merely saying that the more variance in the population, the harder it is to get an accurate estimate of the population mean.

Apart from discretionary changes in the accountant's n, sampling precision will often change with the unknown parameter being estimated. For example, the innate "precision", and also level of bias, of reported earnings is surely affected by the amount of underlying "true earnings" or the general state of the firm. An analogy is that the variance of error in guessing the age of a baby will be lower than when guessing the age of a 50-year old. Models by which the qualities of information change with the unknown parameter θ require specification of subjective likelihood functions like those plotted earlier in this monograph.

The "precision" or general quality of accounting information is bound to change with the state or circumstances of the firm. Leuz and Wysocki (2016, p. 541) saw that as a fundamental problem for empirical research. They explained that changes in accounting quality are likely to be driven by, and hence correlated with, changes in the firm's underlying economics, the two being empirically "inseparable". More specifically, what appears, for example, as an increase in "information risk" might well reflect an increase in the variance or risk of the underlying cash flow, making it hard to separate the effect of perceived information risk from the effect of perceived economic risk on market outcomes like the cost of capital.

14.3 Unknown variance, increasing after observation

If the variance were treated as the only unknown, it would have a posterior distribution that might lie to the right of the prior distribution, indicating that the data adds to uncertainty. The following intuitive explanation of how a perceived variance (or covariance) can increase is due to Robert Winkler:

> Suppose that you're interested in the payoff that will result from a particular decision. You're uncertain about the variance of that payoff, so you gather more information. That information could reduce your uncertainty about the variance, but possibly will cause your probability distribution about the variance to shift so that the posterior distribution for the variance is not only less spread out but also has higher mean. Thus, you have reduced your uncertainty about the variance, but the higher posterior mean for the variance implies that your uncertainty about the payoff itself is increased.

Winkler goes on to say that its "not a bad thing" that you have become more uncertain, rather "its better that you know":

> ...if the variance is truly larger, for whatever reason, its better that you know before you make any decisions than that you have a false sense of security, thinking that the variance of the payoff is smaller than it really is. (Winkler, private communication, 2018)

Again the message is that it is generally better to hold "more accurate" or better informed probability distributions, even if they display a lower and possibly disappointing level of knowledge or certainty.

14.4 Beyer (2009)

Beyer (2009) notes how rarely models in accounting theory treat the variance of the quantity of interest (often a future cash flow) as being unknown. Her model is described to allow both mean and variance to be unknowns, consistent with reality:

> Unlike most analytical models of earnings management, the [Beyer] model assumes that investors are uncertain about the process generating both the mean and the variance of the firm's (present value of) cash flows. This seems to be the realistic case; in practice, investors are uncertain about many aspects of the distribution of a firm's cash flows. In the [Beyer] model, their knowledge of both the mean and variance of this distribution is affected by the firm's management's earnings forecast and the firm's earnings report. Since investors use the information they receive about the firm's earnings to update their beliefs about the unknown mean and variance of cash flows, the manager has incentives to manipulate his earnings forecast and earnings report so that investors perceive the firm's expected cash flows to be high and the variance of cash flows to be low.
>
> ...All of the model's predictions ... arise as a direct consequence of investors' uncertainty about the variance of the

> firm's cash flows. If one were to change the model so that investors knew the variance of cash flows, the predictions of the model would change significantly.... This highlights the importance of taking into account that the variance of cash flows is unknown. (Beyer, 2009, pp. 1713–1715)

Beyer (2009) makes the vital point that the inference problem confronting investors is one of updating beliefs about both the variance and mean of the payoff distribution. Investors, unlike theory models, do not have the crutch of a known variance, and in reality their problem is much harder than accounting theory supposes. The extra difficulty is not simply an extra step of updating two parameters instead of one, it is theoretically a far less tractable problem.[6]

The methodologically important conclusion in Beyer (2009) is that, as an exercise in accounting theory, relaxing the assumption of a known variance in the normal–normal model changes the answers:

> In practice, corporate earnings disclosures do not state explicitly higher moments such as the precision of a signal or the variance of a distribution. Instead, investors observe the earnings realization or a forecast of that realization and, based on these observations, draw inferences about all relevant aspects of future cash flows. This analysis attempts to capture such an updating process that includes both mean and variance.... The model shows that taking into account that the variance of cash flows is unknown to investors significantly affects the equilibrium properties of management forecasts, earnings reports, and stock prices. (Beyer, 2009, p. 1716)

Although Beyer stresses the realism of assuming an unknown variance, the mathematical intractability of this Bayesian inference problem is not clearly shown. Standard models as set out in Bayesian textbooks (e.g. DeGroot, 1970; Gelman *et al.*, 2004) call for inference about the

[6]It is well known in Bayesian statistics that revising joint distributions over multiple parameters leaves very few closed-form solutions.

joint distribution of the unknown mean μ and unknown precision $1/\sigma^2$ of a normal distribution. These parameters are dependent and have to be modelled jointly, rather than one at a time. Posterior beliefs about μ can be expressed as follows: (i) μ has a known normal distribution that depends on the unknown variance σ^2, and (ii) precision $1/\sigma^2$ has a known marginal distribution, that is gamma and depends on the parameters of the prior distribution for μ. The effect of this mathematical complexity is that the joint posterior distribution of μ and unknown precision $1/\sigma^2$ can only be produced numerically (by simulation or by drawing from their joint posterior distribution) and then exhibited as a 2D density plot or, usually less visually successfully, as a 3D histogram.

This innate Bayesian intractability prevents accounting theory from widely testing Beyer's strong view that a model that allows the variance to be unknown will alter substantive conclusions. Testing would be possible only numerically, thus making conclusions generally dependent on input parameters and conditions.

A major repercussion for the literature is that by opting for the tractability that comes with the assumption of a known variance in the "normal–normal" Bayesian model, accounting theory is led to overstate the ability of information, whatever it "says", to resolve uncertainty.

The Dye and Hughes (2018) paper is realistic because it allows for the certainty-reducing nature of much relevant information. A forthcoming paper by Heinle and Smith (2015) opens up discussion about the methodological problem for analytical accounting research in the fact that no simple closed-form Bayesian solutions arise for the normal distribution with both population mean and variance unknown. They evade this problem by assuming that the firm can at its choice and extra cost reduce the perceived variance of its payoff, which is a strong assumption for the reason that greater effort towards revelation of the "true" variance may well reveal it to be higher than first appreciated (i.e. a closer look at the firm's fundamentals might identify new concerns). They set up their model purposely such that a single ($n = 1$) realization of cash flow contains information about the variance as well as the mean.

Another recent paper by Heinle *et al.* (2018) attempts to allow for the natural revision in perceived uncertainty (up or down) that will arise under new information. But again because of the inherent absence of any suitable Bayesian closed form model, no Bayesian revision of the variance occurs. A possible solution or way of letting sample information alter both the mean and variance of the payoff is to invoke mixture distributions, as demonstrated earlier in this monograph. That would require quadratic utility so as to obtain the same mean–variance CAPM form as for normally distributed payoffs and exponential utility.

Latent benefit to users implied by normal–normal

Models under which information necessarily adds to certainty imply an embedded benefit to the information user that will be milked by a self-interested information provider (e.g. by more self-interested manipulation of the signal). That inbuilt value source is the key to Armstrong *et al.* (2016); see below.

A more general Bayesian setup lets the user foresee that she might end up less certain after viewing the accountant's report, in which case she will place compensatory other demands (e.g. lower cost or greater precision).

By assuming the usual normal–normal model with known variance, strategic decision models in accounting do not incorporate the natural real world demands of information users who know that the receipt of information will sometimes, if not frequently, increase their uncertainty. The prospect that uncertainty might increase upon information arrival changes the ex ante expected value of new information, and generally therefore the equilibrium solution of the model. Typically in business contexts, "bad news" (e.g. the firm incurs new high cost levels) adds to investors' doubts about the firm's future profitability. As emphasized throughout this monograph, financial information does not neatly affect perceptions of just one of the parameters or moments of the firm's uncertain payoff.

14.5 Armstrong *et al.* (2016)

Armstrong *et al.* (2016) is one of a number of theory papers that exploits the "normal–normal" Bayesian model for which a more precise signal always brings more certainty. Their strategic disclosure model shows how a manager is motivated to be extra precise in her reporting when the firm is in a "bad state". The argument goes that by being more precise about a result that is expected to lower market perceptions of the firm's mean payoff, she will be rewarded for that precision, because the market will now be more certain of the state of the firm. The drop in market price caused by the "bad" report will thus be kept to a minimum by the market now associating a lower (co)variance or market risk with that payoff, and hence a lower discount rate.

Part of the attraction of this result is that it is so counter-intuitive. The common intuition is that, when the firm is in the bad state, the manager will do more to obscure the truth than to advertise it.

In the insurance analogy, the Armstrong *et al.* finding is equivalent to saying that a 65-year old man with a weak heart, high blood pressure and a poor family tree for longevity, might actively ensure that these facts are made clear to the insurance company, so as to be rewarded with a relatively lower life insurance premium.

There is an element of plausibility to this in at least one Bayesian sense. The insurance company may well be impressed with having such a forthcoming client, not for the news that is handed over, but for the openness in providing it. For other clients, there may well be a greater uncertainty and hence concern about what is not on the record, and hence some extra penalty for "information risk" or the risk of adverse selection.

The end result, however, is that the Armstrong *et al.* (2016) model stands on a strained Bayesian model under which every new data point brings more certainty, no matter what it "says", favorable or unfavorable, and a most unfavorable report can bring a lower cost of capital. The second aspect of that model is that the firm can choose its reporting precision. The problem from a Bayesian perspective is that the precision of an estimate or sample observation depends not only on the discretionary sample size or investigatory effort (which in principle the firm can choose), but also on the underlying population variance, which is usually unknown. It is a simplification to assume that reporting precision can be set at the firm's will, and can be set at any arbitrary level, regardless of the inherent uncertainty or variance in the underlying parameter or population.

Consider for example the depreciation on a factory. There is often a deep question about not just the physical life of the factory but its economic viability (whether its output will remain saleable, or whether it can keep production cost at a viable low). That inherent or fundamental uncertainty is not easily, it at all, reduced by more accounting effort or "better" accounting practice. The accountant can write down a bigger or smaller depreciation expense, but that number remains somewhat of a guess and cannot be forced to be more precise in any clear statistical sense.

Moreover, as mentioned elsewhere in this monograph, greater effort in "activity-based costing" could expose a level of "true" production costs that make the viability of the product less certain, and hence have the opposite statistical effect on certainty than is supposed to follow from more accounting "precision".

14.5.1 Potential Remedy

There is potential, as in Dye and Hughes (2018), that by introducing mixtures of distributions, each with known variances, that accounting theory can achieve both elegant closed-form results and also allow for the practically reality that information often increases users' uncertainty.

By going beyond the usual normal–normal model and adding a mixture element, Dye and Hughes (2018) were obliged to shift away from mean–variance asset pricing, because the mixture distribution is not normal, however they cleverly applied a market-clearing criterion to arrive at CAPM-like closed-form equilibrium asset prices under a theoretical pairing of payoffs with mixture distributions and exponential utility. The alternative, as suggested elsewhere in this monograph, is (risk-averse) quadratic-utility asset prices, which for accounting theory purposes should often suffice, mainly because the primary requirement for realistic theoretical modelling is that the investor is taken as plausibly risk-averse. The only difference in the CAPM under quadratic utility relative to the usual conjunction of normal distributions and exponential utility is that the variance aversion parameter involves a term representing the amount of existing investor wealth. A further alternative with potential for application in analytical research is to assume a lognormal payoff and a log utility or power utility CAPM.

Exponential utility is widely used because of its mathematical congruence when combined with normal distributions, but as an economic model of investor character it is usually not a necessity, and also has its own defects. See Cheynel (2013) on these (e.g. a billionaire or a pauper will hold the same dollar value of risky assets).

15

Bayesian Subjective Beta

Lambert *et al.* (2007) define the payoffs from a set of correlated assets as random variables with joint distribution $f(V_1, V_2, \dots V_n|\Omega)$, where Ω is the available evidence.

In their widely accepted model, assets are priced by CAPM. The relevant parameters of asset j's random payoff V_j are then the conditional mean $E[V_j|\Omega]$, and conditional covariance $\text{cov}(V_j, V_M|\Omega)$. When new information arrives, the assumption is that both parameters are revised.

In the same CAPM context, Lambert and Verrecchia (2010) see information as assisting the subjective assessment of CAPM beta, with the objective of:

> …a more precise estimate of, say, a forward-looking beta than can be extracted from historical returns data or other information. (Lambert and Verrecchia, 2010, p. 5)

Note once more the narrowness of the term "precision" used to summarize the quality of the beta estimate, ignoring the possibility and cost to decision makers of a precise but biased estimate.

While Lambert *et al.* (2007) insist in principle on Bayesian belief revision, a model that accepts that the mean and covariance[1] of random payoff V_j are both unknown, for all assets j in the reference set, leads to a predictive distribution for $(V_1, V_2, \ldots V_n)$ that is not joint normal, and requires Monte Carlo simulation. More promisingly, the Lambert *et al.* (2007) model of Bayesians doing CAPM asset pricing holds under quadratic utility, because the predictive distribution $f(V_1, V_2, \ldots V_n|\Omega)$ does not need to be joint normal to justify CAPM under quadratic utility.

Lambert *et al.* (2007) is essentially concerned with how the cost of capital reacts to new Bayesian beliefs when investors are risk-averse. An assumption of quadratic utility is sufficient for this analysis. An attraction of quadratic utility for Bayesian asset pricing theory is that the posterior predictive distribution of asset payoffs can take any form, and the relevant payoff risk characteristic is still its returns "beta", defined as usual as

$$\begin{aligned}\beta_j|\Omega &= \frac{\text{cov}(r_j, r_M|\Omega)}{\text{var}(r_M|\Omega)} \\ &= \frac{\text{cov}(V_j, V_M|\Omega)}{\text{var}(V_M|\Omega)}\ \frac{P_M}{P_j},\end{aligned} \tag{15.1}$$

where by definition: (i) the return on any asset i relative to its price P_i is $r_i = V_i/P_i - 1$, and (ii) the market return r_M is the price-weighted average of all assets' returns, which is equal to $V_M/P_M - 1$, where $V_M = \sum_i V_i$ and $P_M = \sum_i P_i$. See Lambert *et al.* (2007) for further explanation of these terms.

The standard mean–variance price equations adopted by Lambert *et al.* (2007) for asset j, and for the whole market, are

$$P_j = \frac{E[V_j|\Omega] - c\ \text{cov}(V_j, V_M|\Omega)}{R_f}, \qquad (R_f \equiv 1 + r_f)$$

$$P_M = \frac{E[V_M|\Omega] - c\ \text{var}(V_M|\Omega)}{R_f},$$

[1] i.e. covariance with the market aggregate of all assets.

where c is the market's aversion to payoff variance, and which is merely a different amount under quadratic utility compared to CARA utility.[2]

Substituting for P_j and P_M in Equation (15.1) for beta gives

$$\beta_j|\Omega = \left(\frac{\digamma_j}{\digamma_M}\right)\left(\frac{1 - c\,\digamma_M}{1 - c\,\digamma_j}\right) \tag{15.2}$$

where we define

$$\digamma_j \equiv \digamma_j|\Omega = \frac{\text{cov}(V_j, V_M|\Omega)}{E[V_j|\Omega]}, \tag{15.3}$$

implying $\digamma_M = \frac{\text{var}(V_M|\Omega)}{E[V_M|\Omega]}$.

This expression (15.3) shows how beta responds to new information. The arrival of better information can clearly shift the predictive mean or predictive covariance up or down, and there is no Bayesian model in which these two parameters will naturally remain in the same ratio to one another under every information increment. "Bayesian betas" should naturally therefore go both up and down with new information arrivals.

The only constant is the market price-weighted average beta, which stays equal to one, by construction. That is, letting asset j in Equation (15.2) be the whole market

$$\beta_M|\Omega = \left(\frac{\digamma_M}{\digamma_M}\right)\left(\frac{1 - c\,\digamma_M}{1 - c\,\digamma_M}\right) = 1.$$

A thorough exposition of points of CAPM logic related directly to this chapter is contained in PhD teaching notes made generally available by Jeremy Bertomeu at: *https://drive.google.com/file/d/0B3RTY2KvMiIxTWVFUk1aRG5SU2s/view*.

Bayesian assessment of the covariance

The idea that more information shrinks variances and hence reduces the cost of capital has proved hard to resist in accounting theory and empirical research. So much empirical research stems from that a priori

[2]The only difference is that under quadratic utility, unlike exponential utility, the variance aversion parameter c is affected by the investors' initial wealth (see the quadratic utility equation in Johnstone (2015).

hypothesis. Some empirical studies are believed to support it as an observable correlation, but throughout the literature the evidence is described repeatedly as "mixed".

One way to realistically recast the effect of information on beliefs and the cost of capital is to focus not on certainty (variance) but on covariance or correlation. What effect should better accounting information about firms have on investors' perceptions of how strongly two firms' or two industries' cash flows move together? Clearly, new information might suggest that the underlying commonalities and correlation between two cash flows is higher than was previously believed, or that it is lower. Unlike variance, there is little prior reason to think that the assessed correlation coefficient between two random variables should always decrease.

If there are underlying factors or deeply embedded direct relationships between two firms or industries that put them in the same basket in terms of how they are affected by say wage costs, technology changes, interest rates, consumer trends, taxes or some other input or output variable, then better accounting information should usually help to reveal that source of covariance.

In Bayesian terms, meaning in terms of the laws of probability, the covariance between the two cash flows, V_j and V_k, is

$$\text{cov}(V_j, V_k|\Omega) = \rho_\Omega \sqrt{\text{var}(V_j|\Omega)\ \text{var}(V_k|\Omega)},$$

so even if the variance of one or both cash flows is lower when conditioned on new information, a higher correlation coefficient ρ_Ω between them can lead to a higher covariance, and hence higher cost of capital. This explanation calls for risk to be understood as ultimately payoff covariance rather than merely payoff variance.

Covariance between bivariate normal variables

Consider the following practical case of better information about "fundamentals" bringing an increase in a perceived covariance. Suppose that X and Y are two independent normal variables. It follows that the following two variables

$$A = aX + bY \quad \text{and} \quad B = cX + dY$$

are bivariate normal. The covariance between A and B is found as

$$E[AB] - E[A]E[B] = acE[X^2] + bdE[Y^2] + (ad + bc)E[XY] \\ - (a + c)E[X] - (b + d)E[Y],$$

from which it is easy to see circumstances under which changes in any of the fundamentals a, b, c, d can bring a higher covariance between A and B. The fundamentals a, b, c, d might be things like market share or contribution margins per unit, and can each increase or decrease, thus bringing either increases or decreases in the covariance between two firms' payoffs, A and B. For example, suppose that X is the dollar sales value of the entire market and a and c are the market shares of the two firms. If a and c were both perceived to have increased, there being more than two firms in the market, the covariance between A and B would increase, cet.par.

15.1 Core *et al.* (2015)

There is disagreement surrounding the analysis by Lambert *et al.* (2007), which used the normal–normal model with known variance to find that better accounting information will shrink subjective covariances and hence the cost of capital. Core *et al.* (2015) summarized Lambert *et al.* as follows:

> The information effect [in Lambert *et al.*, 2007] occurs because disclosure quality reduces parameter uncertainty regarding the estimate of expected returns (e.g. Barry & Brown, 1984, 1985; Brown, 1979). Specifically, better disclosure improves investors' prediction of cash flows. Since more of the realization of future cash flows is known, the covariance between the firm's cash flows and the cash flows of stocks in the market portfolio becomes lower, which in turn reduces firm beta and the cost of capital. This effect is not diversifiable because it is present for all covariance terms, and hence lowers systematic risk. ... We note that this prediction from Lambert *et al.* (2007) is not without controversy. For example, Johnstone (2015) shows that if

> information also changes assessments about the mean of firm value [payoff], the cost of capital can *increase* when information precision increases. (Core *et al.*, 2015, p. 4)

I will take the points in this summary one at a time and comment from a Bayesian statistical viewpoint:

Information and parameter uncertainty

We cannot rely on better information making the "true" underlying covariance between two random variables lower. Covariances or correlations between firms are often high and must often increase when firm or market fundamentals change. Better information should at least sometimes reveal that happening. Better accounting might give a truer vision of past and future sales dollars, or costs, with the effect of exposing how closely the firm's financial performance correlates to market downturns or upswings.

Better disclosure improves cash flow prediction

In an economic decision framework, where predictions or probability forecasts are "better" if they turn out to be more accurate, it is false to assume that greater certainty or lower covariances ex ante will turn out to be a better prediction (i.e. a more accurate probability distribution).

Knowing part of a previously unknown cash flow leaves the firm's long-term future "no closer" and possibly still more unpredictable than previously. If the realized cash flow was lower than expected, there may well be a new world of doubt about the future cash flow.

Information risk not diversifiable

This point made by Lambert *et al.* (2007) against some opposition in the literature, is correct, because any recognition of additional or different "types" of uncertainty will change investors' perceived covariances and hence change the cost of capital. No well-diversified portfolio can escape from its market required return being affected by perceived information quality. Again, however, it is not possible to generalize about how different uncertainties melt into a single predictive distribution for the future payoff. Hence, the principle of information risk being undiversifiable says merely that the cost of capital will be affected by assessments of information quality, but not by how much *or in which direction.*

From a Bayesian perspective, a "risk" factor or consideration, here labelled R, is "diversifiable" if and only if the predictive distribution of the market payoff V_M conditioned on that risk

$$f(V_M|..,R,..) \propto f(V_M|\cdot)f(R|..,V_M,..) \quad (V_M = \sum V_j)$$

is such that the risk premium on the market portfolio is unchanged.

In principle, it is unlikely that any material or relevant information about asset payoffs will leave their joint probability distribution such that there is no change in the composition of the market portfolio and its expected return. If information risk has any effect on the rational market portfolio weights, it will virtually always occur that the required return on the revised market portfolio will have changed (either up or down).

Effect of information on mean payoff

Better information can clearly shift the perceived mean payoff of the firm up or down, indeed that is an ex ante desired effect of better information. Note how from Equations (15.2) and (15.3) information about the mean payoff affects the subjective beta of that payoff. This is a little known CAPM fact explained at length in Johnstone (2017).

Lambert *et al.* (2007) and Johnstone (2015), Johnstone (2016), and Johnstone (2017) stressed the effect of a revised mean payoff on ex ante beta and the cost of capital, showing that a lower assessed mean brings a higher CAPM cost of capital. That part of Lambert *et al.* (2007), which is traced by Johnstone (2017) to early work by Fama (1977), goes widely unmentioned in both accounting and finance, which is surprising given its influence and citation count. The relatively very few exceptions include Christensen *et al.* (2010), Gao (2010), Core *et al.* (2015), Amiram *et al.* (2018) and Larson and Resutek (2017).[3]

[3](Gao, 2010, p. 20) note that Lambert *et al.* (2007) showed the formal CAPM effect of the current (post-information arrival) mean payoff on the forward-looking cost of capital, however "they do not link this result directly to disclosure quality".

It is likely that the effects of information and its perceived qualities on assessments of mean future payoffs are the most critical input in asset pricing and portfolio management under a risk-averse utility function. There is much finance literature (e.g. Best and Grauer, 1991; Chopra and Ziemba, 1993) suggesting that portfolio outcomes are highly sensitive to errors in the assessments of mean returns, much more so than to errors in the assessed covariance matrix.

The statistical literature in Bayesian decision analysis has not explored the separate effects of the payoff mean and payoff variance on the certainty equivalent of a future payoff. That is for several reasons, but mainly because no common expected utility function apart from quadratic utility can be written in terms of just mean and variance. It is easy to show, however, for a broad class of risk-averse utility functions, that the minimum required rate of return implied by an investor's certainty equivalent, is decreasing in the mean payoff. That is the remarkably little considered result that was re-discovered under CAPM by Lambert *et al.* (2007), traceable to Fama (1977) and also clear in Hull (1986).

15.2 Verrecchia (2001): Understated influence of the mean

Accounting literature has over-emphasized "second-moment" effects. In particular, the cost of capital is treated as if it is driven by a combination of the assessed ex ante innate payoff variance or covariance and investors' perceptions of information asymmetry. Both are considered second-moment effects. More emphasis is given to investors' uncertainty about the accuracy of their estimated mean than to the estimated mean itself, thus fixating on the variance of the estimate rather than on the first moment effect of a change in the assessed mean.

That omission and many of the critical points raised in this monograph are raised in the caveat at the end of Verrecchia's (2001) "Essays on Disclosure":

> Information asymmetry, like many of the economic consequences posited in these essays, is a "second moment" effect (i.e. a variance effect), and second moment effects may

> be secondary or tertiary in nature when compared against "first moment" effects (i.e. mean effects). For example, one would expect to be able to document that, as a "first moment" effect, "good news" drives prices up and "bad news" drives prices down. Theory-based models, however, commonly characterize information asymmetry as a second-moment effect that is unrelated to means of first moments. (Verrecchia, 2001, p. 174)

The literature's focus on second-moments, either as variances or covariances, goes against what is the far more obvious influence on prices and discount rates of the mean.[4] Verrecchia (2001) goes on to say that accounting has been able to dwell on second-moments because in normal distributions the mean and variance are independent of one another. Uncertainty about the mean can thus be separated from the amount of the mean[5]:

> Information asymmetry is commonly characterized this way because variables are posited to have a normal distribution, which implies two independent moments; obviously, for other (i.e. nonnormal) distributional forms, there may be higher moments and all moments may be related. The problem with second-moment effects is that they are too subtle or obscure to manifest themselves in measurable ways. (Verrecchia, 2001, p. 174)

The last part of this quote is referring to the empirical difficulty of identifying the effects of information asymmetry when the observed price

[4] It was explained above that in a payoffs model, the ex ante mean drives the asset price and its cost of capital. The underlying effect of the mean comes out in the finance portfolio optimization literature where optimal portfolio weights are generally far more sensitive to changes in estimated mean returns than estimated returns covariances (e.g. Best and Grauer, 1991; Chopra and Ziemba, 1993).

[5] Christensen and Feltham (2003, p. 78) state that in this model the investor's observed sample mean affects his posterior mean but not his posterior variance. They say that "this feature simplifies analyses that are based on normal distributions" (p. 78). Verrecchia's point is that also can mislead researchers to overlook the role of the observed mean.

and returns effects are more largely due to changes in first moments (e.g. good versus bad news about the amount of sales or costs). This is another way to make my earlier point that the market is influenced by what the information "says" directly as well as by how "precisely" or reliably it says it.

Basic insight for accounting research

The following illustration offers a basic insight for all accounting researchers. It shows, in the simplest possible asset pricing application, the primary effect of what the information "says", or specifically in whether the news about the payoff mean is directly "good" or "bad". Very similarly, see Veronesi (1999, pp. 977–978).

The random payoff is $V \in \{0, 1\}$ with a Bernoulli distribution with index $\pi = p(V = 1)$. The payoff mean is $E[V] = \pi$ and its variance is $\text{var}(V) = \pi(1 - \pi)$. Let the asset's ex ante market price V be

$$P = E[V] - c\,\text{var}(V), \tag{15.4}$$

assuming a zero risk-free rate. The price-implied cost of capital is $E[r] = E[V]/P - 1$.

The plot in Figure 15.1 shows curves for the cost of capital under good news and bad news. Good news is defined as news that makes $\pi > 0.5$ and bad news is news that makes $\pi < 0.5$. The cost of capital is plotted as a function of $\text{var}(V) = \pi(1 - \pi)$. The left-hand point where the two curves coincide occurs at $\pi = 0.5$ where the variance is 0.25. Going to the right in both the curves corresponds to stronger news. For good news, $\pi \to 1$ and for bad news $\pi \to 0$, both sending the variance towards zero.

This plot shows how the required return on capital is greatly affected by news about the mean π. With good news (π rising from 0.5 towards one), lower variance is associated with a lower cost of capital (be careful to note that on the horizontal axis variance is reducing towards the right, as the news becomes stronger).

That positive relationship between variance and the cost of capital is as usually expected, but gives a false impression. It is not the lower variance that is driving the lower cost of capital, it is as much or more

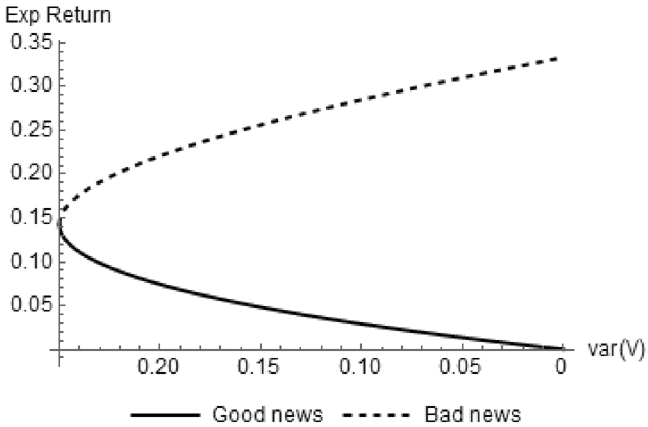


Figure 15.1: Plot of expected return as a function of $\mathrm{var}[V] = \pi(1-\pi)$ (with $c = 0.25$).

the higher mean! We know this from the plot for bad news. In its case, lower variance brought by π heading away from 0.5 towards zero brings a *higher* cost of capital, not a lower cost of capital as would be presumed. The limit of the cost of capital as news worsens is

$$\begin{aligned}
&\lim_{\pi \to 0} \left(\frac{E[V]}{P} - 1 \right) \\
&= \lim_{\pi \to 0} \left(\frac{\pi}{E[V] - c\,\mathrm{var}(V)} - 1 \right) \\
&= \lim_{\pi \to 0} \left(\frac{\pi}{\pi - c\pi(1-\pi)} - 1 \right) \\
&= \frac{c}{1-c} = 33\%.
\end{aligned}$$

The only parameter that we can rely on for having a monotonic relationship with the cost of capital is the mean. Specifically, and I suggest generally throughout investment under uncertainty, the cost of capital is decreasing in the expected or mean payoff, and is generally, or at least often, more affected by the payoff mean than by the variance or risk. This may be contrary to convention, but is easily demonstrated.

For that reason, the caveat in Verrecchia (2001) on the relative insignificance of second-moments is very well justified. In general, Verrecchia is saying that empirical studies of the effect of second-moment or risk effects are likely to be futile whenever first moment effects dominate.

Note that the equivalent of my Fama ratio (15.3) for the pricing equation (15.4) is

$$\frac{\text{var}(V)}{E[V]},$$

because from Equation (15.4)

$$E[R] = \frac{E[V]}{P} = \left[1 - c\frac{\text{var}(V)}{E[V]}\right]^{-1}.$$

Again therefore the cost of capital must be understood as caused by *risk per unit of mean*, rather than merely by "risk".

For the Bernoulli distribution, the risk per unit of mean is

$$\frac{\text{var}(V)}{E[V]} = \frac{\pi(1-\pi)}{\pi} = (1-\pi),$$

which increases whenever bad news brings lower π. Bad news will therefore, for this distribution, *always* bring a higher cost of capital.

Note that a fundamental implication for empirical research is that it is not justified in theory to posit a universally positive association between risk and the cost of capital. To test for the effect of risk per se, it is necessary to control for the effect of the mean, observing that virtually any relevant information will alter market assessments of both at once.

As a rough rule of thumb, stronger (i.e. more convincing) negative news should commonly bring a higher cost of capital, and stronger positive news should very generally result in a lower cost of capital. The ideal but likely illusive metric driving the cost of capital in a mean–variance asset pricing framework is the Fama ratio, or ratio of variance/covariance to mean.

A highly contrary point for empirical research is that when accounting information is seen to be better and the cost of capital lower, that effect is prone to have been driven not by the higher information quality or "precision", but by the information prompting a higher ex ante mean payoff in the market's Bayesian probability revisions.

Leuz and Wysocki (2016)

By allowing for the effect of the mean, the curves in Figure 15.1 show the cost of capital as a function of uncertainty reduction (information gets better and resolution greater going to the right). Note how the required return can be either increasing or decreasing with better resolution, depending on whether resolution is good or bad news. It would follow therefore that in general there will be no monotonic or unambiguous relationship between information quality and the cost of capital.

Leuz and Wysocki (2016) provide a detailed synthesis of the empirical literature and concede that results are "mixed":

> In sum, the evidence on the relation between disclosure and reporting and the cost of capital is fairly mixed and still evolving. The empirical results appear to be sensitive to a number of factors, including the cost-of-capital measures (i.e. realized returns vs. ex ante cost of capital), firm size, and the types of disclosures or earnings attributes... (Leuz and Wysocki, 2016, p. 550)

Compounding the difficulty of isolating the effect of information quality on the cost of capital is the Bayesian issue that "information uncertainty" or "information risk" is merely part of investors' overall uncertainty about the outcomes of the firm's economic activities. Not being able to separate these uncertainties makes the problem of separating their individual effects harder still. Leuz and Wysocki (2016) describe the various proxies that have been used as measures of disclosure quality and conclude that the "separation problem" has plagued the empirical literature:

> ...all commonly used proxies for disclosure and reporting are likely to comingle the firm's underlying economics and the reporting (quality) constructs that they are trying to measure. (Leuz and Wysocki, 2016, p. 541)

> ...it is possible that these studies do not illustrate the effects of accruals quality or earnings smoothing, but instead reflect

> differences in firms' operating and economic risks. (Leuz and Wysocki, 2016, p.548)

These concessions hark back to the Bayesian tenet that "information uncertainty" is inseparable from overall uncertainty. When de Finetti laid the foundations of Bayesian statistics, he said that objective probability "does not exist". By that he meant that an uncertain event does not have some objective or physical uncertainty that can be separated from our uncertainty about our measurement instrument (e.g. do we have enough sufficiently representative data, is our model correct, are we seeing what we think, . . . ?).[6]

Matters which seem to fall under the heading of exogenous or physical uncertainties are likely in our subjective understanding to be part of what affects the measurement process and its randomness. For example, if the firm perceives a "physical" change in its economics, it might decide on a different reporting mode (e.g. earnings manipulation) that reacts to that perception, thus letting the underlying "reality" affect the unobservable quality of its representation:

> Making matters worse, managers may endogenously respond to performance shocks by manipulating disclosures and reporting numbers, which creates the additional issue of distinguishing between the properties of manipulated and "neutral" or "normal" earnings. (Leuz and Wysocki, 2016, p. 541)

Part of the failure of empirical research to identify how users of accounting disclosure benefit from or respond to "better quality" accounting must surely be that empirical studies have not controlled directly for whether the good/bad quality information being reported is "good"/"bad" in terms of "what it says" (i.e. whether the news is favorable or unfavorable). That experimental control would categorize disclosures two-by-two as both "good news"/"bad news" and good/bad quality. By not controlling for the content (i.e. meaning) as well as

[6] See Subramanyam (1996, p. 208) on why investors are always uncertain about the "precision" of the earnings number they receive.

quality (i.e. veracity) of news, empirical research omits the potentially very dominant "first moment effect" that is correctly raised by Verrecchia (2001), albeit somewhat late in the piece given as Verrecchia says that models have for decades dwelt almost entirely on second-moments.

Omission of the mean in the literature

Although Lambert *et al.* (2007) and several others have explained how information affects the cost of capital through revisions of the mean payoff, that effect is very widely unrecognized. In his authoritative literature survey, Kothari (2001) focusses on the role of accounting information in fundamental analysis and forecasting of future cash flows. Like Lambert *et al.* (2007) his model assumes asset pricing based on ex ante assessment of the subjective probability distributions of future cash flows. Rather than looking empirically for an explanation of expected returns in factors that occur to be correlated with empirical returns, Kothari (2001) goes back to thinking about the fundamentals of cash flows (understood as lotteries) and their probability distributions. He seems to depart however from Lambert *et al.* (2007) by explaining beta or systematic risk in terms of solely the payoff's covariance, with no allowance for its mean:

> Risk here refers to the systematic (or non-diversifiable or the covariance) component of the equity cash flows' volatility. Single- or multi-beta versions of the CAPM imply that the equity discount rate increases in the equity cash flows' systematic risk. (Kothari, 2001, pp. 124–125)

This passage represents the everyday understanding of what CAPM tells researchers about the parameter of the cash flow distribution that drives the discount rate. It omits to mention the first moment or mean cash flow, and hence does not present the same CAPM logic as Lambert *et al.* (2007). There are a great many similarly incomplete interpretations in the literature (e.g. Strobl, 2013, p. 465). Indeed, interpretations centred strictly on variance and never on mean are so entrenched in textbooks that the Lambert *et al.* (2007) interpretation would be seen as "wrong" by most finance students.

A subtle point of principle is that if Kothari's description is about the cash dividends and capital growth flowing to the stockholder, and these are understood as *returns* (by being divided by opening equity value), then the quote is quite correct. Under CAPM equilibrium prices, equity *returns* are a function of the conventional returns beta (covariance) only. That is, the only driver of returns is the returns covariance or risk. It is only once we go to the more fundamental cash payoff explanation of CAPM equilibrium returns that we see that equity returns are driven by both the mean and covariance of the dollar payoff of the firm's business activity.

This is not an easy distinction to teach, which may be a large part of why it is so widely overlooked, and was played down somewhat in Lambert *et al.* (2007). It is based on the CAPM equilibrium condition, under which two results hold together:

(i) a risky asset is priced as a function of its payoff mean and payoff covariance; and

(ii) the price of a risky asset is such that its price-implied cost of capital can be written equivalently as either a function of its *returns covariance* (standardized as beta) or as a function of its *payoff mean and covariance*.

This point of CAPM principle is emphasized because accounting is concerned with how information affects asset prices and returns, thus begging the "finance" question of which of an asset's parameters are involved fundamentally in determining those economic outcomes.

15.3 Decision analysis effect of the mean

Bayesian decision analysis subsumes mean–variance analysis (Meyer, 1987; Johnstone and Lindley, 2013) and is generally skeptical of finance models that treat uncertain payoffs as if they can be fully summarized by just their mean and variance. It is useful therefore to take a general decision analysis view of whether the mean payoff should affect the investor's required return when buying that random future payoff.

Does a mere shift to the right or left in the payoff distribution, with no change in its shape (e.g. variance) affect the investor's rational required return? A simple but general proof shows how the economically "natural" required rate of return decreases with a higher mean payoff.

An asset pays $V = 1$ or $V = 0$. The investor has a piecewise linear utility function $U(W)$ with slope m_1 below initial wealth W_0 and slope m_2 above W_0 (under risk aversion $m_1 > m_2 > 0$). The investor's indifference price P, given probability assessment $p = p(V = 1)$, is found therefore by solving

$$U_0 = p(U_0 + m_2(1 - P)) + (1 - p)(U_0 - m_1 P),$$

where $U_0 = U(W_0)$. Hence

$$P = \frac{pm_2}{m_1(1 - p) + pm_2},$$

implying a minimum required expected return equal to

$$E[R] = \frac{E[V]}{P} = \frac{(1 - p)m_1 + pm_2}{m_2}. \tag{15.5}$$

Hence,

$$\frac{dE[R]}{dp} = \frac{m_2 - m_1}{m_1},$$

which is negative for all risk-averse bettors (who by definition have $m_1 > m_2$). Thus, a risk-averse bettor will require or willingly accept a lower expected return ex ante whenever she has a higher probability p and thus higher mean.

To be clearer, we must allow for the simultaneous effect of that higher p on $\text{var}(V)$. If we write the required expected return in terms of the ratio of the payoff variance to the payoff mean (here called $\digamma$)

$$\digamma \equiv \frac{\text{var}(V)}{E[V]} = \frac{p(1 - p)^2 + (1 - p)p^2}{p} = (1 - p),$$

Equation (15.5) becomes

$$E[R] = 1 + \digamma \left(\frac{m_1}{m_2} - 1 \right),$$

proving how the investor's indifference or minimum required rate of return is set by her assessment of the ratio of the payoff variance per unit of mean. Also,

$$\frac{dE[R]}{d\digamma} = (m_2 - m_1)/m_2,$$

revealing how a more risk-averse investor reacts more strongly to ratio F.

This simple yet general analysis is reassuring in its consistency with the Lambert *et al.* (2007) demonstration of how a higher mean payoff must be discounted at a lower rate, all else equal. That general finding is implicit in decision analysis under any risk-averse utility, and is not unique to CAPM.

As an important economic insight, the effect of the mean and the F ratio is possibly little known because decision analysis, unlike finance and accounting, does not focus on price-implied rates of return (implicit discount rates or "costs of capital") and is also not generally focussed on only the first two moments of the payoff distribution (rather than the whole distribution).

16

Other Bayesian Points of Interest

Discussions in the accounting literature sometimes vacillate between Bayesian and frequentist positions. In this chapter I make comment from a Bayesian perspective on points that arise commonly in accounting theory and in motivating and interpreting empirical accounting research.

16.1 Accounting input in prediction models

A good way to grasp the information objective of inducing accurate probability formation, while still thinking conventionally about accounting information, is to imagine accounting disclosures as inputs into a conventional bankruptcy prediction model, like Ohlson (1980). Any accounting report that assists the model to attach higher probabilities of failure to ex post bankrupt firms, and lower probabilities to non-bankrupt firms, is economically advantageous. Those desirable shifts in probability assessments will of course sometimes imply that the resulting probabilities are closer to 0.5 (maximum uncertainty) rather than to 0 or 1.

Note that common forms of bankruptcy prediction model are based on frequentist versions of logistic regression, like the Ohlson (1980) model. A Bayesian user's perspective is that the output from such models is still "information", possibly very informative information. The user might generally prefer that the modelling had been Bayesian from the start, but if a non-Bayes model or heuristic attached a probability estimate $\widehat{p}$ to bankruptcy, the Bayesian would work with that "signal" by finding her subjective probability of bankruptcy, $p(Bankrupt|\widehat{p})$. The number $\widehat{p}$ is just a piece of information, as per Demski. Note that the frequentist properties of the logistic regression would come into the Bayesian calculation, because the Bayesian would need to know enough about the origins of $\widehat{p}$ to form a likelihood function, including for example how often the model attaches high probabilities to non-Bankrupt firms, and vice versa.

16.2 Earnings quality and accurate probability assessments

Notions of "accrual quality" or "earnings quality" fit easily with an underlying objective of improving the accuracy of user's probability assessments. High-quality "accrual accounting" calls in different circumstances for either an increase or decrease in the firm's income or net assets, and is not bound to heighten investor certainty or uncertainty. Earnings "quality" is ultimately a notion of "accuracy" or "truth" or even "bankability" (Francis *et al.*, 2006; Dechow and Schrand, 2004), to be compared against related outcomes like future earnings and future cash flows and stock prices.

> Our focus is on earnings quality from the perspective of the analyst. ... From this perspective, a high quality earnings number is one that accurately reflects the company's current operating performance, is a good indicator of future operating performance, and is a useful summary measure for assessing firm value. (Dechow and Schrand, 2004, p. 5)

The identification of accruals "red flags" in Dichev *et al.* (2013) is aimed at improving accounting and financial statement analysis in ways that will often intentionally add to investor uncertainty about a firm's future cash flows or stock price. That occasional increase in market uncertainty is a natural by-product of "better" or more accurate information. Accounting, doing its job, will sometimes have that effect.

16.3 Expected variance as a measure of information

Fischer and Stocken (2001), and also Beyer *et al.* (2013), suggest that the "expected precision" from a signal S can be treated as an ex ante measure of its informativeness. The signal S is designed to be informative about value or parameter θ, and its informativeness is assessed by $E[\text{var}(\theta|S)]$.

This approach has an interesting interpretation following from the law of total variance,

$$\text{var}\,(\theta) = E[\text{var}(\theta|S)] + \text{var}(E[\theta|S]).$$

The implications of this law are:

(i) we always "expect" to gain greater certainty from signal S about unknown θ, in the sense that on average S brings a lower variance

(ii) the expected reduction in the perceived variance of θ is

$$\text{var}\,(\theta) - E[\text{var}(\theta|S)] = \text{var}(E[\theta|S]).$$

So, interestingly, the expected reduction in variance equals the variance of the expected mean, which tells us that a highly informative signal is one that is expected to bring a large shift in the perceived mean, up or down. For example, suppose that S has two possible values s_1 and s_2. The "experiment" is highly informative if $\text{var}(E[\theta|S])$ is large, which occurs when the conditional means, $E[\theta|s_1]$ and $E[\theta|s_2]$, are wide apart. So a highly informative signal is one that can bring either a "big shift left" in the mean or a "big shift right" in the mean.

Note that this analysis applies only in contexts like those raised in the two papers cited, where the posterior distribution is normal, and hence the relevant notion of "uncertainty" is captured by purely variance. In other distributional forms, like a beta distribution, for example, other measures of uncertainty or information, like entropy, are required.

16.4 Information stays relevant

Today's Bayesian posterior is tomorrow's prior, so, for example, an updated forecast or earnings report does not make the earlier statement irrelevant. We might learn that the earlier signal was wrong or was "misstated" or biased, but that itself is new information and today's probability reacts accordingly. Barth *et al.* (2001) make the following suggestion, which is questionable in Bayesian principle:

> ...accounting can be value relevant but not decision relevant if it is superseded by more timely information. (Barth *et al.*, 2001, p. 80)

Today's inference will be a probability distribution $f(V|x_1, x_2)$ for the future share price based on an interaction between the earlier signal x_1 and the new one x_2. As found in Simpson's paradox and similar Bayesian examples, the interaction between two or more information items can be highly synergistic. For example, x_2 might suggest implicitly that x_1 was fraudulent or fudged, raising new doubts when the two bits of information occur together, despite neither signal of itself meaning much. Any correction to an earlier signal can raise doubts over the credibility of even the latest signal. Similarly, even if both signals are credible and precise, a reversal or any real change is often informative about some underlying process variance or volatility, or inbuilt measurement variation.

The way that different individual signals can combine to reverse or exacerbate the effect of one another appears in calculations of the value of information. Howard (1966) showed how the value of two items of information can exceed the sum of their individual values. Similarly, Samson *et al.* (1989) revealed how the values of different signals are not additive.

16.5 Bayesian view of earnings management

Any form of accrual accounting involves by definition earnings "management" or "manipulation", since cash flows are discretionally "allocated" or "matched" into period-by-period accounting "profit" numbers. Earnings management and earnings quality are sometimes seen as incompatible,[1] however the Bayesian view is that signals that come not

[1] Lo (2008) describes "earnings management" as an accounting research euphemism for alterations by managers, possibly aided by their accountants, meant to mislead or reach self-interested contractual benchmarks.

from nature alone, but via human processes, have the advantage that they can be tailored to be inherently statistically "better" signals. After all, optimal signal design with respect to objectives is the subject of the strategic disclosure literature in accounting.

Accountants and auditors, among all those involved with the firm and its financial and operating circumstances, are well placed to understand and invoke notions of "true" or "permanent" income, or to send out information that is: (i) factually based on the firm's fundamentals, (ii) relevant to users, and (iii) open about of its own weaknesses, and timely. Earnings "manipulation" can thus potentially add to the Bayesian value of an earnings report, which is how accounting practitioners traditionally imagine their own role and expertise (Dechow, 1994, pp. 4–5).

Barth and Taylor (2010) described accounting discretion similarly, as not necessarily undesirable:

> Wherever there is discretion in accounting — which is essentially everywhere — there is the opportunity for earnings management. The relation between managerial discretion and investor welfare is ambiguous and likely varies by setting. (Barth and Taylor, 2010, p. 32)

Similarly, all the empirical studies in accounting of stock price reaction to earnings announcement are predicated on accounting having incremental information content. That is essentially also the Demski–Bayesian view, despite Demski dismissing in principle the traditional "normative school". His effective position is that accounting rules or standards can be first designed, and then expertly put into effect, so as to make the reported information broadly useful to its Bayesian users, but unfortunately not better for every user, or in the same degree.

16.6 Numerator versus denominator news

A distinction is made in accounting research (e.g. Hodder *et al.*, 2014; Botosan and Plumlee, 2013) between "numerator news" and "denominator news", but is not easy to rationalize in a Bayesian model.

The motivation for this distinction is the well-known shortcut used in finance practice to value an uncertain future cash payoff, where the subjective mean cash payoff (numerator) is divided by a chosen discount factor (denominator). Numerator news is supposed to be news about the mean payoff and denominator news is about the discount rate. Since the discount rate depends on the asset's payoff covariance with the market of all assets, denominator news is understood as news that alters that subjective covariance.[2]

The Bayesian issue is that information that affects the joint payoff distribution will not usually, or possibly ever, affect only one of the first two payoff moments. There are so many ways in which both parameters will move at once, and often dependently. Realistic conditions under which one will move without any change in the other are not easy to imagine. For example, if the firm's expected sales units double, it is hard to imagine that there would be no change in the subjective variance or perceived potential for wide swings, and greater sensitivity to the market. Similarly, any change in the firm's operations (e.g. operating leverage) or marketing will affect both parameters, as would a change in firm management or business strategy. A change in operating leverage causes naturally disproportionate changes in fixed and variable costs, so both mean and (co)variance will move.

In some statistical distributions, the two parameters cannot move independently. For example, in a Bernoulli distribution with index θ, the mean is θ and the variance is $\theta\,(1-\theta)$. A similar issue occurs in the more complicated but standard Bayesian inference model for a normal distribution where both parameters are unknown. The predictive mean and variance are dependent.

A Bayesian picture of relevant information about a firm's cash flows is that there will virtually always be implications in that information for the whole payoff distribution, including all of its moments and all of its covariances with other assets.

Another "more finance" point that came out of Lambert *et al.* (2007) is that under a correct rather than shortcut expression of CAPM, the required rate of return or discount rate on an asset hinges on its mean payoff as well as on its payoff covariance. Numerator news is therefore

[2] Denominator news might also include anything that affects economy-wide discount rates, such as a generally greater reluctance to invest.

always also denominator news. Cheynel (2013) correctly captures this CAPM fact by treating payoff risk or variance as "per unit of mean".

16.7 Mixtures of normals

In mean–variance portfolio theory, the payoff from a weighted portfolio of assets is a sum of normal variates, which is normal. But in Bayesian inference and decision-making generally, the usual situation is that the probability distribution of a payoff from any single asset is best modelled as a mixture distribution of payoffs. For example, if sales dollars have one distribution under a cold winter and another under a warm winter, then the unconditional distribution, relevant to any inference made before we know how the season turns out, is a mixture distribution of the two conditional distributions.

The Bayesian issue for much accounting research is not that mixture distributions are the natural way to understand the variable of interest, typically for example a payoff. The problem for modelling is that mixtures of normals are not themselves normal. They are bimodal and potentially often a very long way from being normal distributions.

There would seem to be many papers that treat mixture distributions of normals as normal distributions, apparently so as not to lose the tractability of a Bayesian model based on normal distributions. See for example Armstrong *et al.* (2016). That is a mistake of statistics in the model, but might be excused as an approximation. Whether results are sensitive to such an approximation is a good research question, given the general applicability in real-world contexts of mixture distributions. Baron (1977, p. 1692) and Liu (2004, p. 233) explain that probability mixtures of different assets or payoffs occur naturally throughout business.[3] See also Winkler (1973, p. 399). Barth (2006a, p. 280) gives a specific example to do with estimating future cash flows when they depend on whether or not a new contract is entered into, and the estimated probability of that occurring.

[3] For discusssion on the behavior of mixture assets on the mean–variance plane, see Baron (1977) and Johnstone and Lindley (2013).

16.8 Information content

Verrecchia (1990, p. 365) distinguishes correctly between whether a signal carries "good" or "bad" news, and whether it carries strong news. Bad news is often depicted in markets as generally stronger than good, for lots of reasons (e.g. the firm would not give out bad news if it was false, whereas good news might commonly be inflated). Verrecchia notes the mistake of confusing the signal "realization" with its quality per se:

> Presumably, information quality involves the distributional characteristics of an event (e.g. its variance) whereas a realization is simply the outcome of the event itself. (Verrecchia, 1990, p. 365)

In Bayesian calculations, "signal content", as in the traditional accounting expression "information content", is fully evident only once the observer knows both the signal observed (e.g. this years earnings are −$20 mill.) and its apparent informational quality.

It is taken for granted in some accounting research that better quality ("more precise") accounting information tends of itself resolve uncertainty and hence bring a lower-risk premium on capital investment, no matter what that information actually "says". Johnstone (2016) explained that the effect of information on certainty (and the cost of capital) depends not merely on signal quality (e.g. sample size) but on what the signal says in either its stated amount (−$20 mill.) or at least in terms of its "direction" (e.g. good news or bad news). Bayesian "information content" is an inseparable mix of what a signal "says" and how strongly or credibly it says it.

Statistically, a binomial sample of 100 tosses of a typical coin is usually quite enough when the observed frequency of heads is 0.505 to be very certain that the coin's "parameter" is very close to 0.5, and hence to remain sure that we are extremely unsure what the next toss will be. If however the observed frequency of heads in 100 tosses was 90% then we would be much more sure what the next toss will be (i.e. it will very probably be heads — and the coin is biased).

Information, therefore, brought by a large sample, comes from a combination of the quality (e.g. unbiased and high precision) of the experiment combined with its observed outcome. Much of course still depends on where the user started in terms of prior knowledge. Information that contradicts prior beliefs, but not too strongly, will virtually always add to uncertainty, often frustratingly so. The point of principle is that by changing users' beliefs, information proves its "information content". That change in beliefs might sometimes be a confirmation of them, in the sense that the posterior distribution becomes tighter in the same location (i.e. around the same parameter value). Confirmation can be as informative as contradiction.

16.9 Fundamental versus information risk

In a Bayesian setting, all uncertainty is homogeneous or fungible, in the sense that it is impossible to know a difference between uncertainty due to nature and uncertainty attributable to the observer's own limitations. Yee (2006, p. 837) drew a distinction between natural or "physical" uncertainty in a dividend stream and uncertainty about future dividends brought by poor information or "earnings quality". Similarly, Zhang (2013) separates uncertainty caused by suspected accounting errors from uncertainty caused by unforeseeable events. However, more in keeping with the Bayesian view, Francis *et al.* (2004) and Francis *et al.* (2007) and others regard "fundamental risk" and "information risk" as inseparable.

A Bayesian decision maker forecasting a future cash flow has only information to go on, she does not observe nature other than through information or models, so in a sense all her uncertainty is "information uncertainty", particularly given that observers will be generally uncertain about how "precise" (e.g. reliable) their own received signals are:

> ... the market is unlikely to have perfect knowledge of the signal precision ex ante. ... Note that uncertain precision does not imply that the market has no ex ante information regarding signal quality, it merely implies that the market does not have perfect ex ante information. (Subramanyam, 1996, p. 208)

16.10 When information adds to information asymmetry

Different levels of conditioning can lead the same piece of information to have opposite Bayesian implications (e.g. see the discussion on Simpson's paradox). Essentially, the finding is that a given signal x_1 can have opposite meanings depending on: (i) its perceived correlation with another piece of information x_2, and (ii) the perceived error characteristics of that other signal. If a third signal is introduced to the Bayesian analysis, the same kind of reversal can happen to one or both of the first two signals. In principle, in infinite populations, there is no end to this possible reversal, as more information is added piece by piece and all joint error probabilities are assessed and allowed for in formal Bayesian calculations. Only the limits of current knowledge and methodological perseverance will decide where conditioning ends, and, from there, only the results of the ensuing decisions will give a hint of whether that level of conditioning was adequate.

The arrival of new information can often add to information asymmetry. Signal x_2 can mean different things altogether to those users who already have signal x_1 relative to those who have only x_2, as was shown disconcertingly by Simpson's paradox. The possibility of better information exacerbating information asymmetry is clear in the Bayesian tenet (Bernardo and Smith, 1994, p. 298) that data cannot "speak for itself", but instead might lead to very different conclusions when combined with prior beliefs and related information.

Different agents with different priors and different additional information, possibly even different models and thus likelihood functions, will not interpret the same data as having the same evidential weight or necessarily even the same evidential direction. This contradicts the usual presumption that more or better public information reduces information asymmetry.

Ultimately, in a stationary world, and where users all have the same model, they will be brought together by more data ("data swamps prior"), but in dynamic contexts like the stockmarket, different information users' beliefs can have different foundations and go in opposite directions and perhaps never converge. Du and Huddart (2017) have recently made that point.

16.11 Value of independent information sources

Penman (2009) contests the idea that accounting information, to be accurate in some sense, should mimic actual market outcomes. For example, on that proposition it would be held that the firm's reported net equity is doing something right if it tends to follow the market value of the firm's stock. Penman's intuition is that accounting information can be made more valuable by having a conceptual and practical foundation that is divorced from market valuations and finance methodologies. For example, old-fashioned "matching" of costs and revenues, and accounting "accruals" judged on conventional and possibly conservative lines, can yield the contrarian signal that "saves the day", or at least moderates the possibility of an information cascade.[4]

The Bayesian theory for combining signals is well illustrated in the task of combining expert opinions, where each opinion is expressed as a probability forecast, labelled $\widehat{p}$. Suppose that n forecasters each report their subjective probabilities of firm X going bankrupt. And suppose that the empirical frequency of bankruptcy is about 5% per year. If all of the forecasters report that firm X is in dire bother and has an 80% or more chance of bankruptcy, a Bayesian like Lindley (1982b) and Lindley (1983) would interpret that combination of forecasts in one of several very different ways, depending strongly on how independent the forecasters' prediction are seen to be.

In the limit, the forecasters might be seen as completely dependent. Suppose that the probability that expert j says 0.8 is one, for all j, once it is known that any one of the other n forecasters said 0.8. In that case, the n forecasts are really only one, and the receiver's final assessment will be some Bayesian translation of their agreed 0.8. For example, if the experts are known to be overconfident, and thus not be well calibrated, the receiver's final subjective probability might be only 0.7 or lower.

[4] One way to describe a Bayesian information cascade is a run of past signals such that the next signal cannot, whatever it says, change the inference enough to change the decision, and is therefore never drawn and seen as "valueless".

More interestingly, suppose that the forecasters are independent, and each happens to forecast $\widehat{p} = 0.25$. To make the calculations easier, suppose that forecasts are in buckets and all of the forecasters gave a probability $\widehat{p} \in (0.2, 0.3)$. Remarkably, that set of forecasts can easily turn into a final receiver's probability assessment of nearly one. A simple Bayesian argument shows how. In a world where the empirical frequency of bankruptcy is only 5%, it will be a relatively unusual event for a forecaster to give $\widehat{p}$ in the 0.2 to 0.3 interval. The likelihood ratio of that event might realistically be something like

$$\frac{f(\widehat{p} \in (0.2, 0.3)|bankrupt)}{f(\widehat{p} \in (0.2, 0.3)|not\ bankrupt)} = \frac{0.05}{0.02} = 2.5.$$

The likelihood ratio when say $n = 3$ independent identical forecasters all report $\widehat{p} \in (0.2, 0.3)$ is then $(2.5)^3 = 15.625$ and the corresponding posterior odds given prior odds of 1/19 are 0.822, giving a posterior probability of $odds/(1 + odds) = 0.822/1.822 = 0.45$. If there were five independent forecasters, that probability is 0.84.

Examples like this are emphasized by Lindley (1982b) as an illustration of how intuition is often out of step with Bayesian probability, and how independent information sources can be so much more informative when combined than when taken individually. The natural intuition for some users is to feel that there appears to be a confirmed or consensus belief at a bankruptcy probability of around 0.2 to 0.3, but that inference does not include information about the past statistical "error characteristics" of the forecasters, including their covariance.

Note how a probability statement made by an expert is just another "signal" in a Bayesian analysis, and can be interpreted "for what its worth". Its Bayesian interpretation via a subjective likelihood function can allow for its issuer's motivations and past record, and any suspicion of miscalibration, over-confidence or incompetence on the part of the forecaster. See the Bayesian literature on recalibrating probability forecasts (e.g. Clemen and Winkler, 1987; Clemen and Winkler, 2007). Treatment of another person's probability as just grist for the Bayesian updating mill is essentially the way that Demski treated accounting "numbers".

16.12 How might market probabilities behave?

An elegant way to picture how a market populated by a good number of Bayesian learners might behave is to think of the football betting market raised by Watts (1976, p. 677). Any prediction market will do, and equally so will the market for binary options, where the normalized asset payoff is zero or one, and the time to expiry is finite and usually short (like the length of a football game). Watts (1976, p. 677) conjectured that the price of a prediction market contract — in an efficient market — will be a good probability estimate since otherwise a trader with a model that estimates probabilities more accurately will have an incentive to trade.

In the decades following Watt's comments, a vast literature on the efficiency of betting markets and their ability to aggregate the information and individual probability forecasts of the traders. In principle, these are markets in uncertainty reduced to their essence, so their analogy with the stockmarket is direct and obvious.

I raise this point to introduce Figure 16.1 which plots the price of a contract paying 0 or 100 on the result of a famous US baseball game. The main point is to show how during the game, the market probability (represented by its price) fluctuated widely up and down with the passage of events and how those events were perceived and interpreted by traders.

Figure 16.1 displays how there is no natural monotonic path to certainty. Certainty in any given outcome naturally tends to wax and wane over time. In the stockmarket, there is no end date, so uncertainty about the "end result", or any future dated stock price, is perpetual.

This simple example shows graphically how more information does not mechanically add to a market's certainty. What better and more current information can there be than the real-time state of the game, and the real-time market consensus probability. That information could be seen as an accounting ideal.

There are Bayesian models and assumptions by which certainty accrues with each new single observation, but a real-life market example shows how those models are not natural in their descriptive validity. They apply to stationary well-defined problems, like drawing from an

Figure 16.1: Real-time betting on the Chicago Cubs (vs Florida Marlins) 2003 NLCS, Game 6 (Plot provided by Justin Wolfers).

urn, but not descriptively to many realistic market or asset valuation contexts.

There can hardly be better information than watching the game and the running market price, yet certainty about the contract payoff comes to the market only by the expiry of time, not by any substantive resolution of which team has better "fundamentals" or of why one team should win. During the game there are lengthy periods where certainty is increasing (probability tending towards one or zero) and similar time intervals during which it is decreasing (towards 0.5).

There is a methodological principle for accounting as an information source for investors in this betting market price path. If an investor held a contract on one of these teams winning, and paid for the "fundamental analysis" of an expert in real time, so that she could constantly re-decide whether to realize her position before the game ends (like selling a stock mid-stream), the investor would not fault the expert for becoming less certain of the final outcome at many times during the game. If at one point the expert held that the outcome was a "coin toss", the investor would not infer that she had no expertise. Expertise need not bring certainty, indeed it should often bring circumspection and even deeper uncertainty. Experts or analysts motivated by a need for higher resolution, rather than more "accuracy" at whatever loss of resolution, will often be far wrong.

16.13 "Idiosyncratic" versus "undiversifiable" information

Some attempts (e.g., Gao and Verrecchia, 2012) are made to distinguish between information about a firm's idiosyncratic risk and information about its diversifiable risk. One interpretation is that information about the firm's payoff covariance with the market concerns its systematic or undiversifiable risk, whereas information about its payoff variance that does not change perceptions of its payoff covariance concerns only its idiosyncratic or diversifiable risk.

In realistic Bayesian models of the firm's uncertain payoff, that distinction will rarely work. In general, any relevant news will affect both parameters. As a simple example, suppose that the firm's probability of success is θ where θ is affected by what other firms do. Let there be two market states, A and B, and let the conditional values of θ be θ_A and θ_B. If the probability of state A is p then the prior probability of the firm succeeding is $p\theta_A + (1-p)\theta_B$. So if another firm takes an action that changes p, then even with fixed θ_1 and θ_2, the firm has a new variance and a new covariance with the market, not to mention a new mean. Similarly, a change in either θ_1 or θ_2 will have the same two effects, of itself.

In everyday terms, there is no realistically fundamental news about a firm that does not affect rational perceptions of both its own payoff distribution (variance and mean) and its payoff correlation with the market. Even for example something as idiosyncratic as a change of senior management or a new costing-for-pricing system will have at least some of both Bayesian effects.

Lastly, and possibly most damaging of all, it follows by Lambert *et al.* (2007) that any information about the firm's mean payoff will also affect its beta, and thus be at least partly "undiversifiable" in the sense that it affects the cost of capital. That information might concern only the firm's idiosyncratic activities (e.g. its peculiar "firm-specific" R&D products) and might therefore be archetypically "idiosyncratic news", but, by affecting the firm's ax ante mean payoff, it affects its cost of capital (up or down). This CAPM insight, attributable to Fama (1977) originally and explained in Lambert *et al.* (2007), appears to undo a vast amount of routine finance and accounting discourse.

17

Conclusion

Issues of how information affects beliefs, certainty, decisions and rational risk premia have been the subject of Bayesian theory since the 1950s when the founding Bayesians wrote the first textbooks on Bayesian business decision-making under uncertainty. That connection between Bayesian theory and financial decision-making was cemented in the early literature on Bayesian portfolio optimization, where fundamental Bayesian insights such as the use of predictive distributions were applied to decision problems characterized by innate parameter uncertainty.

Although avowedly Bayesian in principle, accounting theory after Demski, Feltham and others largely detached itself from the source Bayesian literature, and even from the Bayesian finance literature. This monograph is intended to assist PhD students and researchers to re-make that connection.

A traditional understanding of accounting information under efficient markets theory says that better information alters investors beliefs and trades and tends to have a stronger influence, up or down, on the stockmarket. Even confirmatory evidence fits that description, because it tightens investors' belief distributions around the same mean.

By the traditional view, accounting information serves investors in the same way as the sports pages serve gamblers on football games. Investors qua gamblers use the information available to revise their probability assessments. Their new assessments may not prove to be more successful in all cases, but on average they assist towards more profitable betting or investment outcomes.

An appealing viewpoint, underlying much contemporary empirical accounting research, suggests that better information, such as higher quality earnings, reduces uncertainty and hence also reduces the risk premium or market cost of capital. On that understanding, financial reporting standards are evaluated by whether firms disclosing "more" or "more precise" information seem to be "charged" a lower cost of capital. The older and less idealistic Bayesian view is that information which raises new doubts about an asset's future viability and payoff, and causes its ex ante discount rate to increase, is desirable in the sense that "it is always better to know".

Bayes fits

Starting with Demski and Feltham, Bayesian logic has been shown to fit elegantly with the idea of accounting as information for decision-making under uncertainty. The rules of Bayesian logic are nothing but the probability calculus, part of which is Bayes theorem. A Bayesian in the mathematical sense is merely someone who applies the laws of probability, to revise and reconcile beliefs.

Critics of Bayesian inference traditionally balked at the subjectivity of the prior belief, usually overlooking the innate subjectivity of any model. The Bayesian response is that all beliefs have a starting point and are subjective, so why not express that subjectivity openly and use it advantageously to incorporate factors in the inference and decision that would otherwise be relegated to afterthoughts, and possibly go financially unhedged. As a conceptual framework for understanding uncertain inference in markets and the value and limitations of information, textbook subjectivist Bayesian theory is as good an ideal as exists.

"Accurate", not certain, beliefs

Might accounting have followed a false lead by regarding information as a way to resolve uncertainty rather than to expose and gauge it? The premise that information should, merely by being informative, resolve uncertainty, led appealingly to a proposition that better accounting information would be naturally reflected by the market in lower assessments of risk and lower-risk premia.

Any call for more certainty demands probability assessments that are "sharper" or higher "resolution", but does not require those probability assessments to be "well-calibrated". As an objective, "greater certainty" prioritizes certainty and a low discount rate (cost of capital) over calibration and the "right" discount rate.

The "right" discount rate is analogous to a well-calibrated forecast, it puts the risky asset in its "natural" cost of capital bracket, whether that is high or low. Thus, just as a well-calibrated probability forecaster is able to say accurately that today is a 20–25% day in terms of the frequency of rain, a "well-calibrated" cost of capital is one that categorizes firms into perceived risk bins that are borne out by empirical ex post frequencies (e.g. firm failure rates). For example, of the subset of firms classed as having about a 10% subjective probability of bankruptcy, about 10% go bankrupt.

The mistake of seeking an ever lower-risk assessment (and lower cost of capital) for every firm is evident when it comes to assessing the probability of a firm defaulting on its debt. Any information which leads the market to see all firms in the "near zero probability of default" bracket is not what is expected or even desired by financial information users. Instead, the user or market seeks to identify those firms that are (much) more probable to default, so as to buy some and sell others.

A Bayesian view of information is that sometimes, perhaps often, the more we learn about a risk, the wider or flatter our posterior probability distribution becomes. From this perspective, which applies to doctors, lawyers, security analysts, sports pundits and every profession dealing with uncertain outcomes, not just accounting, the expert (e.g. auditor) is manifestly not doing her job if her disclosures do not sometimes or often add greatly to users' uncertainty and discomfort.

Economic Darwinism

Probabilities near 0 or 1 are always of interest, and an individual investor would like to be able to accurately attach either a 0 or 1 to the probability of each firm in the market staying solvent. But any model or information source that purports to do anything like that will almost surely lose out very quickly to a model that produces well-calibrated probabilities distributed over the full unit interval.

An implication for accounting, with respect to its ability to assist decision makers, is that an investor encouraged to form unrealistically confident beliefs will likely quickly lose money. Economic Darwinism generally rewards more accurate probabilities, not more certain ones, depending of course on how aggressively the user acts upon those probabilities.

For example, if the market maker in ignorance posts odds of exactly one, or a market price of 0.5, on a 0/1 binary outcome, then a log utility investor with subjective belief $\pi = p(1) = 0.8$ expects a compound return from betting against the market maker equal to

$$\exp\left[\pi \ln\left(\frac{0.8}{0.5}\right) + (1-\pi)\ln\left(\frac{0.2}{0.5}\right)\right] - 1,$$

which will turn out positive on (geometric) average only if the frequency of 1's in the realized series exceeds 66.1%. Put another way, if the frequency of 1's is just 66.1% and the log utility investor believes any probability equal to or greater than 0.8, her average realized return will be negative. This example shows how an investor can assess a probability nearer to the actual observed frequency than the market probability and yet still lose. That underlines how essential it is for decision makers to make accurate probability assessments.

Numerical examples like this clarify and reinforce an ex post perspective on the value of information. Our log utility investor, had she believed probability $p(1) = 0.9$ ex ante, would have been delighted with her ex ante expected utility or expected capital gain, but would have been less well pleased with her compound return ex post.

Accounting theory, understandably given its focus on the design of accounting rules and disclosure contracts, is pre-occupied with ex ante expected utility. But decision analysis within the theory of economic

Darwinism shows how ex ante expectations are only worthwhile if ultimately realized (in dollars).

A very favorable result for Bayesian users of accounting information is that an astute investor does not need a great degree of certainty to "beat the market". Instead, just a small information edge, played well in the knowledge that it is only a small edge, is sufficient for capital growth.

A Bayesian investor is typically accepting of highly uncertain beliefs, and is often not confident that more information will bring more resolution. That circumspection can be the key to investment success.

Information about information

"Information uncertainty", so-called, is merely one contributor to overall uncertainty. Different "types" of uncertainty are fungible in the sense that they blur into one predictive distribution when combined using the laws of probability. Under Bayes theorem, different grounds for uncertainty do not simply accumulate to more or deeper uncertainty, but instead combine in sometimes counter-intuitive and unexpected ways.

Each new cause for uncertainty incorporated in beliefs, including uncertainty about the accuracy or other qualities of existing information, amounts merely to another level of Bayesian conditioning or learning. At each step, further conditioning can reinforce and accentuate previous beliefs, or leave them unchanged, or possibly even weaken or reverse them, regardless of how relevant or perhaps irrelevant that new information appears of itself.

The intuition that information uncertainty must exacerbate fundamental uncertainty is misconceived. It holds formally in some limited and tractable model forms, as is well understood, but it does not hold as a general Bayesian law. Since information must sometimes increase uncertainty, allowance for the "information risk" of that information, or for the possibility that it is wrong, can sometimes restore overall certainty.

If Bayesian logic is to guide equity investors and others in their inference and decision-making, allowance has to be made for the reality that the variances and covariances between assets' or firms' cash flows are unknown and uncertain. This alone means that better information — call it reduced "information uncertainty" — can lead to upward revision of the predictive payoff covariances, and hence greater uncertainty.

Think no further than a strictly Bayesian statistical model in which a larger data set increases the precision of the estimated population variance (or covariance), and therefore reduces "information uncertainty", but produces a higher posterior mean for that unknown variance. In principle, we now know the variance more accurately, and we know that it is higher than previously believed. Analogously, a more expert fundamental analysis of a firm's operations, costs and revenues, can reveal that the firm's net cash flows are more correlated with market conditions than was first understood. Note that in both of these cases, the variance or covariance is treated as an unknown parameter, unlike in the formative finance parameter risk models where the covariance matrix is taken as known or given.

Accounting has little choice but to be broad minded in its models of firm cash flows and investment returns. Firm fundamentals have all manner of statistical distributions, including natural mixture distributions. If the payoff from an investment hinges for instance on the discrete outcome of a regulatory or court decision, then its probability distribution is inevitably a mixture distribution, and can be more correlated with the market under one regulatory outcome than another. Even if its payoff, conditional on a given level of regulation, is normal, its mixture distribution is not, and none of the distribution parameters are known with certainty.

Signs of good accounting information

The clearest ex post indication that accounting information is perceived as relevant and reliable is the stock price change associated with the disclosure. Asset prices under CAPM are a function of the predictive distribution of future payoffs and hence are a composite of (at least) the first and second-moments of that distribution. It would be a deep insight for standard setters if research were to reveal whether prices are generally more sensitive to revisions in payoff mean than revisions in payoff (co)variance (risk). It has been suggested in the portfolio optimization literature in finance that optimal portfolios are far more sensitive to estimated mean returns than estimated covariances. The possible effects of such an insight on standard setting are clearly complicated, but some simple conclusions would follow.

For example, if accurate Bayesian assessment of the mean cash flow is more critical, then investors will be relatively more concerned to correctly assess the "direction" (i.e. "good" or "bad")/ of the accounting news, than its precision. Analysts might quibble over whether the reported earnings should have been a bit less or a bit more, and the normative accounting objective of better "earnings quality" might suggest that the reported figure is (say) insufficiently conservative, but the investor is basically concerned with whether the earnings report points to a higher expected future cash flow or a lower one. The task would be one of getting the "location" of the posterior predictive distribution right, rather than only shrinking its variance. Put another way, should the mean be revised up (buy) or down (sell)? That is the primary question in valuation. The secondary question is "by how much", and that is where the perceived variance plays a greater role.

Acknowledgements

The author thanks Mark Tippett, Tom Dyckman and Jim Ohlson for helpful comments and suggestions.

References

Aitkin, M. (2010). *Statistical Inference: An Integrated Likelihood/Bayesian Approach*. Boca Raton: Chapman and Hall.

Alchain, A. (1950). "Uncertainty, evolution and economic theory". *Journal of Political Economy*. 58: 211–221.

Amiram, D., W. R. Landsman, E. L. Owens, and S. R. Stubben (2018). "How are analysts' forecasts affected by high uncertainty?" *Journal of Business Finance and Accounting*. 45: 295–318.

Armstrong, C. S., D. J. Taylor, and R. E. Verrecchia (2016). "Asymmetric reporting". *Journal of Financial Reporting*. 1: 15–32.

Artiach, T. C. and P. M. Clarkson (2011). "Disclosure, conservatism and the cost of capital: A review of the foundational literature". *Accounting and Finance*. 51: 2–49.

Avramov, D. (2002). "Stock return predictability and model uncertainty". *Journal of Financial Economics*. 64: 423–458.

Avramov, D. and G. Zhou (2010). "Bayesian portfolio analysis". *Annual Review of Financial Economics*. 2: 25–47.

Banks, D. L. (1996). "A conversation with I. J. Good". *Statistical Science*. 11: 1–19.

Baron, D. P. (1977). "On the utility theoretic foundations of mean-variance analysis". *Journal of Finance*. 32: 1683–1697.

Barry, C. B. (1978). “Effects of uncertain and nonstationary parameters upon capital market equilibrium”. *Journal of Financial and Quantitative Analysis*. 13: 419–433.

Barry, C. B. and S. J. Brown (1985). “Differential information and security market equilibrium”. *Journal of Financial and Quantitative Analysis*. 20. 4-7-422.

Barry, C. B. and R. L. Winkler (1976). “Nonstationarity and portfolio choice”. *Journal of Financial and Quantitative Analysis*. 11: 217–235.

Barth, M. E. (2006a). “Including estimates of the future in today’s financial statements”. *Accounting Horizons*. 20: 271–285.

Barth, M. E. (2006b). “Research, standard setting, and global financial reporting”. *Foundations and Trends in Accounting*. 1: 71–165.

Barth, M. E. and D. J. Taylor (2010). “In defense of fair value: Weighing the evidence on earnings management and asset securitizations”. University of Pennsylvania. Wharton Faulty Research Paper 2-2010.

Barth, M., W. H. Beaver, and W. R. Landsman (2001). “The relevance of the value relevance literature for financial accounting standard setting: Another view”. *Journal of Accounting and Economics*. 31: 77–104.

Basu, S. (1997). “The conservatism principle and the asymmetric timeliness of earnings”. *Journal of Accounting and Economics*. 24: 3–37.

Basu, S. (2013). “Devil’s advocate: The most incorrect beliefs of accounting experts”. *Accounting Horizons*. 27: 841–846.

Basu, S. (2015). “Is there a scientific basis for accounting? Implications for practice, research, and education”. *Journal of International Accounting Research*. 14: 235–265.

Bawa, V. S., S. J. Brown, and R. W. Klein (1979). *Estimation Risk and Optimal Portfolio Choice*. Amsterdam: North-Holland.

Beaver, W. (1968). “The Information content of annual earnings announcements”. *Journal of Accounting Research. Supplement*. 6: 67–92.

Beck, K., B. Niendorf, and P. Peterson (2012). “The use of Bayesian methods in financial research”. *Investment Management and Financial Innovations*. 9: 68–75.

Bedford, N. M. and V. Baladouni (1962). “A communication theory approach to accountancy”. *The Accounting Review*. 37: 650–659.

Benjamin, D. J., J. O. Berger, and V. E. Johnson (2018). “Redefine statistical significance”. *Nature Human Behaviour*. 2: 6–10.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. New York: Springer-Verlag.

Berger, J. O. (2003). “Could Fisher, Jeffreys and Neyman have agreed on testing?” *Statistical Science*. 18: 1–32.

Berger, J. O. and T. Sellke (1987). “Testing a point null hypothesis: The irreconcilability of p-values and evidence”. *Journal of the American Statistical Association*. 82: 112–122.

Berger, J. O. and R. L. Wolpert (1988). *The Likelihood Principle*. 2nd ed. Haywood, California: IMS Lecture Notes — Monograph Series.

Bernardo, J. M. and A. F. M. Smith (1994). *Bayesian Theory*. New York: Wiley.

Bertomeu, J. (2013). “Discussion of earnings manipulation and the cost of capital”. *Journal of Accounting Research*. 51: 475–493.

Bertomeu, J., A. Beyer, and R. A. Dye (2011). “Capital structure, cost of capital, and voluntary disclosures”. *The Accounting Review*. 86: 857–886.

Bertomeu, J., M. N. Darrough, and W. Xue (2017). “Optimal conservatism with earnings manipulation”. *Contemporary Accounting Research*. 34: 252–284.

Best, M. J. and R. R. Grauer (1991). “On the sensitivity of mean-variance efficient portfolios to changes in asset means: Some analytical and computational results”. *The Review of Financial Studies*. 4: 315–342.

Beyer, A. (2009). “Capital market prices, management forecasts, and earnings management”. *The Accounting Review*. 84: 1713–1747.

Beyer, A., D. A. Cohen, T. Z. Lys, and B. R. Walther (2010). “The financial reporting environment: Review of the recent literature”. *Journal of Accounting and Economics*. 50: 296–343.

Beyer, A., D. A. Cohen, T. Z. Lys, and B. R. Walther (2013). "Conservatism and aggregation: The effect on cost of equity capital and the efficiency of debt contracts". *Rock Center for Corporate Governance at Stanford University Working Paper No. 120. Available at SSRN: https://ssrn.com/abstract=2103776 or http://dx.doi.org/10.2139/ssrn.2103776*.

Bhattacharya, N., F. Ecker, P. Olsson, and K. Schipper (2012). "Direct and mediated associations among earnings quality, information asymmetry, and the cost of equity". *The Accounting Review*. In press.

Bickel, J. E. (2007). "Some comparisons among quadratic, spherical, and logarithmic scoring rules". *Decision Analysis*. 4: 49–65.

Bickel, J. E. and J. E. Smith (2006). "Optimal sequential exploration: A binary learning model". *Decision Analysis*. 5: 116–128.

Birchler, U. and M. M. Bütler (2007). *Information Economics*. London: Routledge.

Blackwell, D. (1969). *Basic Statistics*. New York: McGraw-Hill.

Blume, L. and D. Easley (2006). "If you're so smart, why aren't you rich? Belief selection in complete and incomplete markets". *Econometrica*. 74: 929–966.

Botosan, C. (1997). "Disclosure level and the cost of equity capital". *The Accounting Review*. 72: 323–349.

Botosan, C. A. and M. A. Plumlee (2002). "A re-examination of disclosure level and the expected cost of equity capital". *Journal of Accounting Research*. 40: 21–40.

Botosan, C. A. and M. A. Plumlee (2013). "Are information attributes priced?" *Journal of Business Finance and Accounting*. 40: 1045–1067.

Brown, S. J. (1979). "The effect of estimation risk on capital equilibrium". *Journal of Financial and Quantitaive Analysis*. 14: 215–220.

Burgstahler, D. (1987). "Inference from empirical research". *Accounting Review*. 62: 203–214.

Butterworth, J. E. (1972). "The accounting system as information function". *Journal of Accounting Research*. 10: 1–27.

Cabrales, A., O. Gossner, and R. Serrano (2013). "Entropy and the value of information for investors". *American Economic Review*. 103: 360–377.

Chen, Q., T. Hemmer, and Y. Zhang (2007). “On the relation between conservatism in accounting standards and incentives for earnings management”. *Journal of Accounting Research*. 45: 541–565.

Chen, Q. and K. Schipper (2016). “Comments and observations regarding the relation between theory and empirical research in contemporary accounting research”. *Foundations and Trends in Accounting*. 10: 314–360.

Cheng, M., D. S. Dhaliwal, and M. Neamtiu (2011). “Asset securitization, securitization recourse, and information uncertainty”. *The Accounting Review*. 86: 541–568.

Cheynel, E. (2013). “A theory of voluntary disclosures and cost of capital”. *Review of Accounting Studies*. 18: 987–1020.

Chopra, V. K. and W. T. Ziemba (1993). “The effect of errors in means, variances, and covariances on optimal portfolio choice”. *Journal of Portfolio Management*. 19: 6–11.

Christensen, J. A. and J. S. Demski (2003). *Accounting Theory: An Information Content Perspective*. New York: McGraw-Hill.

Christensen, J. A. and J. S. Demski (2007). “Anticipatory reporting standards”. *Accounting Horizons*. 21: 351–370.

Christensen, P. O., L. E. de la Rosa, and G. A. Feltham (2010). “Information and the cost of capital: An ex ante perspective”. *Accounting Review*. 85: 817–848.

Christensen, P. O. and G. Feltham (2003). *Economics of Information, Volume 1: Information in markets*. New York: Springer-Verlag.

Clarkson, P., J. Guedes, and R. Thompson (1996). “On the diversification, observability, and measurement of estimation risk”. *Journal of Financial and Quantitative Analysis*. 31: 69–84.

Clemen, R. T. and R. L. Winkler (1987). “Calibrating and Combining Precipitation Probability Forecasts”. In: *Probability and Bayesian Statistics*. Ed. by R. Viertl. Springer: Boston, MA.

Clemen, R. T. and R. L. Winkler (2007). “Aggregating Probability Distributions”. In: *Advances in Decision Analysis: From Foundations to Applications*. Ed. by E. Ward, R. F. Miles, and D. von Winterfeldt. Cambridge: Cambridge University Press. 154–176.

Coles, J. L. and U. Loewenstein (1988). “Equilibrium pricing and portfolio composition in the presence of uncertain parameters”. *Journal of Financial Economics*. 22: 279–303.

Coles, J., U. Loewenstein, and J. Suay (1995). “On equilibrium pricing under parameter uncertainty”. *Journal of Financial and Quantitative Analysis*. 30: 347–364.

Core, J. E., W. R. Guay, and R. S. Verdi (2008). “Is accruals quality a priced risk factor?” *Journal of Accounting and Economics*. 46: 2–22.

Core, J. E., L. Hail, and R. S. Verdi (2015). “Mandatory disclosure quality, inside ownership and cost of capital”. *European Accounting Review*. 24: 1–29.

Cremers, J.-H., M. Kritzman, and S. Page (2005). “Optimal hedge fund allocations: Do higher moments matter?” *The Journal of Portfolio Management*. 31: 70–81.

Dawid, A. P. (1982). “The well-calibrated Bayesian”. *Journal of the American statistical Association*. 77: 605–610.

de Finetti, B. (1937/1964). “La prevision: ses lois logiques, ses sources subjectives”. *Annales de l’Institut Henri Poincaré*. 7: 1–68. Translated in Kyburg, H. E. and Smokler, H. E. (eds.) 1964 *Studies in Subjective Probability*. Wiley: New York.

Dechow, P. M. (1994). “Accounting earnings and cash flows as measures of firm performance: The role of accounting accruals”. *Journal of Accounting and Economics*. 18: 3–42.

Dechow, P. M., W. Ge, and C. M. Schrand (2010). “Understanding earnings quality: A review of the proxies, their determinants and their consequences”. *Journal of Accounting and Economics*. 50: 344–401.

Dechow, P. M. and C. M. Schrand (2004). *Earnings Quality*. Research Foundation of CFA Institute. 2004.

DeFond, M. L. (2010). “Earnings quality research: Advances, challenges and future research”. *Journal of Accounting and Economics*. 50: 402–409.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

DeGroot, M. H. (1986). “A conversation with David Blackwell”. *Statistical Science*. 1: 40–53.

DeGroot, M. H. and S. E. Fienberg (1982). "Assessing probability assessors: Calibration and refinement". In: *Statistical Decision Theory and Related Topics III*. Ed. by S. S. Gupta and J. O. Berger. New York: Academic Press. 291–314.

DeGroot, M. H. and S. E. Fienberg (1983). "The comparison and evaluation of forecasters". *The Statistician*. 32: 14–22.

Demski, J. S. (1969). "Predictive ability of alternative performance measurement models". *Journal of Accounting Research*: 96–115.

Demski, J. S. (1973). "The general impossibility of normative accounting standards". *The Accounting Review*. 48: 718–723.

Demski, J. S. (1980). *Information Analysis*. 2nd ed. Reading, MA: Addison Wesley.

Demski, J. S. and G. A. Feltham (1972). "Forecast evaluation". *The Accounting Review*. 47: 533–548.

Dichev, I. D., J. R. Graham, C. R. Harvey, and S. Rajgopal (2013). "Earnings quality: Evidence from the field". *Journal of Accounting and Economics*. 56: 1–33.

Du, K. and S. Huddart (2017). "Reporting system, Investor learning, and stock regularities". URL: www.personal.psu.edu/sjh11/Papers/regularities.pdf.

Dyckman, T. R. (2016). "Significance testing: We can do better". *Abacus*. 52: 319–342.

Dye, R. A. (1985). "Disclosure of non-proprietary information". *Journal of Accounting Research*. 23: 123–145.

Dye, R. A. (1990). "Mandatory versus voluntary disclosures: The cases of financial and real externalities". *The Accounting Review*. 65: 1–24.

Dye, R. A. (2001). "An evaluation of "essays on disclosure" and the disclosure literature in accounting". *Journal of Accounting and Economics*. 32: 181–235.

Dye, R. A. (2017). "Some recent advances in the theory of financial reporting and disclosures". *Accounting Horizons*. 31: 39–54.

Dye, R. A. and J. S. Hughes (2018). "Equilibrium voluntary disclosures, asset pricing, and information transfers". *Journal of Accounting and Economics*. In press.

Dye, R. A. and S. S. Sridhar (2002). “Resource allocation effects of price reactions to disclosures”. *Contemporary Accounting Research*. 19: 385–410.

Dye, R. A. and S. S. Sridhar (2007). “The allocational effects of the precision of accounting estimates”. *Journal of Accounting Research*. 45: 731–769.

Ewart, R. and A. Wagenhofer (2011). “Earnings management, conservatism, and earnings quality”. *Foundations and Trends in Accounting*. 6: 65–186.

Ewart, R. and A. Wagenhofer (2017). “Increasing internal controls incentives and welfare through conservative accounting. MEAFA 2017”. URL: http://sydney.edu.au/business/research/meafa/meeting/2017.

Fama, E. F. (1977). “Risk adjusted discount rates and capital budgeting under uncertainty”. *Journal of Financial Economics*. 5: 3–24.

Fama, E. F. and M. Miller (1972). *The Theory of Finance*. Hinsdale, Illinois: Dryden Press.

Feltham, G. A. (1968). “The value of information”. *The Accounting Review*. 43: 684–696.

Feltham, G. A. (2007). “Joel S. Demski: A leader in accounting scholarship”. In: *Essays in Accounting Theory in Honour of Joel S. Demski*. Ed. by R. Antel, F. Gjesdal, and P. Jinghong Liang. New York: Springer. 1–32.

Feltham, G. and J. Demski (1970). “The use of models in information evaluation”. *The Accounting Review*. 45: 623–640.

Fenton, N., M. Neil, and D. Berger (2016). “Bayes and the law”. *Annual Review of Statistics and Its Application*. 3: 51–77.

Fienberg, S. E. (2006). “When did Bayesian inference become Bayesian?” *Bayesian Analysis*. 1: 1–40.

Fienberg, S. E. (2008). “The early statistical years: 1947–1967. A conversation with Howard Raiffa”. *Statistical. Science*. 23: 136–149.

Fischer, P. E. and P. C. Stocken (2001). “Imperfect information and credible communication”. *Journal of Accounting Research*. 39: 119–134.

Fischer, P. E. and R. E. Verrecchia (2000). “Reporting bias”. *The Accounting Review*. 75: 229–245.

Fishburn, P. C. (1986). “The axioms of subjective probability”. *Statistical Science*. 1: 335–345.

Foster, G. (1978). *Financial Statement Analysis*. 1st ed. Englewood Cliffs, N.J: Prentice-Hall.

Foster, N. (2003). *The FASB and the Capital Markets*. The FASB Report. Norwalk, CT.: FASB.

Francis, J., R. LaFond, P. Olsson, and K. Schipper (2004). “Costs of equity and earnings attributes”. *The Accounting Review*. 79: 967–1010.

Francis, J., R. LaFond, P. Olsson, and K. Schipper (2005). “The market pricing of accruals quality”. *Journal of Accounting and Economics*. 39: 295–327.

Francis, J., R. LaFond, P. Olsson, and K. Schipper (2007). “Information uncertainty and post-earnings announcement drift”. *Journal of Business Finance and Accounting*. 34: 403–433.

Francis, J., P. Olsson, and D. J. Nanda (2008). “Voluntary disclosure, earnings quality, and costs of capital”. *Journal of Accounting Research*. 46: 53–99.

Francis, J., P. Olsson, and K. Schipper (2006). “Earnings quality”. *Foundations and Trends in Accounting*. 1: 259–340.

Gao, P. (2010). “Disclosure quality, cost of capital and investor welfare”. *The Accounting Review*. 85: 1–29.

Gao, P. (2013a). “A measurement approach to conservatism and earnings management”. *Journal of Accounting and Economics*. 55: 251–268.

Gao, P. (2013b). “A two-step representation of accounting measurement”. *Accounting Horizons*. 27: 861–866.

Gao, P. (2015). “Optimal thresholds in accounting recognition standards”. Chicago Booth Research Paper No. 15-50.

Gao, P. and R. E. Verrecchia (2012). “Economic consequences of idiosyncratic information in diversified markets”. Chicago Booth Research Paper No. 12-09.

Gao, P. and G. Zhang (2018). “Accounting manipulation, peer pressure, and internal control”. *The Accounting Review*. Forthcoming.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004). *Bayesian Data Analysis*. 2nd ed. Boca Raton, FL.: CRC Press.

Gigler, F. B. and T. Hemmer (2001). “Conservatism, optimal disclosure policy, and the timeliness of financial reports”. *The Accounting Review*. 76: 471–493.

Giocoli, N. (2013). “From Wald to Savage: Homo economicus becomes a Bayesian statistician”. *Journal of the History of the Behavioral Sciences*. 49: 63–95.

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007). “Probabilistic forecasts, calibration and sharpness”. *Journal of the Royal Statistical Society: Series B*. 69: 243–268.

Gneiting, T. and A. E. Raftery (2007). “Strictly proper scoring rules, prediction, and estimation”. *Journal of the American Statistical Association*. 102: 359–378.

Gonedes, N. J. (1975). “Information production and the capital market”. *Journal of Finance*. 30: 841–864.

Gonedes, N. J. (1976). “The capital market, the market for information, and external accounting”. *Journal of Finance*. 31: 611–630.

Gonedes, N. and Y. Ijiri (1974). “Improving subjective probability assessment for planning and control in team-like organizations”. *Journal of Accounting Research*. 12: 251–269.

Good, I. J. (1950). *Probability and the Weighing of Evidence*. Griffin: London.

Good, I. J. (1952). “Rational decisions”. *Journal of the Royal. Statistical Society. Series B*. 14: 107–114.

Good, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. MIT Press.

Gox, R. and A. Wagenhofer (2010). “Optimal precision of accounting information in debt financing”. *European Accounting Review*. 19: 579–602.

Grushka-Cockayne, Y., V. R. J. Jose, and K. C. Lichtendahl (2017). “Ensembles of overfit and overconfident forecasts”. *Management Science*. 63: 1110–1130.

Guay, W. and R. Verrecchia (2006). “Discussion of an economic framework for conservative accounting and Bushman and Piotroski (2006)”. *Journal of Accounting and Economics*. 42: 149–165.

Hakansson, N. and W. Ziemba (1995). "Capital Growth Theory". In: *Handbooks in Operations Research and Management Science: Finance*. Ed. by R. Jarrow, V. Maksimovic, and W. Ziemba. Elsevier. 65–86.

Harvey, C. R. (2017). "Presidential address: The scientific outlook in financial economics". *Journal of Finance*. 72: 1399–1440.

Harvey, C. and G. Zhou (1990). "Bayesian inference in asset pricing tests". *Journal of Financial Economics*. 26: 221–254.

Healy, P. M. and K. G. Palepu (2001). "Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature". *Journal of Accounting and Economics*. 31: 405–440.

Heinle, M. S. and K. C. Smith (2015). "A theory of risk disclosure". *Review of Accounting Studies*: In press.

Heinle, M. S., K. C. Smith, and R. E. Verrecchia (2018). "Risk-factor disclosure and asset prices". *The Accounting Review*. 93: 191–208.

Hirshleifer, J. and J. G. Riley (1992). *The Analytics of Uncertainty and Information*. Cambridge: Cambridge University Press.

Hitz, J.-M. (2007). "The decision usefulness of fair value accounting — A theoretical perspective". *European Accounting Review*. 16: 323–362.

Hodder, L., P. Hopkins, and K. Schipper (2014). "Fair value measurement in financial reporting". *Foundations and Trends in Accounting*. 8: 143–270.

Holthausen, R. W. and R. L. Watts (2001). "The relevance of the value-relevance literature for financial accounting standard setting". *Journal of Accounting and Economics*. 31: 3–75.

Howard, R. A. (1966). "Information value theory". *IEEE Transactions on Systems, Science and Cybernetics*. 2(1): 22–26.

Howson, C. and P. Urbach (2005). *Scientific Reasoning: The Bayesian Approach*. 3rd ed. Chicago: Open Court.

Hribar, P. and N. T. Jenkins (2004). "The effect of accounting restatements on earnings Revisions and the estimated cost of capital". *Review of Accounting Studies*. 9: 337–356.

Hughes, J. S., J. Liu, and J. Liu (2007). "Information asymmetry, diversification, and the cost of capital". *The Accounting Review*. 82: 705–729.

Hull, J. C. (1986). “A note on the risk-adjusted discount rate method”. *Journal of Business Finance and Accounting*. 13: 445–450.

Ijiri, Y. and R. Jaedicke (1966). “Reliability and objectivity of accounting measurements”. *The Accounting Review*. 41: 474–483.

Jiang, G., C. M. Lee, and Y. Zhang (2005). “Information uncertainty and expected returns”. *Review of Accounting Studies*. 10: 185–221.

Johnstone, D. J. (1986). “Tests of significance in theory and practice (with discussion by D.V Lindley and G. Barnard)”. *Journal of the Royal Statistical Society. Series D (The Statistician)*. 35: 491–504.

Johnstone, D. J. (1987a). “On the interpretation of hypothesis tests following Neyman and Pearson”. In: *Probability and Bayesian Statistics*. Ed. by R. Viertl. Boston, MA: Springer. 267–277.

Johnstone, D. J. (1987b). “Tests of significance following R. A. Fisher”. *The British Journal for the Philosophy of Science*. 38: 481–499.

Johnstone, D. J. (1997). “Comparative classical and Bayesian interpretations of statistical compliance tests in auditing”. *Accounting and Business Research*. 28: 53–82.

Johnstone, D. J. (2011). “Economic interpretation of probabilities estimated by maximum likelihood or score”. *Management Science*. 57: 308–314.

Johnstone, D. J. (2015). “Information and the cost of capital in a mean-variance efficient market”. *Journal of Business Finance and Accounting*. 42: 79–100.

Johnstone, D. J. (2016). “The effect of information on uncertainty and the cost of capital”. *Contemporary Accounting Research*. 33: 752–774.

Johnstone, D. J. (2017). “Sensitivity of the discount rate to the expected payoff in project valuation”. *Decision Analysis*. 14: 126–136.

Johnstone, D. J. and D. V. Lindley (1995). “Bayesian inference given data ‘significant at α’: Tests of point hypotheses”. *Theory and Decision*. 38: 51–60.

Johnstone, D. J. and D. V. Lindley (2013). “Mean–variance and expected utility: The Borch paradox”. *Statistical Science*. 28: 223–237.

Jose, V. R., R. F. Nau, and R. L. Winkler (2008). “Scoring rules, generalized entropy, and utility maximization”. *Operations Research*. 56: 1146–1157.

Kadane, J. B. (2011). “Partial-Kelly strategies and expected utility: Small-edge asymptotics”. *Decision Analysis*. 8: 4–9.

Kalymon, B. A. (1971). “Estimation risk in the portfolio selection model”. *Journal of Financial and Quantitative Analysis*. 6: 559–582.

Kelly, J. L. (1956). “A new interpretation of information rate”. *Bell System Technical Journal*. 35: 917–926.

Kim, D. and Y. Qi (2010). “Accrual quality, stock returns, and macroeconomic conditions”. *The Accounting Review*. 85: 937–978.

Klein, R. W. and V. S. Bawa (1976). “The effect of estimation risk on optimal portfolio choice”. *Journal of Financial Economics*. 3: 215–231.

Klein, R. W. and V. S. Bawa (1977). “The effect of limited information and estimation risk on optimal portfolio diversification”. *Journal of Financial Economics*. 5: 89–111.

Kothari, S. P. (2001). “Capital markets research in accounting”. *Journal of Accounting and Economics*. 31: 105–231.

Kravet, T. and T. Shevlin (2010). “Accounting restatements and information risk”. *Review of Accounting Studies*. 15: 264–294.

Lambert, R. (2010). “Discussion of implications for GAAP from an analysis of positive research in accounting”. *Journal of Accounting and Economics*. 50: 287–295.

Lambert, R. A. and R. E. Verrecchia (2010). “Cost of capital in imperfect competition settings”. URL: https://www8.gsb.columbia.edu/faculty-research/sites/faculty-research/files/accounting/RICHARDLAMBERT.pdf.

Lambert, R., C. Leuz, and R. E. Verrecchia (2007). “Accounting information, disclosure, and the cost of capital”. *Journal of Accounting Research*. 45: 385–420.

Larson, C. R. and R. J. Resutek (2017). “Types of investor uncertainty and the cost of equity capital”. *Journal of Business Finance and Accounting*. 44: 1169–1193.

Laux, V. and P. C. Stocken (2017). “Accounting standards, regulatory enforcement, and innovation”. URL: https://ssrn.com/abstract=2713170.

Lee, L. C. and N. M. Bedford (1969). “An information theory analysis of the accounting process”. *Accounting Review*. 44: 256–275.

Leuz, C. and P. Wysocki (2008). "Economic consequences of financial reporting and disclosure regulation: A review and suggestions for future research". URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1105398.

Leuz, C. and P. Wysocki (2016). "The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research". *Journal of Accounting Research*. 54: 525–622.

Lewellen, J. and J. Shanken (2002). "Learning, asset-pricing tests, and market efficiency". *Journal of Finance*. 57: 1113–1145.

Lindley, D. V. (1957). "A statistical paradox". *Biometrika*. 44: 187–192.

Lindley, D. V. (1965). *Introduction to Probability and Statistics*. Parts 1 and 2. Cambridge University Press.

Lindley, D. V. (1980). "L. J. Savage — his work in probability and statistics". *Annals of Statistics*. 8: 1–24.

Lindley, D. V. (1982a). "Scoring rules and the inevitability of probability". *International Statistical Review*. 50: 1–26. (with discussion).

Lindley, D. V. (1982b). "The improvement of probability judgements". *Journal of Royal Statistical Society Series A*. 145: 117–126.

Lindley, D. V. (1983). "Reconciliation of probability distributions". *Operations Research*. 31: 866–880.

Lindley, D. V. (1984). "A Bayesian lady tasting tea". In: *Statistics An Appraisal*. Ed. by H. A. David and H. T. David. Ames: Iowa State University Press. 455–485.

Lindley, D. V. (2014). *Understanding Uncertainty*. New York: John Wiley.

Lindley, D. V. and L. D. Phillips (1976). "Inference for a Bernoulli process (a Bayesian view)". *American Statistician*. 30: 112–119.

Lindsay, R. M. (1995). "Reconsidering the status of tests of significance: An alternative criterion of adequacy". *Accounting, Organizations and Society*. 20: 35–53.

Liu, Y. (2004). "A new foundation for the mean-variance analysis". *European Journal of Operations Research*. 158: 229–242.

Lo, K. (2008). "Earnings management and earnings quality". *Journal of Accounting and Economics*. 45: 350–357.

Luenberger, D. (1998). *Investment Science*. New York: Oxford University Press.

MacLean, L. C., R. Sanegre, Y. Zao, and W. T. Ziemba (2004). “Capital growth with security”. *Journal of Economic Dynamics and Control*. 28: 937–954.

MacLean, L. C., W. T. Ziemba, and G. Blazenko (1992). “Growth versus security in dynamic investment analysis”. *Management Science*. 38: 1562–1585.

MacLean, L. and W. T. Ziemba (1999). “Growth versus security tradeoffs in dynamic investment analysis”. *Annals of Operations Research*. 85: 193–225.

MacLean, L., W. T. Ziemba, and Y. Li (2005). “Time to wealth goals in capital accumulation and the optimal trade-off of growth versus security”. *Quantitative Finance*. 5: 343–357.

Markowitz, H. (1991). “Foundations of portfolio theory”. *Journal of Finance*. 46(2): 469–477.

Markowitz, H. (2006). “de Finetti scoops Markowitz”. *Journal of Investment Management*. 4: 5–18.

Mattessich, R. V. (2003). “The rise and significance of modern analytical methods in accounting. Part I. A review essay of accounting theory: An information content perspective, of John A. Christensen and Joel Demski”. *Energeia*. 2: 133–144.

Mattessich, R. V. (2006). “The information economic perspective of accounting: Its coming of age”. *Canadian Accounting Perspective*. 5: 209–236.

Mcgrayne, S. B. (2011). *The Theory that Would not Die: How Bayes’ Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. Yale University Press.

Meyer, J. (1987). “Two-moment decision models and expected Utility”. *The American Economic Review*. 77: 421–430.

Mock, T. (1971). “Concepts of information value and accounting”. *The Accounting Review*. 46: 765–778.

Mock, T. (1973). “The value of budget information”. *The Accounting Review*. 48: 520–534.

Morrison, D. E. and R. E. Henkel (1970). *The Significance Test Controversy*. Chicago: Aldine.

Murphy, A. H. (1966). “On expected-utility measures in cost-loss ratio decision situations”. *Journal of Applied Meteorology*. 8: 989–991.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.

Neururer, T., G. Papadakis, and E. Riedl (2016). “Tests of investor learning models using earnings innovations and implied volatilities”. *Review of Accounting Studies*. 21: 400–437.

Neyman, J. (1950). *First Course in Probability and Statistics*. New York: Henry Holt.

Ng, D. S. (1978). “An information economics analysis of financial reporting and external auditing”. *The Accounting Review*. 53: 910–920.

Ng, J. (2011). “The effect of information quality on liquidity risk”. *Journal of Accounting and Economics*. 52: 126–143.

O’Hagen, A. (1994). *Kendall’s Advanced Theory of Statistics. Vol. 2B: Bayesian Inference*. London: Edward Arnold.

Ohlson, J. A. (1980). “Financial ratios and the probabilistic prediction of bankruptcy”. *Journal of Accounting Research*. 19: 109–131.

Ohlson, J. A. (2015). “Accounting research and common sense”. *Abacus*. 51: 525–535.

Ohlson, J. A. (2018). “Researchers’ data analysis choices: An excess of false positives?” URL: https://ssrn.com/abstract=3089571.

Ottaviani, M. and P. Sorensen (2015). “Price reaction to information with heterogeneous beliefs and wealth effects: Underreaction, momentum, and reversal”. *American Economic Review*. 105: 1–34.

Pastor, L. and P. Veronesi (2009). “Learning in financial markets”. *Annual Review of Financial Economics*. 1: 361–381.

Pearl, J. (1999). “Simpson’s paradox: An anatomy”. *UCLA Cognitive Systems Laboratory, Technical Report.*

Penman, S. H. (2009). “Accounting for intangible assets: There is also an income statement”. *Abacus*. 45: 358–371.

Penman, S. H. (2010). “Financial forecasting, risk and valuation: Accounting for the future”. *Abacus*. 46(2): 211–228.

Perezgonzalez, J. D. (2015). “Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing”. *Frontiers in Psychology*. URL: https://doi.org/10.3389/fpsyg.2015.01293.

Pressacco, F. and P. Serafini (2007). “The origins of the mean-variance approach in finance: Revisiting de Finetti 65 years later”. *Journal of Decisions in Economics and Finance*. 30: 19–49.

Raiffa, H. (1968). *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*. Reading, MA: Addison Wesley.

Raiffa, H. (2002). “Decision analysis: A personal account of how it got started and evolved”. *Operations Research*. 50: 179–185.

Rajgopal, S. and M. Venkatachalam (2011). “Financial reporting quality and idiosyncratic return volatility”. *Journal of Accounting and Economics*. 51: 1–20.

Ramsey, F. P. (1926). “The foundations of mathematics”. *Proceedings of the London Mathematical Society*. s2-25: 338–384.

Riedl, E. J. and G. Serafeim (2011). “Information risk and fair values: An examination of equity betas”. *Journal of Accounting Research*. 49: 1083–1122.

Robert, C. P. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. 2nd ed. New York: Springer.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods*. 2nd ed. Springer-Verlag.

Rogers, J. L., D. J. Skinner, and A. Van Buskirk (2009). “Earnings guidance and market uncertainty”. *Journal of Accounting and Economics*. 48: 90–109.

Ross, J. F. (2016). “The information content of accounting reports: an information theory perspective”. *Information*. 7: 48.

Royall, R. (1997). *Statistical Evidence: A Likelihood Paradigm*. London: Chapman and Hall.

Rubinstein, M. (2006). “Bruno de Finetti and mean-variance portfolio selection”. *Journal of Investment Management*. 4: 3–4.

Samson, D., A. Wirth, and J. Rickard (1989). “The value of information from multiple sources of uncertainty in decision analysis”. *European Journal of Operational Research*. 39: 254–260.

Samuelson, P. A. (1965). “Proof that properly anticipated prices fluctuate randomly”. *Industrial Management Review*. 6: 41–50.

Sandroni, A. (2000). “Do markets favor agents able to make accurate predictions?” *Econometrica*. 68: 1303–1342.

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

Savage, L. J. (1962). *The Foundations of Statistical Inference*. London: Methuen.

Savage, L. J. (1971). "Elicitation of personal probabilities and expectations". *Journal of the American Statistical Association*. 66: 783–801.

Schlaifer, R. O. (1959). *Probability and Statistics for Business Decisions: An Introduction to Managerial Economics Under Uncertainty*. New York, NY: McGraw-Hill.

Scott, W. R. (1979). "Scoring rules for probabilistic reporting". *The Journal of Accounting Research*. 17(1): 156–178. (Spring, 1979).

Seidenfeld, T. (1979). *Philosophical Problems of Statistical Inference : Learning from R.A. Fisher*. Boston: D. Reidel.

Seidenfeld, T. (1985). "Calibration, Coherence, and Scoring Rules". *Philosophy of Science*. 52: 274–294.

Shanken, J. (1987). "A Bayesian approach to testing portfolio efficiency". *Journal of Financial Economics*. 19: 195–215.

Smith, K. (2017). "Option prices and disclosure". URL: https://ssrn.com/abstract=3068855.

Stocken, P. C. (2013). "Strategic accounting disclosure". *Foundations and Trends in Accounting*. 7: 197–291.

Strobl, G. (2013). "Earnings manipulation and the cost of capital". *Journal of Accounting Research*. 51: 449–473.

Subramanyam, K. R. (1996). "Uncertain precision and price reactions to information". *The Accounting Review*. 2: 207–220.

Swieringa, R., M. Gibbins, L. Larsson, and J. Sweeney (1976). "Experiments in the heuristics of human information processing". *Journal of Accounting Research* (Supplement). 14: 159–187.

Tang, V. W. (2011). "Isolating the effect of disclosure on information risk". *Journal of Accounting and Economics*. 52: 81–99.

Theil, H. (1969). "On the use of information theory concepts in the analysis of financial statements". *Management Science*. 15: 459–480.

Tobin, J. (1969). "Comment on Borch and Feldstein". *Review of Economic Studies*. 36: 13–14.

Verdi, R. S. (2012). "Discussion of Financial reporting frequency, information asymmetry, and the cost of equity". *Journal of Accounting and Economics*. 54: 150–153.

Veronesi, P. (1999). “Stock market overreactions to bad news in good times: A rational expectations equilibrium model”. *Review of Financial Studies*. 12: 975–1007.

Veronesi, P. (2000). “How does information quality affect stock returns?” *Journal of Finance*. 55: 807–837.

Verrecchia, R. E. (1979). “On the theory of market efficiency”. *Journal of Accounting and Economics*. 1: 77–90.

Verrecchia, R. E. (1982). “The use of mathematical models in financial accounting”. *Journal of Accounting Research*. 20: 1–42.

Verrecchia, R. E. (1983). “Discretionary disclosure”. *Journal of Accounting and Economics*. 5: 179–194.

Verrecchia, R. E. (1990). “Information quality and discretionary disclosure”. *Journal of Accounting and Economics*. 12: 365–380.

Verrecchia, R. E. (2001). “Essays on disclosure”. *Journal of Accounting and Economics*. 32: 97–180.

Wagenhofer, A. (2011). “Towards a theory of accounting regulation: A discussion of the politics of disclosure regulation along the economic cycle”. *Journal of Accounting and Economics*. 52: 228–234.

Wagenhofer, A. (2014). “Trading off costs and benefits of frequent financial reporting”. *Journal of Accounting Research*. 52: 389–401.

Wagenmakers, E.-J., M. Marsman, T. Jamil, A. Ly, J. Verhagen, J. Love, R. Selker, Q. F. Gronau, M. Šmíra, S. Epskamp, D. Matzke, J. N. Rouder, and R. D. Morey (2018). “Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications”. *Psychonomic Bulletin and Review*. 25: 35–57.

Watts, R. (1976). “The capital market, the market for information, and external accounting: Discussion”. *The Journal of Finance*. 31: 677–679.

Winkler, R. L. (1973). “Bayesian models for forecasting future security prices”. *Journal of Financial and Quantitative Analysis*. 8: 387–405.

Winkler, R. L. (1989). “Combining forecasts: A philosophical basis and some current issues”. *International Journal of Forecasting*. 5: 605–609.

Winkler, R. L. (2003). *An Introduction to Bayesian Inference and Decision*. 2nd ed. Sugar Land, Texas: Probabilistic Publishing.

Winkler, R. L. and C. B. Barry (1975). "A Bayesian model for portfolio selection and revision". *The Journal of Finance*. 30: 179–192.

Winkler, R. L., Y. Grushka-Cockayne, K. C. Lichtendahl, and V. R. J. Jose (2019). "Averaging probability forecasts: Back to the future". *Decision Analysis*. Forthcoming.

Yee, K. K. (2006). "Earnings quality and the equity risk premium: A benchmark model". *Contemporary Accounting Research*. 23: 833–877.

Zeff, S. A. (2008). "The contribution of the Harvard Business School to Management Control, 1908–1980". *Journal of Management Accounting Research*. 20(Special Issue): 175–208.

Zellner, A. and V. Chetty (1965). "Prediction and decision problems in regression models from the Bayesian point of view". *Journal of the American Statistical Association*. 60: 608–616.

Zhang, G. (2013). "Accounting standards, cost of capital, resource allocation, and welfare in a large economy". *The Accounting Review*. 88: 1459–1488.

Zhang, X. F. (2006). "Information uncertainty and stock returns". *The Journal of Finance*. 61: 105–137.