The True Cost of Living: 1974-1991

Bruce W. Hamilton[1]

May 19, 1998

I. INTRODUCTION

Engel’s Law states that food’s budget share is inversely related to household real income. As Houthakker (1987) states, “Of all empirical regularities observed in economic data, Engel’s law is probably the best established….(but) like most economic laws, it holds only *ceteris paribus*; prices, among other things, are assumed constant.” Algebraically,

Equation 1

$$\omega_f = g(y/p) + \sum X_i \theta_i$$

where ω_f is food’s budget share, y/p is nominal income deflated by a cost of living deflator, the X_i’s are the *ceteris* to be held *paribus*, and the $_i$’s are their coefficients.

With a micro cross-section/time-series data set such as the Panel Study of Income Dynamics, it is relatively straightforward to observe or estimate everything in equation 1 **except** the cost of living deflator **p**. My strategy is to track temporal movement in the true cost-of-living deflator, **p**, by observing ω_f, y and the vector of X’s, estimating the parameters, and solving for the implied time path of **p.**

As Nakamura (1996) has noted in his controversial critique of the CPI, food’s budget share declined throughout the “stagnant” 1970s, suggesting that income growth was more robust during the ‘70s than suggested by the CPI figures. I find that food’s budget share has declined far too much since 1974 to be explained by CPI-deflated income growth, or by any other regressors.

In spirit, this approach is similar to that of Nordhaus (1996), who in effect created a CPI adjusted to make real income growth consistent with consumers’ perceptions of their financial well-being. Nordhaus measures consumers’ perceptions of their well-being by taking the difference between those who report themselves “better off than last year” and those who report themselves “worse off” in the University of Michigan Survey Research Center’s survey of consumer behavior.

My approach offers several advantages relative to Nordhaus. First, my dependent variable is continuous rather than the crude trichotomous survey variable (“better off than last year,” “worse off,” “about the same”).[2] Second, actions do speak louder than words; under my approach a consumer must put his money where his mouth is, not just respond to a survey. Third, as I will show, the Krueger/Siskind (1998) criticisms of Nordhaus (shifts in the income distribution and life-cycle effects) are very easily handled.

My approach also has some kinship with the extensive work of Jorgenson and Slesnick (1997), who estimate a complete system of demand functions, and then use information on prices and the demand-system coefficients to back out annual costs and standards of living. They aggregate individual-good prices into 5 broad-category translog price indices, which constitute the price data for their demand system. For purposes of estimating CPI bias, the main difficulty with their approach is that they must assume that the individual-good prices are reported correctly, and that they properly capture the array of available goods. Whereas their method elegantly handles substitution bias, it has no way of handling such problems as new-product bias or unmeasured quality improvement.

Most CPI research takes the brute-force approach of examining the various components of the index for specific instances of bias, and attempting to determine the magnitude of these biases (see Moulton (1996), Shapiro and Wilcox (1996), Boskin, Dulberger, Gordon, Griliches and Jorgenson (1996)). One problem with this approach is that it is impossible to know when we are “finished;” the most we can ever say is that we have corrected all of the biases that we have found. A related problem is that for many important adjustments it is extremely difficult to quantify the required adjustment; much of the discussion, even in a study as careful as the Boskin report, takes the form of (highly educated) guesstimates.

To take one example, Gordon (1990) presented extremely careful and painstaking estimates of the true cost of automobile services and made several important discoveries including evidence of serious bias in the CPI for used cars. But even after this careful work he neglected one important component of the cost of auto services: the longevity of cars themselves. The life expectancy of 1960 cars was just over 9 years; that of 1979 vintages was over 11 years. Over this timespan both Gordon and the CPI missed a substantial source of decline in the cost of automobile services.

---

[1] Johns Hopkins University. I thank Carl Christ, Jennifer Hunt, Robert Moffitt, and Matthew Shapiro for helpful comments on earlier drafts.

[2] See also Krueger and Siskind (1998), who do not find significant bias after applying some refinements to the Nordhaus approach.

Another problem with the standard approach is that it is quite difficult to examine the magnitude of the bias historically; frequently the data required to correct CPI components are available only for the relatively recent past, so it is difficult to learn how significant the bias has been in the more remote past.

Whatever new problems may be introduced by the Engel-curve method of estimating the cost of living, it has the virtue of bypassing a host of virtually insurmountable problems. To get a feel for the usefulness of food as an indicator of well-being, turn to Figures 1 and 2. Data for both figures are national aggregates from the National Income Accounts.[3] Treating food's share as an indicator for standard of living, one would conclude that there was a period of stagnation from the late 1960s through the mid-'70's, but that for the 20 years since then real income growth has been about what it was in the prior two decades.

Figure 1

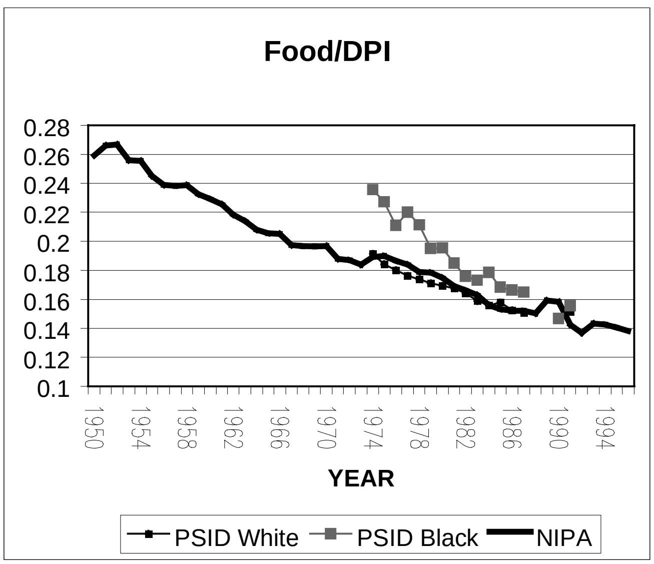


Figure 2

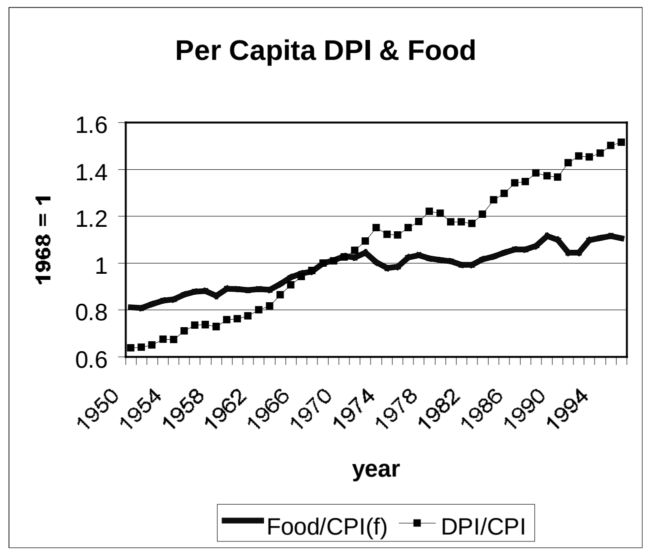


Figure 2 shows the components of this ratio. Per capita food consumption (food expenditure deflated by the food component of the CPI)[4] rose about 25%, almost linearly, from 1950 through 1972. After drifting down, then up and then down again, per-capita food consumption in 1992 was back at

[3] Figure 1 also shows the average value of food's budget share (at home plus away from home) respectively for the black and white PSID samples upon which I base my estimation below. Whereas my estimates are based on food at home, here I show food at home and away from home to show the comparability with the NIPA numbers.

approximately its 1972 level, despite a 30% rise in per-capita CPI-deflated DPI from 1972 to 1992 and a 10% decline in the relative CPI of food.[5] Figure 2 appears to be internally inconsistent, and also inconsistent with figure 1 (which suggests that real income growth after 1975 may have been comparable to the growth pattern from 1950 to 1970). Nevertheless, one could imagine a variety of explanations for the apparent inconsistency between these figures. It could have been caused by an unmeasured decline in the relative price of food,[6] the widening of the income distribution, the decline in family size, increasing female labor force participation or a change in the volatility of income, for example.

The first purpose of this paper is to utilize the PSID to see whether the anomalies of Figures 1 and 2 can be attributed to some non-CPI cause such as demographics or changes in the distribution of income. The second purpose is to offer a more refined estimate of CPI bias. Third, I will present evidence of strikingly different inflation rates by race.

Using the PSID, I estimate a demand function for food at home[7] for 1974[8] through 1991[9]. Using a standard measure of real income (total family income after federal taxes, the PSID's best continuously available approximation of disposable income)[10] deflated by the CPI, this demand function has shown consistent drift over the sample period; I attribute this drift to unmeasured growth in real income, and in turn I attribute the mismeasurement of income to CPI bias.

---

[4] Throughout this paper, when referring to "the CPI," I will mean the CPI-U and its components.

[5] If income and price elasticities are respectively ± 0.5, measured changes in income and prices should have resulted in approximately a 20% increase in food consumption.

[6] I have already accounted for the measured decline, by deflating food expenditure by the food CPI.

[7] Food away from home would be interesting as well. However, in some years it is reported only by fairly broad band. And for many consumers it is zero. Below I discuss the role of food away from home.

[8] Inasmuch as the PSID began in 1968, it is unfortunately not possible to extend this technique back to the early 1960s or the 1950s. In addition, the food data appear to be fairly erratic from 1968-72 (and not reported for 1973). Thus I must begin my analysis in 1974.

[9] When I began this research, the last wave of the PSID available in final form was 1992, which reports 1991 income. Thus 1991 is the last year for which I can calculate my dependent variable, food's share of income. Even as more data become available, it will be difficult to extend this work forward due to a change in PSID coverage (see footnote 10 below).

[10] In the 1992 wave, PSID does not report federal income taxes (for 1991 income). To estimate federal taxes for 1991 I regress 1990 taxes on income, income squared, family size and size of home mortgage. I then assume that the same relationship (R^2 = .988) explains the relationship between taxes and the other regressors for 1991, with 1991 income deflated by CPI growth to account for indexing of the tax code. Unfortunately, this extrapolation becomes progressively worse as it is extended forward in time; thus even when the data become available it will be difficult to extend this method forward in time.

In a nutshell, the results are as follows: On average, in 1974 the PSID sample[11] of white households spent 16.64% of its income on at-home food. By 1991 this share had fallen to 12.04%. Measured per-household income grew 7% over this time span, explaining just over half a point of the food share decline. Decline in the relative CPI of food is sufficient to explain perhaps as much as 1 percentage point of decline in food's share. Other regressors accounts for less than 0.1 point of additional decline; thus about 3 points of the food-share decline are left to be explained by CPI bias. I estimate that this bias is about 2.5% per year from 1974 through 1981, and slightly under 1% per year since then.

For blacks, food's share fell from 21.17% to 12.44%. Approximately .8 point of the decline can be explained by measured income growth, and another point by movement in other regressors, and up to another 1 point by the decline in the food CPI. Thus the food-share decline left to be explained by measurement error is 5.9 points. I estimate the bias to be approximately 4% per year from 1974 through 1981 and about 3% per year since then.

II. ESTIMATION APPROACH

In this section I describe the empirical method that I will use to estimate CPI bias. As stated, the fundamental approach is to infer well-being by observing food's share.

a. Why Food?

Food is the only consumption item regularly tracked in the PSID, and the PSID is the only micro data set with good income data which covers a sufficient time span to properly carry out this exercise. But in fact food was a fortuitous choice; it is perhaps the ideal indicator good for inferring inflation.

First, any indicator good must have an income elasticity substantially different from unity; otherwise the budget share is insensitive to income and thus to mismeasurement of income.

Second, almost unique among consumer goods, food has no durability whatsoever; food expenditure is virtually identical to food consumption. There is no stock/flow issue to confound measurement of consumption.

[11] Actually this a subset of the PSID data base, in which I have eliminated some outliers and other pathological cases. Below I will discuss the selection criteria.

Third, food is a pretty straightforward good. Other candidate indicator goods, such as recreation (see Costa (1997)), involve tricky definitional problems over and above the issue of durability. (How, for example, do we count books?)

Fourth, it is fairly natural to assume, as I will, that food is strongly separable from nonfood in consumers' utility functions. With separability comes the result that CPI bias in such prime suspects as personal computers will not affect food's budget share through some peculiar (unmodeled) complementarity or substitutability, but only through the channels anticipated in the model.

b. Basic Estimating Structure

I begin with the basic demand structure of Working (1943) and Leser (1963):

Equation 2

$$\omega_{i,j,t} = \phi + \gamma\left(\ln P_{f,j,t} - \ln P_{n,j,t}\right) + \beta\left(\ln Y_{i,j,t} - \ln P_{j,t}\right) + \sum_x \theta_x \cdot X_{i,j,t} + \mu_{i,j,t}$$

where

$\omega_{i,j,t}$	Food's share of family income for person i in SMSA j in year t
$P_{f,j,t}$	True but unobservable price of food in SMSA j in year t
$P_{n,j,t}$	True but unobservable price of nonfood in SMSA j in year t
$P_{j,t}$	True but unobservable general price level in SMSA j in year t
$Y_{i,j,t}$	Consumer i's nominal income in year t
$X_{i,j,t}$	Background variable vector for person I in SMSA j in year t

The true cost of living, $P_{j,t}$, is a weighted average of the prices of food and nonfood[12]:

Equation 3

$$\ln P_{j,t} = \alpha \ln P_{f,j,t} + (1-\alpha)\ln P_{n,j,t}$$

P_f , P_n , and P are measured with (CPI-bias) error:

[12] I take averages of logs, not of absolute levels. In part this is for algebraic simplicity down the road, but it is justified in that a cost-of-living index which is a weighted average of logs reflects elasticity of substitution equal to unity, rather than the zero elasticity of substitution assumed in a fixed-weight index.

Equation 4

$$\ln\left(P_{g,j,t}\right)=\ln\left(P_{g,j,0}\right)+\ln\left(1+\Pi_{g,j,t}\right)+\ln\left(1+\mathrm{E}_{g,j,t}\right);\quad g=f,n,\_$$

where

$P_{g,j,0}$	Unobservable true price of g (f,n, or _) in SMSA j in year 0
$\Pi_{g,j,t}$	Cumulative percent increase in the (CPI) measured price of g in SMSA j from year 0 to year t
$E_{g,t}$	Year-t percent measurement error in cumulative good-g inflation (assumed constant across SMSA's)

Now, to save the carrying of log notation, I adopt the following:

$\ln P_g$	p_g
$\ln Y$	y
$\ln (1+\Pi_g)$	π_g
$\ln (1+E_g)$	ε_g

Throughout the paper, **P** and **p**, with appropriate subscripts, refer to true but unobservable prices; $\boldsymbol{\Pi_t}$ and $\boldsymbol{\pi_t}$ refer to CPI-measured cumulative inflation until year t, and $\mathbf{E_t}$ and $\boldsymbol{\varepsilon_t}$ with appropriate product subscripts, refer to cumulative CPI error until year t. Combining,

Equation 5

$$\varepsilon_t=\alpha\cdot\varepsilon_{f,t}+(1-\alpha)\cdot\varepsilon_{n,t}$$

Now substitute into equation (2), and collect terms respectively involving ε and p_0:

Equation 6

$$\begin{aligned}&\omega_{i,j,t}=\phi+\gamma\left(\pi_{f,j,t}-\pi_{n,j,t}\right)+\beta\left(y_{i,j,t}-\pi_{j,t}\right)+\sum_X\theta_X X_{i,j,t}\\&+\gamma\cdot\left(\varepsilon_{f,t}-\varepsilon_{n,t}\right)-\beta\cdot\varepsilon_t\\&+\gamma\left(p_{f,j,0}-p_{n,j,0}\right)-\beta\left(p_{n,j,0}\right)+\mu_{i,j,t}\end{aligned}$$

where (6) assumes (as I will) that CPI bias does not vary geographically. Now suppose we have a cross-section/time-series database (the PSID) with micro data on income and food expenditure (as well as other variables such as family composition which would influence food expenditure), as well as cross-section CPI for all consumption, for food, and for nonfood, over the entire data period for a sample of SMSA's. Empirical specification is straightforward:

Equation 7

$$\omega_{i,j,t} = \phi + \gamma\left(\pi_{f,j,t} - \pi_{n,j,t}\right) + \beta\left(y_{i,j,t} - \pi_{j,t}\right) + \sum_X \theta_X X_{i,j,t}$$
$$+ \sum_{t=1}^{T} \delta_t \cdot D_t + \sum_j \delta_j \cdot D_j + \mu_{i,j,t}$$

Although $p_{f,j,t}$ and $p_{n,j,t}$ are measured with error, this is not a source of errors-in-variables bias in the demand coefficients γ, β and θ. Each year's CPI error itself is perfectly correlated with the corresponding year dummy; thus CPI bias does not bias our estimates of the demand coefficients.

c. Geographic Price Variation

The relative CPI of food declined about 10% over the sample period; the first term on the right-hand side of (7) accounts for the effect of this (measured) price decline upon ω. Were we to omit the price term, that part of the decline in ω due to the price decline would be improperly ascribed to CPI bias. Thus identification of CPI bias requires inclusion of (π_f - π_n) and estimation of the price coefficient, γ. But γ can only be estimated if the relative price varies independently of the other regressors. Clearly, time-series variation using national CPI data will not do; any movement in the relative price is perfectly correlated with the year dummies. The price must vary must vary over both time and space if γ is to be identified.

BLS does not report cross-section data on price levels; SMSA CPI's (reported for 25 SMSA's) are measured only relative to same-city CPI's in the base year. Thus, with SMSA specific CPI data we can calculate relative (measured) inflation rates across cities, but we cannot compare price levels. So long as there is sufficient geographic/temporal variation in measured relative inflation rates this may be sufficient to identify γ. Here I describe the estimation technique employed with the geographic-variation sample; below I describe the compromises which must be made when we rely on the larger national data base.

Equations (6) and (7) presume data on geographic price variation. Interpretation of dummy coefficients is clean and informative:

Equation 8

$$\delta_t \quad = \quad \gamma \cdot \left(\varepsilon_{f,t} - \varepsilon_{n,t}\right) - \beta \cdot \varepsilon_t$$

Equation 9

$$\delta_j \quad = \quad \gamma \cdot \left(p_{f,j,0} - p_{n,j,0}\right) - \beta \cdot p_{j,0}$$

Next assume that the relative bias as between food and nonfood is constant across all years:

Equation 10

$$\varepsilon_{f,t} = r \cdot \varepsilon_{n,t} \qquad \forall t$$

where r is the (unknown) ratio of ε_f to ε_n. The parameter estimates from (7) identify the CPI bias, ε_t, up to the unknown ratio **r**:

Equation 11

$$\delta_t = \varepsilon_t \cdot \left(-\beta - \frac{\gamma(1-r)}{1-\alpha(1-r)} \right)$$

where α is food-price's share in the cost of living index. If r = 1 (food and nonfood are equally biased)

Equation 12

$$\varepsilon_t = \frac{\delta_t}{\beta}$$

More generally, Equation (12) is approximately correct so long as either γ or (1-r) is close to zero. If r < 1 as seems plausible (food is less badly biased than nonfood), then (12) understates the bias, increasingly as r falls increasingly below unity. Thus we will argue below that (12) gives an approximate lower bound for the CPI bias.

The SMSA dummy coefficient, given in equation (9), is a weighted average of the SMSA's base-period general price level and its relative price of food. We will discuss interpretation of these coefficients in the empirical section below.

d. Current vs Permanent Income

Ever since Friedman (1957) economists have understood that consumption patterns are best explained by permanent, not current income. If income is measured with error (either simple measurement error or failure to properly capture permanency) then the estimate of β is biased, and (by equation (12)) this imparts bias to the estimated CPI error[13].

Qualitatively, the concern over permanent income is not a major issue. Food's budget share declined by approximately 25% over the sample period; if little of this decline can be statistically explained by movements in current income or other regressors, significant bias in the CPI is a very likely candidate for explaining the movement. But our estimate of the magnitude of the CPI bias may be estimated with error if we do not properly capture the effects of permanent income.

In this section I introduce two sets of regressors which will mitigate the problems associated with the non-permanency of current income. The first, which will appear in all specifications, is a set of income-growth variables:

Equation 13

$$\begin{aligned} \dot{y}_{t,+} &= y_t - y_{t-1} \quad \textit{if} \quad y_t - y_{t-1} > 0 \quad ; \\ &= 0 \quad \textit{otherwise} \\ \dot{y}_{t,-} &= y_t - y_{t-1} \quad \textit{if} \quad y_t - y_{t-1} < 0 \quad ; \\ &= 0 \quad \textit{otherwise} \end{aligned}$$

where y_t is income after taxes, deflated by the CPI.[14] I allow the growth coefficients to be different depending on whether the growth is positive or negative, to address the possibility that consumers respond differently to a rise than to a fall in income.

With the income-growth variables in the regression, the current-income coefficient relates food's share to current income, holding the one-year growth of current income constant. If these regressors

[13] The direction of bias, both to β and the ultimate CPI-bias estimate, is unclear, since (mismeasured) income appears on both sides of the equation. β is likely biased toward –1 (the elasticity of demand biased toward 0). I thank Matthew Shapiro for clarifying this bias discussion.

[14] In another specification, not reported, I included two-year lags as well; these were uniformly insignificant.

properly complete the relationship between current and permanent income, then we can interpret the current-income coefficient in the new regression as a permanent-income coefficient.

Second, I replace current CPI-deflated income with the household's average value of this variable over the past three years. I run each specification below respectively with current income and with this permanent-income measure. I report results for this permanent-income regression only for one specification.

III DATA

In this section I discuss the formation of my data sets from the PSID panel data set.

a. Selection Criteria

I begin with the entire 1968-1992 PSID family file, and eliminate cases for a variety of reasons. The rationale for elimination takes two forms:

- Good reason to believe that something is wrong with the data;
- Reason to suspect that the pattern of food demand might be unusual for the excluded category.

Exclusion criteria are as follows:

- The PSID does not report food consumption for 1973, 1988, or 1989; thus I exclude these years. I also exclude 1968-1972 because the food data appear to be unreliable. PSID for those years reports food at home, food away from home, and total food. The first two series do not add up to the third. In addition, food-at-home's share of income shows an extraordinarily and implausibly sharp decline from 1968 through 1972, from about 25% of income to about 15%, for whites; and a similar but larger fall for blacks.
- The PSID reports food consumption for the current year but income for last year. Thus the dependent variable $\omega_{i,j,t} = \frac{food_{i,j,t}}{income_{i,j,t}}$ cannot be constructed for the most recent year in the PSID. For this reason I am unable to use 1992 (except inasmuch as it reports 1991 income).
- All households with one, or 3 or more, major adults;
- The poverty sample and all households receiving AFDC or food stamps.
- Families for whom CPI-deflated family after-tax money income is less than $150, not reported or top-coded.

- Families for whom husband and wife federal income taxes were not reported
- Families who spent more than 80%, or less than 2%, of income on food
- Families which reported changes in composition over the previous year.
- Families for which either the head or wife was less than 21 years old.

b. Geographic and National Samples

I work with two data sets. The first is the entire national data set for the PSID, for blacks and whites (other races are excluded). The virtue of this data set (relative to the second one, described below) is size; there are approximately 1800 valid observations per year for whites and just under 100 for blacks.

The second data set includes only observations from SMSA's for which BLS reports local Consumer Price Indices. Tables 1 and 2 gives summary statistics for the 1974 and 1991 components of the sample respectively, for blacks and for whites, for both samples.

Table 1
Summary Statistics of Full PSID Sample:
1974 and 1991

	WHITE 1974	WHITE 1991	BLACK 1974	BLACK 1991
Number	1450	1901	71	85
"Real" After-tax Income[15]	10411	11120	8767	9243
# Children	1.15	1.03	1.32	1.26
County Unemployment	5.51	5.57	4.96	5.37
Husband's Hrs Work	1984	1902	2012	1841
Wife's Hrs Work[16]	877	1150	1161	1350
Husband's Education	12.27	13.21	10.04	12.12
Wife's Education	12.00	13.04	10.77	12.65
Husband's Age	44.83	45.09	44.45	43.98
Wife's Age	42.01	42.80	40.22	41.04
Fd Share @ Home	.1664	.1204	.2117	.1244
Fd Share Restaurant	.025	.031	.024	.031

[15] "Real" family income is family money income less husband-and-wife federal income taxes, deflated by the ratio of the current Consumer Price Index to the 1968 CPI.
[16] PSID reports wife's average weekly hours; I have multiplied by 50.

Table 2
Summary Statistics of Geographic-Variation PSID Sample
1974 and 1991

	WHITE 1974	WHITE 1991	BLACK 1974	BLACK 1991
Number	409	403	31	22
"Real" After-tax Income[17]	12008	12784	9571	11065
# Children	1.05	.92	1.35	1.13
County Unemployment	6.43	5.59	5.49	6.17
Husband's Hrs Work	1957	1863	1894	1905
Wife's Hrs Work[18]	855	1044	1280	1512
Husband's Education	13.07	13.83	10.51	12.79
Wife's Education	12.27	13.43	11.10	13.04
Husband's Age	44.10	45.69	42.76	43.83
Wife's Age	41.68	43.28	38.49	41.54
Food Price[19]	1.10 (.015)	.972 (.039)	1.10 (.018)	.994 (.048)
# SMSA's	23	25	23	25
Fd Share @ Home	.1572	.1196	.2128	.1181
Fd Share Restaurant	.027	.030	.030	.033

Unfortunately the second data set is extremely small, particularly for blacks. The black data set yields approximately one observation per SMSA per year. I restrict my attention to whites when working with the smaller geographic-variation sample of Table 2.

IV RESULTS

a. Geographic Price Variation

Tables 3 and 4 below give the regression results based on the geographic-variation sample for whites:

[17] "Real" family income is family money income less husband-and-wife federal income taxes, deflated by the ratio of the current CPI to the 1968 CPI. To convert to 1996 dollars, multiply by 4.51.
[18] PSID reports wife's average weekly hours; I have multiplied by 50.
[19] Standard Deviation in parentheses.

Table 3
Background Coefficients
Whites: Geographic Sample
(t-ratios in parentheses)

Variable	Coefficient	t-Statistic
Constant	.980	(29.98)
Age of Head	.0008	(4.88)
Age of Wife	.0002	(1.36)
# of Children	.0212	(32.06)
County Unemp	-.00002	(-.059)
Hd ann hrs x10^3	.0064	(6.84)
Wife ann hrs x 10^3	-.0013	(-1.66)
Head education	-.00089	(-2.78)
Wife education	-.0011	(-3.07)
ln Income	-.101	(-53.80)
ln Rel Food Price	.0369	(1.46)
Income Growth +	.0022	(1.17)
Income Growth -	-.025	(-9.28)
Fdshare-rest	.063	(5.13)
Adj R^2		.538

Table 4
Time and SMSA Coefficients
Whites: Geographic Sample
(t-ratios in parentheses)

YEAR	COEFFICIENT	CUMUL. BIAS ESTIMATE	SMSA	COEFFICIENT	P_0*
1975	-.0055 (-1.48)	.053 (-1.43)[20]	New York	.0293 (6.588)	1.49
1976	-.0086 (-2.34)	.082 (-2.29)	Miami	.0215 (3.91)	1.38
1977	-.0097 (-2.66)	.092 (-2.61)	Los Angeles	.0183 (3.881)	1.34
1978	-.0142 (-4.00)	.131 (-3.94)	Buffalo	.0182 (1.347)	1.34
1979	-.0169 (-4.71)	.154 (-4.65)	San Francisco	.0178 (1.929)	1.33
1980	-.0218 (-5.51)	.194 (-5.43)	Portland, OR	.0173 (1.61)	1.32
1981	-.0245 (-5.66)	.215 (-5.58)	Chicago	.0134 (2.411)	1.27
1982	-.0241 (-5.15)	.212 (-5.08)	Cincinnati	.0130 (-1.28)	1.27
1983	-.0274 (-5.50)	.238 (-5.44)	Houston	.0117 (1.784)	1.25
1984	-.0277 (-5.63)	.240 (-5.56)	San Diego	.0094 (.690)	1.22
1985	-.0279 (-5.34)	.241 (-5.28)	Washington, D. C.	.0083 (1.127)	1.21
1986	-.0295 (-6.11)	.253 (-6.04)	Philadelphia	.0080 (.997)	1.21
1987	-.0291 (-5.88)	.250 (-5.81)	Milwaukee	.0072 (.219)	1.20
1988			Detroit	.0069 (1.049)	1.19
1989			Boston	.0063 (.831)	1.19
1990	-.0303 (-6.20)	.259 (-6.12)	Baltimore	.0056 (.876)	1.18
1991	-.0312 (-6.31)	.266 (-6.23)	Denver	.0055 (2.297)	1.18
			Pittsburgh	.0032 (.446)	1.15
			Kansas City	.0019 (.472)	1.14
			Seattle	.0006 (-.724)	1.12
			Atlanta	0 (omitted)	1.12
			St. Louis	-.0020 (-.825)	1.09
			Cleveland	-.0041 (-1.32)	1.07
			Dallas	-.0110 (-1.77)	1
			Minneapolis St. Paul	-.0110 (-2.90)	1

Below I discuss the interpretation of the time dummies, and the income and price coefficients (which are necessary to map the time dummies into estimates of CPI bias). Discussion of all other coefficients is postponed until the end of the empirical section.

i. Income and Price Coefficients

The income and price coefficients imply approximate income and price elasticities[21] of respectively +.33 and -.65. Not surprisingly, given the very small variation in measured relative price of

[20] T-statistics estimated for ε_t as follows: First run equation (7) via OLS. Solve for ε_t as described above. With these estimated coefficients as initial points, run equation (6) via NLLS (with (ε_f - ε_n) suppressed, incorporating the maintained hypothesis that r = 1). SPSS generates point estimates and standard errors for ε_t.
[21] Formulas for income and price elasticities are respectively

food, the price coefficient is not estimated precisely. And the implied price elasticity seems high. In work below, when I need an assumed value for γ, I will assume alternatively the .037 point estimate from table 3 and .07; the latter implies a price elasticity of approximately .45.[22]

To some readers, the income elasticity looks low; it is much lower, for example, than that estimated by Tobin (1950).[23] Two points are in order: First, in that specification below in which I report results based on a 3-year average of income, the elasticity rises to about 0.4. Second, my estimated elasticity is for food at home, which is demanded more inelastically than total food.[24]

ii. Interpretation of Year Coefficients

If the assumptions underlying equation (11) are satisfied, then the cumulative CPI bias for each year is found by dividing the year coefficient by the income coefficient. The estimated cumulative percent bias over the entire data period is

Equation 14

$$1 - EXP(\varepsilon_{74-91}) = 1 - EXP\left(\frac{d91}{\beta}\right) \quad = \quad -.266$$

Annual bias is easily found by taking the difference between successive cumulative bias estimates. Figure 3 gives point estimates for cumulative bias, from 1975 through 1991. In most but not all cases the difference between successive-year dummy coefficients is significant. Generally two-year differences are significant. The figure tells a striking story: The average annual bias from 1974 through 1981 is 3.54%; for

---

$$\eta_y = 1 + \beta/\omega$$

$$\eta_p = -1 + (\gamma - \alpha\beta)/\omega$$

where α is the share of the food price in the cost of living index.

[22] The income coefficient is fairly close to that of Deaton and Muellbauer (1980), but the price coefficient is substantially lower than theirs.

[23] I thank Matthew Shapiro for bringing Tobin's paper to my attention, and for sharpening my own approach to permanent income.

[24] If 20% of food is at restaurants (about right for 1991), and restaurant-food demand is unit elastic, the overall elasticity of demand for food rises from 0.4 to 0.52 (or from 0.33 to 0.36).

the period 1982-1991, average annual bias is 0.67%.[25] If these estimates are even close to correct, the productivity slowdown of the 1970s is a chimera.

Recall that this estimate is valid only up to the unknown value of **r**, the ratio of the bias in food to nonfood. If food is less badly biased than nonfood, equation (14) understates the CPI bias.

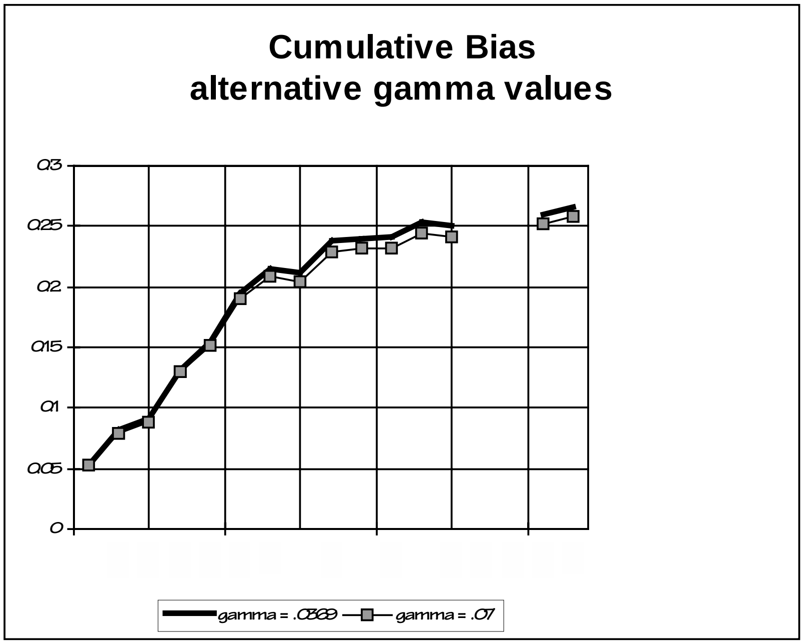
iii. A Caveat: The Relative CPI of Food

The relative CPI of food fell by about 13% from 1975 through 1985, and then rose by about 3% to 1990. By itself, this accounts for a decline in food's share from 1975 to 1985, and a rise thereafter. In principle, this price movement has been accounted for by the inclusion of the relative CPI of food in the regression. But if the price coefficient is biased, then so is the portion of temporal movement in food's share which the regression attributes to intertemporal movements in the price of food. Thus the portion left to be explained by CPI bias is also biased.

If the true value of γ is higher than the estimated value of .0369, then part of the decline in food's share from 1975 through 1985 which we attribute to CPI bias is really due to the decline in the measured price of food, and our estimate of CPI bias for this period is biased upward. By contrast, our estimate of bias since 1985 is biased downward. The second series in Figure 3 shows cumulative bias under the alternative assumption that $\gamma = .07$. As can be seen, the effect is quite modest, and the overall story is essentially unchanged: annual bias is around 3% from 1975 through 1981 and just under 1% since.

[25] The diminution of bias after 1981 appears in all specifications and for both blacks and whites. I will discuss possible reasons for this diminution at the end of the empirical section.

Figure 3



b. The National Sample: No Price-Variation Data

The limited size of the geographic sample, dictated by the limited coverage of local CPI deflators, precludes four important lines of inquiry regarding Engel curves and their implications for CPI bias:

- Functional form: Perhaps the true relationship between food's share and log income is nonlinear; if so it is possible that the series of year dummies is really picking up unmeasured nonlinearity rather than unmeasured inflation. As we will see, the estimating equation required to explore this question requires several hundred dummy variables under the geographic-variation regime.
- Race: I wish to explore the possibility that CPI bias is different for blacks than for whites. My basic hypothesis (confirmed below) is that the true cost of living has grown more slowly for blacks than for whites over the sample period, reflecting a gradual erosion in general price discrimination against blacks. This issue cannot be meaningfully explored with the geographic sample, because (see Table 2) there are so few black observations per year.

- Age: Many people are concerned about the possibility that the true cost of living grows at a different rate for the elderly than the prime-aged population. With a sufficient sample size (i.e., the national sample) I can simply estimate Engel-curve equation separately for the prime-age population and the over-62 population. I have carried out this exercise but to not report the results. There is no material difference between the estimated inflation rates for prime-age and old people.
- Fixed effects: If any omitted household-specific fixed effects are correlated with the regressors, then OLS regression coefficients are biased. As is well known, fixed-effect biases are eliminated by taking first differences. Given the importance of income growth, we need three successive years of data on a household (rather than the usual two) in order to construct an observation. Particularly in light of my selection criterion requiring no change in family composition, this substantially reduces the number of observations. In order to retain an adequate sample size, I elected to estimate the first-difference version of equation (7) only with the white subset of the national sample . In fact, both the background coefficients and the year dummies were virtually identical to the OLS coefficients, so I do not report fixed-effects results.

To explore these questions I use the entire PSID sample (subject to the criteria described above), and estimate equation (7) and appropriate variants, omitting the relative price term and the SMSA dummies. I refer to equation (7) but without the price term or the SMSA dummies as equation (7a).

i. Basics

The form without the food price or the SMSA dummies is misspecified. A simple if inelegant test of the importance of specification bias is to run equation (7a) – the one without price and SMSA dummies – over the geographic sample, and to compare the results with those of equation (7). When I do this[26], the adjusted R^2 falls from .538 to .501. Both the background and time-dummy coefficients (corrected for omission of time-series variation in the relative measured price of food, see below) are very close to the equation-(7) specifications.

By omitting the food-price term from (7a) , we eliminate not only cross-section variation in the relative price of food, but time-series variation as well. The year dummy coefficients pick up not only the CPI bias of equation (8) but also the effect of intertemporal variation in the relative (CPI-measured, but omitted because of perfect correlation with the year dummies) price of food. Proper interpretation of the year dummy coefficient is given by equations (15) and (16). CPI bias is found by subtracting $\gamma(\pi_{f,t} - \pi_{n,t})$ from the dummy coefficient, and then dividing by β (assuming, as discussed above, that $\gamma(1\text{-}r)$ is close to zero). As the relative CPI of food declined by about 10% from 1974 through 1991, part of the decline in ω_f is due to the (measured) food-price decline and not to any unmeasured rise in income

Equation 15

$$\delta_t = \varepsilon_t \cdot \left(-\beta - \frac{\gamma(1-r)}{1-\alpha(1-r)} \right) \qquad \left\{ + \gamma \cdot \left(\pi_{f,t} - \pi_{n,t} \right) \right\}$$

Equation 16

$$\varepsilon_t = -\frac{\delta_t \qquad \left\{ - \gamma \cdot \left(\pi_{f,t} - \pi_{n,t} \right) \right\}}{\beta}$$

ii. Log Linear Results by Race

There is a substantial literature suggesting that blacks' economic difficulties extend beyond their inability to earn income at the same rate as whites. Unfortunately, there is no systematic evidence on the magnitude of this discrimination, nor what has happened to this magnitude over the past two decades. In this section I hope to fill this void.

The most obvious "direct" method of estimating the cost-of-living premium faced by blacks would be to reproduce BLS's CPI analysis separately for blacks. Even at a point in time the task would be enormous; there is no hope of using this approach to track any trend in price discrimination. And even if it could be done up to BLS's standards, it would still suffer from CPI bias.

The approach developed above is much more parsimonious; it infers movement in the true cost of living based on food consumption. With adequate data, this method can isolate inter-racial differences in the growth rate of the cost of living. To do this, I estimate equation (7a) separately for blacks and whites.[27]

[26] These results are not shown, but are available on request.
[27] As noted above, the geographic-variation sample has far too few black observations (approximately one per city per year).

Table 5 gives the estimated coefficients for the two races, as well as a separate set of results for whites, in which current CPI-deflated income is replaced by its (past) three-year average.

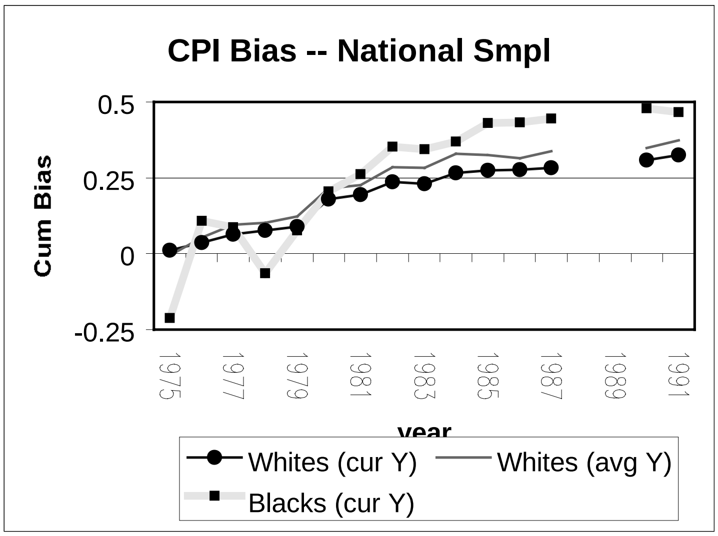
Table 5
Regression Coefficients by Race
Equation 7a: National Sample
(t-ratios in parentheses)

	WHITES	**WHITES 3 YR AVG INC**	**BLACKS**
VARIABLE	COEFFICIENT	COEFFICIENT	COEFFICIENT
Constant	.997 (123.68)	.906 (128.78)	1.028 (23.41)
Age of Head	.0005 (5.67)	.00023 (6.46)	.0007 (1.86)
Age of Wife	.0002 (2.44)	.00068 (20.95)	-.0001 (-0.25)
# of Children	.020 (54.75)	.022 (56.83)	.016 (9.01)
County Unemp	.0004 (3.07)	.00051 (3.61)	.0013 (1.55)
Head ann hrs x10^3	.0052 (10.58)	.003 (5.23)	.0034 (1.20)
Wife ann hrs x10^3	-.0017 (-4.08)	-.0035 (-3.10)	-.0054 (-2.12)
Head education	-.00004 (-.22)	-.0014 (-9.54)	.0016 (1.99)
Wife education	-.0013 (-6.24)	.00084 (6.04)	-.0054 (-4.97)
Ln Income	-.097 (-101.84)	-.089 (-106.8)	-.096 (-16.88)
Inc Growth +	.0014 (1.60)	-.026 (-15.8)	-.010 (-1.10)
Inc Growth -	-.056 (-28.30)	-.125 (-65.45)	-.082 (-7.49)
Fdshare-out	.067 (5.26)	.043 (4.14)	.060 (.798)
1975	-.00561 (-2.57)	-.00404 (-1.88)	.0147 (1.13)
1976	-.00668 (-3.10)	-.00802 (-3.78)	-.0141 (-1.12)
1977	-.00951 (-4.48)	-.0121 (-5.780	-.0118 (-0.960)
1978	-.0118 (-5.60)	-.0137 (-6.60)	.0018 (0.149)
1979	-.0131 (-6.27)	-.0156 (-7.59)	-.0116 (-.948)
1980	-.0210 (-10.07)	-.0235 (-11.45)	-.0238 (-1.95)
1981	-.0217 (-10.02)	-.0235 (-11.02)	-.0299 (-2.32)
1982	-.0260 (-11.93)	-.0297 (-13.80)	-.0415 (-3.20)
1983	-.0246 (-11.15)	-.0289 (-13.22)	-.0397 (-3.04)
1984	-.0291 (-13.63)	-.0345 (-16.27)	-.0432 (-3.33)
1985	-.0297 (-13.78)	-.0334 (-15.57)	-.0526 (-4.07)
1986	-.0306 (-14.61)	-.0327 (-15.71)	-.0534 (-4.18)
1987	-.0316 (-15.39)	-.0359 (-17.84)	-.0559 (-4.50)
1990	-.0356 (-17.14)	-.0380 (-18.87)	-.0627 (-5.00)
1991	-.0374 (-18.27)	-.0409 (-20.60)	-.0596 (-4.82)
Adj R^2	.504	.475	.512

First, note the estimates based respectively on current and 3-year average (CPI-deflated) income respectively, for whites. With three-year average income, the income elasticity rises from about .33 to about .4 as would be expected. Most of the background coefficients show little change. The estimated dummy coefficients are somewhat larger and the estimated CPI biases are also a bit higher. (Cumulative estimated CPI bias from 1974 through 1991 is 19% greater under the 3-year-average than under the current-income specification.) The fit is somewhat worse under the 3-year-average specification; on this criterion I discuss only the results based on current income.

Figure 4 shows estimated cumulative CPI bias respectively for whites and blacks, calculated according to equation (14) (assuming $\gamma = .0369$ as estimated in table 3). Average annual bias for whites is 2.75% from 1974-1981 and 1.51% thereafter; for blacks it is 4% from 1974-1981 and 2.99% thereafter. Over the complete sample period the black cost of living premium over whites has declined by 14 percentage points.

Figure 4



King and Mieszkowski (1971) find about a 10% black housing-price premium in New Haven. Kain and Quigley (1970) find only about a 5% premium in St. Louis. In a careful survey of grocery price discrimination, Sexton (1971) reports that the collective wisdom of the studies is inconclusive. There was generally no evidence of a ghetto price premium holding constant the type of store (i.e., chain, large independent, small independent). However, prices at chains were lower than at independents (particularly small independents) and chains were disproportionately located outside the ghetto.

In neither housing nor food is there independent evidence of the roughly 14% black premium in the early 1970s that is implied by my findings, though in both the grocery and housing literature there is (inconclusive) evidence of somewhat smaller price discrimination. Both sets of studies were subject to substantial criticism at the time of their publication. As it seems impossible to refine these studies after a lapse of 25 years, or to extend them to other goods, one is left to decide whether my own indirect evidence is sufficient to support the conclusion of an approximate 14-percentage-point decline in price discrimination against blacks since 1974.

The finding of a dramatic difference in apparent CPI bias by race does not demonstrate that the level of pure race-based price discrimination is declining, or that it was high in the late 1970s. One possibility is that, whereas blacks did face higher prices in the 1970s than whites, the cause was not race *per se* but something correlated with race. I explore two such possibilities below.

- The effect of migration: King and Mieszkowski found that black families who had recently migrated to the current urban area generally faced higher housing costs than otherwise similar families. It is plausible that recent migrants face a higher cost of living in general than natives, because it takes time to acquire the knowledge to shop efficiently. To explore this possibility I have re-run the regression of Table 5 with two added regressors (entered both separately and jointly). The first is a dummy variable equal to 1 if the household lived in a different state last year; the second is a dummy equal to 1 if the household lived in a different county last year. For both blacks and whites, both of these variables are insignificant and have essentially no effect on the other coefficients.
- Central-city costs: Living costs are generally higher in central cities than suburbs or outside SMSA's, for reasons that will be discussed when I examine the SMSA dummy coefficients below. Perhaps blacks' gain on whites in terms of cost of living is not due to eroding racial price barriers *per se*, but to the increasing ability of blacks to move to the suburbs, where prices are lower for all. To test for this, I have re-run the regression underlying Table 5, for both blacks and whites, including a dummy variable equal to 1 if the household lives in an SMSA central city and another dummy if it lives in an SMSA suburb (thus living outside an SMSA is the excluded group). For both races, food's share is about 1.5 points higher in SMSA's than out of SMSA's, but there is no significant difference between city and suburb, and the other coefficients were unaffected.

b. Functional Form

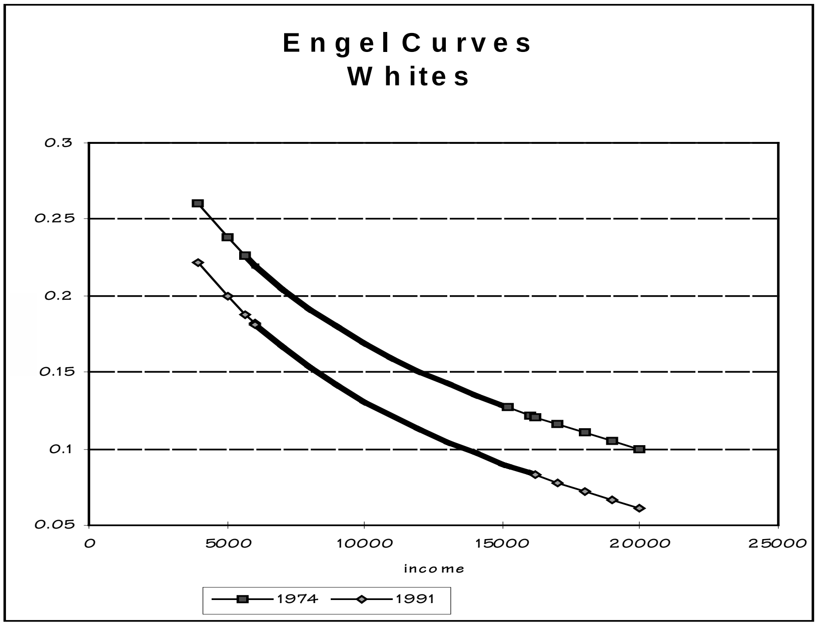
The estimating equations, (7) and (7a), presume that ω_f is linear in the log of real income. But if the true relationship is nonlinear,[28] the year dummy coefficients in (7) is in part picking up unmeasured nonlinearity in the relationship between the ω_f and income rather than unmeasured growth in income. [29]

Even if the true demand function is quite nonlinear in income, it is unlikely that such nonlinearity gives rise to much of the estimated drift in the Engel curves reported above, as can be seen in Figure 5. This figure plots the Engel curves for whites, based on the coefficients from Table 5[30] for 1974 and 1991. The Engel-curve drift apparent in this graph is what I attribute to CPI bias. Functional-form bias might arise if most of the 1974 observations come from the left portion of that fitted curve and the 1991 observations come disproprortionately from the right portion. Then one could imagine that there is but one "true" Engel curve, which the regression failed to find because of the imposed log-linearity.

The data-point configuration envisioned above does not occur. The heavy marker-free portions of the two Engel curves in figure 5 depict the sample (within-year) mean plus and minus one standard-deviation. Since the observations do not come from different levels of CPI-deflated income, it is unlikely that unmeasured nonlinearity accounts for the time dummies (i.e., for the drift in estimated Engel curves).

[28] Indeed, since a budget share is are bounded between 0 and 1, it cannot be linear in log income as income grows without limit.
[29] Several authors have found higher-order terms in log income to be significant.
[30] Calculated at sample average values of the regressors other than income and the year dummies

Figure 5



I have checked for nonlinearity by exploring various functional forms within a given year.[31] By restricting my "fishing" to a single year (I chose 1976) I eliminate CPI-bias induced mismeasurement of income; thus any nonlinearity that I find (or do not find) is due to true variation in income. Among the forms attempted, I obtained the best fit with

Equation 17

$$\omega = \phi + \beta_1 (y_i - p) + \beta_2 (y_i - p)^2 + \sum \theta_X \cdot X_i + \mu_i$$

As a further check I ran equation (17) separately for each year. For both races the quadratic fit the data better than the linear equation, with adjusted R^2 rising an average of 12 percentage points. For whites, all of the coefficients including income and income squared were quite similar across years. For blacks, all of the coefficients **except** income and income squared were stable across years. However, the income and income squared coefficients were very unstable, likely reflecting the fact that there were relatively few very high black family incomes. Inasmuch as the income coefficients for blacks and whites were quite similar to one another in the prior specification, in the quadratic specification below I assume that the black and white income coefficients are identical; however, I allow all of the other coefficients to vary by race.

The next task with the nonlinear form is to incorporate the errors-in-variables hypotheses concerning the CPI, as represented in equations (4) and (5). Making these substitutions and collecting terms involving the various CPI errors, equation (7) becomes,

Equation 18

$$
\begin{aligned}
&\omega_{i,j,t} = \phi + \beta_1\left(y_{i,j,t} - \pi_{j,t}\right) + \beta_2\left(y_{i,j,t} - \pi_{j,t}\right)^2 && \\
&+ \gamma\left(\pi_{f,j,t} - \pi_{n,j,t}\right) && \\
&+ \gamma\left(\varepsilon_{f,t} - \varepsilon_{n,t}\right) - \beta_1\varepsilon_t + \beta_2\varepsilon_t^{\,2} && [\delta_1 \cdot D_t] \\
&- 2\beta_2\left(y_{i,j,t} - \pi_{j,t}\right)\cdot \varepsilon_t && [\delta_2\left(y_{i,j,t} - \pi_{j,t}\right)\cdot D_t] \\
&- 2\beta_2\left(y_{i,j,t} - \pi_{j,t}\right)\cdot p_{j,0} && [\delta_3\left(y_{i,j,t} - \pi_{j,t}\right)\cdot D_j] \\
&- 2\beta_2 p_{j,0} \cdot \varepsilon_t && [\delta_4 \cdot D_j \cdot D_t] \\
&+ \gamma\left(p_{0f,j} - p_{0n,j}\right) - \beta_1 p_{j,0} + \beta_2 p_{j,0}^{\,2} && [\delta_5 \cdot D_j]
\end{aligned}
$$

The square-bracketed terms at the end of the lines indicate the dummy-variable configurations that in principle would capture each of the error contributions. Unfortunately, this specification yields 509 dummy variables (17 time dummies, 17 time dummies interacted with income, 25 city dummies, 25 city dummies interacted with income, and 425 city-time cross dummies). I did not attempt to estimate this equation over the geographic sample, with the price variable and the SMSA dummies included.

I estimate the quadratic form only with the national sample, thus gaining observations and shedding variables (the relative price of food and the SMSA dummies). In addition, I impose more structure on the time path of CPI bias. In place of a completely free set of year dummies I assume that the cumulative bias for each component of the CPI grows piecewise linearly:

[31] In addition to the quadratic term reported here, in other unreported specifications I added cubic and

Equation 19

$$\varepsilon_t = \varepsilon_0 \cdot t + \hat{\varepsilon}_0 \cdot \hat{t}$$

$$\textit{where} \qquad t = year - 1973$$

$$\textit{and} \qquad \hat{t} \quad \begin{array}{ll} = year - 1980 & \textit{if} \quad year \geq 1980 \\ = 0 & \textit{otherwise} \end{array}$$

Under this form the annual (incremental) bias is equal to ε_0 from 1974 through 1980, and is equal to $\varepsilon_0 + \hat{\varepsilon}_0$ thereafter. This specification allows for the deceleration of bias after 1980 that appeared in Table 4 and Figure 3.

Using the national sample and omitting the food-price and SMSA dummies, the specification is

Equation 20

$$\begin{aligned}
&\omega_{i,t} = \phi + B \cdot \phi_B + + \beta_1 \left(y_{i,t} - \pi_t\right) + \beta_2 \left(y_{i,t} - \pi_t\right)^2 + \sum X_i \theta_x + \sum B \cdot X_i \cdot \theta_{BX} \\
&+ \rho_1 \cdot t + \rho_2 \cdot \left(y_{i,t} - \pi_t\right) \cdot t \\
&+ \rho_{B1} \cdot B \cdot t + \rho_{B2} \cdot B \cdot \left(y_{i,t} - \pi_t\right) \\
&+ \hat{\rho}_1 \cdot \hat{t} + \hat{\rho}_2 \cdot \left(y_{i,t} - \pi_t\right) \cdot \hat{t} \\
&= \hat{\rho}_{B1} \cdot \hat{t} + \hat{\rho}_{B2} \cdot \left(y_{i,t} - \pi_t\right) \cdot \hat{t} \\
&+ \mu_{i,j,t}
\end{aligned}$$

where B = 1 if the household is black. Thus coefficients subscripted by a $_B$ represent the differential effect of the variables on blacks.

The time dummies, both free-standing and interacted, have been replaced by a simple time trend. Interpretation of the time coefficients is given below:[32]

---

quartic terms.

[32] After 1980, replace ρ_1 with $\rho_1 + \hat{\rho}_1$, etc.

Equation 21

$$\rho_1 = \gamma\left(\overline{\pi}_{f,t} - \overline{\pi}_{,n,t}\right) + \gamma\left(\varepsilon_{f0} - \varepsilon_{n0}\right) - \beta_1\varepsilon_0 + \beta_2\varepsilon_0^2$$
$$\rho_2 = -2\beta_2\varepsilon_0$$

where $(\overline{\pi}_{f,t} - \overline{\pi}_{n,t})$ is the average rate of CPI-measured relative food price drift over the sample period.

The coefficient ρ_2 enables us to identify ε_0, the average annual rate of CPI bias. And borrowing our estimate of the price coefficient γ from Table 3, we can identify (ε_{f0} - ε_{n0}), the annual bias in the relative price of food. Note that with the nonlinear form we do not need an estimate of γ to identify the overall CPI bias, ε_0.

Table 6 reports OLS estimation results for equation 20:

Table 6
Regression Results from Equation 20

	WHITES		**BLACKS**	
VARIABLE	COEFFICIENT	T-STAT	DUMMY COEFFICIENT	T STAT
Const	4.643	56.37	.066	3.76
Husband Age	.00036	4.17	-.00024	-0.77
Wife Age	.00012	1.30	-.000007	-0.20
# Children	.020	56.50	-.0041	-2.86
Husb Education	-.00008	-0.51	.0006	1.04
Wife Education	-.0013	-6.61	-.0032	-3.78
Unemp	.00056	4.37	.00076	1.23
Husb Hrs	.005	10.45	-.0027	-1.21
Wife Hrs	-.0026	-6.57	-.0030	-1.49
Ygrow	-.000006	-0.69	-.014	-1.90
Yshrink	-.045	-22.84	-.022	-2.67
Fdshare-out	.061	5.32	-.0003	-0.87
β_1	-.888	-49.76		
β_2	.043	43.69		
time·ln(atyr) (ρ_2)	.00215	4.97	.00250	2.66
time (ρ_1)	-.0230	-5.79	-.025	-2.90
$\hat{\rho}_1$	.0138	2.55	-.0079	-0.50
$\hat{\rho}_2$	-.00135	-2.28	.00085	0.49
adj R^2	.538			

Table 7

Annual CPI Bias From Quadratic Form, by Race

COEFFICIENT	MEANING	WHITE	BLACK
ε_0	All items, 1974-80	-.0251	-.0542
$\varepsilon_0 + \hat{\varepsilon}_0$	All items, 1981-1991	-.0093	-.0384
$\varepsilon_{0,f} - \varepsilon_{0,n}$	Food – Nonfood 1974-1980	-.0041	.0067
$\left(\varepsilon_{0,f} + \hat{\varepsilon}_{0,f}\right) - \left(\varepsilon_{0,n} + \hat{\varepsilon}_{0,n}\right)$	Food – Nonfood 1981-1991	-.0058	.0056

The second line of (21) yields point estimates of ε_0 and $\hat{\varepsilon}_0$. Point estimates are given in Table 7 above. The results are broadly consistent with all of our prior CPI bias estimates. For whites the bias fell from about 2.5% per year in 1974-1980 to 1% thereafter. The annual bias for blacks is 3 points higher than whites throughout the time period, about twice the gap estimated in the log-linear forms. For whites, food is biased about a half-point more badly than nonfood; for blacks the opposite is true.

c. Diminishing Bias

Across all specifications, and for both blacks and whites, the results suggest that the average bias was at least one point lower after 1981 than before.[33] Two explanations suggest themselves. First, Baily (1981) has shown that there was much more relative price volatility in the 1970s than in the 1960s. If

---

[3333] However, the deceleration for blacks is not statistically significant; see the t-statistic for $\hat{\rho}_2$ for blacks in table 6.

relative price volatility diminished in the 1980s with the decline in overall inflation, then it is likely that substitution bias in the CPI also declined.

A substantial part of the explanation surely also lies with technical progress at BLS.[34] Perhaps the most significant was the 1983 modification of the homeowner component of the housing cost index.

IV. BACKGROUND COEFFICIENTS

In the above regressions, I have included a series of background regressors, some of which are of considerable interest in their own right. With the exception of the SMSA dummies, which appear only in the smaller geographic sample, the same background regressors appear in all equations; as their coefficients are similar across these equations, I restrict my discussion to their values in the final regression.

a. The SMSA Coefficients

The SMSA dummy coefficient's interpretation is given by equation (9) and reported in Table 4. We can only extract SMSA-specific estimates of p_0 from these estimates after making some assumptions concerning the relationship between p_{0f} and p_{0n}. Two natural assumptions concerning geographic variation in the prices of food and nonfood present themselves:

i Food is transportable; it is sold on a national market and commands the same price everywhere[35];

ii. Geographic price differences occur equally across commodities; the **relative** price of food and nonfood is the same across SMSA's.

Under either assumption we can use equation (9) to convert SMSA dummies into cost-of-living indices. The calculated CPI bias based on assumption (ii) is presented in the last column of Table 4. Under this more conservative[36] assumption (only the price of nonfood varies across SMSA's), the most expensive SMSA (New York) is 49% more expensive than the least expensive cities (Dallas and Minneapolis).

The pattern of SMSA cost-of-living variation is amazingly reasonable. Standard urban models tell us that the cost of living rises with city size, and that money incomes should vary in the same proportion to compensate for inter-SMSA cost of living differences. Qualitatively, then, we expect to see food's share of income higher, correcting for money income, in big cities than small.

---

[34] Greenlees and Mason (1996) provide a tabular history of technical improvements in the CPI.
[35] This is the assumption of standard urban models.
[36] Under assumption (ii) calculated interurban cost of living differences are about 40% smaller than under assumption (i).

A simple regression of my calculated SMSA cost of living on SMSA population yields the following result:

Equation 22

$$\ln(p_{j,0}) = .15 + .0170 \cdot POP/1000000 \qquad R^2 = .289$$
$$(3.05)$$

The cost of living rises 1.70% per million population. Independent evidence on the effect of city size on cost of living is sketchy, but consistent with this estimate. In an admittedly back-of-the-envelope calculation based on a simple urban model and reasonable assumptions regarding commuting cost and other parameter values, Mills and Hamilton (1994) calculate that the cost of living rises by approximately 1.5% per million residents. Rosen’s (1979) interurban wage regressions indicate that a 10% population increase leads to approximately a 0.5% increase in hourly wages, presumably to compensate for a cost of living increase. A population increase from 3 million to 4 million (i.e., 30%) thus leads to a 1.5% increase in wages. The equation-22 coefficient of 1.70% is very much in line with this other evidence. The pattern of SMSA dummy coefficients above is both qualitatively and quantitatively consistent with the urban models; recipients of these high wages do not act rich (i.e., do not spend a low fraction of their incomes on food); they act like consumers whose high wages are offset by high prices.

The ability of the Engel-curve approach to identify inter-urban cost of living differences gives some reason to believe that estimates of year-to-year cost of living differences might be accurately estimated by the same approach.

b. Education

For both races (Table 6), husband’s education has a small and insignificant effect on food consumption. For wives, particularly black wives, the effect is large and significant. A 4-year rise in a white wife’s education reduces food’s share by 0.52 points, about the same as a 5% rise in income. A similar rise in a black wife’s education reduces food’s share by 1.8 points, the same effect as is generated by an 18% rise in income.

One possible explanation for this finding is that a rise in female education shifts family preferences away from food; another is that the rise in education shifts intra-family bargaining power toward wives, and that they respond by reducing food expenditure. But these arguments would seem to lead in the opposite direction; an increase in female education should lead the family to devote more resources to food. [37]

There is a third explanation, consistent with the hypothesis of this paper; a rise in female education reduces the cost of living. According to this explanation, food's share falls with female education because real income rises. More educated females shop more efficiently, or to put it differently more educated females run more efficient household-production technologies. According to this hypothesis, a 4-year rise in female education enables a white household to get about 5% more out of its expenditure dollars and a black family about 18% more. If this interpretation is correct, the rate of return to black female education is approximately 9%, counting only the reduced-cost-of-living effect.[38] But of course some or all of this effect might be a return to (unobserved) ability, omitted but correlated with education.

c. Dynamics

Table 8 below illustrates the dynamics of food consumption as income changes. First, I calculate food consumption (in dollars) at new steady-state income = $11,000 (in 1968 dollars) and at the sample means of all of the other regressors, setting both income-growth variables equal to zero (this gives steady-state value of ω_f, which I solve for food consumption). Then I simulate respectively a 10% rise in income (to the new steady state) and a 10% decline in income (to the same new steady state).

[37] Recall that the wife's hours of work are included in the regression; the income coefficient is not proxying for hours.

[38] The calculation is done as follows: normalize husband's and wife's earnings to 1 (each). Let the cost of a year of education be 1 (i.e., a year of the wife's income foregone). Four years of education costs 4 and reduces food's share by 1.8 points, which implies a rise in family income of 18%. Since income = 2, the rise in income is .36. The rate of return is .36/4 = 9%.

Table 8
Food-Consumption Response
to 10% Income Change
WHITES

COL. #	1	2	3	4
	INCOME	RISES	INCOME	FALLS
	FOOD	INCOME	FOOD	INCOME
Old Steady State	1354	10000	1441	12220
Transition	1395	11000	1440	11000
New Steady State	1395	11000	1395	11000

Reading down Columns 1 and 2, income was originally $10,000 and food consumption was $1354. In row 2 income jumps 10% (to $11,000); food consumption immediately rises to $1395. Row 3 reveals that the new steady state food consumption is $1395. When income rises consumers immediately jump to the new steady state, apparently regarding the income increase as permanent. By contrast, when income falls in Columns 3 and 4 (from $12,220 to $11,000) white consumers make no adjustment in the first year; they appear to regard a fall in income as completely transitory.

The behavior of blacks is quite different, as is revealed in Table 8a below:

Table 8a
Food-Consumption Response
to 10% Income Change
BLACKS

COL #	1	2	3	4
	INCOME	RISES	INCOME	FALLS
	FOOD	INCOME	FOOD	INCOME
Old Steady State	1722	9000	1891	11111
Transition	1788	10000	1870	10000
New Steady State	1803	10000	1803	10000

When blacks' incomes rise, the first-year response is only about 80% of the steady state; when income falls they immediately close about 25% of the steady-state gap. Whereas whites treat a rise in income as permanent and a fall as so temporary as to be ignored, blacks take a fall in income more seriously and give less credence to a rise.

d. Eating Out

If home-cooked and restaurant meals are perfect substitutes then the dependent variable should be the budget share of total food; otherwise the relative price of restaurant meals should enter on the right-hand side. Unfortunately neither restaurant food expenditure nor its price seems to be available in high-quality form. The former is reported by the PSID but only in fairly broad bands and many zero entries, and the latter has the same problems as the food CPI discussed above. I have approached the problem in two ways.

First, despite data-quality concerns I ran the regressions of Table 6 above with total food's share as the dependent variable. None of the point estimates was materially affected, but the fit was considerably worse. [39] I do not report these results. Second, as can be seen, I added the budget share of food away from home as a right-hand side regressor. Inclusion has no material effect on other coefficients. The coefficient is always small and positive (!), and significant.

It is implausible that home-cooked and restaurant meals are complements as implied by the coefficient; perhaps the restaurant variable is picking up a fixed gourmand effect. In any event the effect is very small; a 1-point rise in restaurant-food's share increases home-food's share by 0.06 points.

V CONCLUSION

In this paper I estimate Engel curves for black and white households in the PSID. If the Engel curves are properly specified, and if there are no systematic errors in the data, the coefficients of these curves should not move from year to year. By observing the annual drift in these estimated Engel curves, I am able to infer the rate of bias in the CPI, which was used to deflate the income variable.

For all of the above specifications, we find evidence of substantial CPI bias, which is generally much more severe during the 1970s than later, and for blacks than whites. Table 9 pulls together the annual-bias estimates from all of the various functional forms and data sets:

[39] As predicted, the income elasticity rises somewhat (β falls); this is almost exactly offset by declines in the time dummy coefficients, leaving the estimated CPI bias virtually unchanged.

Table 9
Summary of Annual Bias Estimates

	WHITE 1974-1981	WHITE 1981-1991	BLACK 1974-1981	BLACK 1981-1991
Geographic Sample	-.0353	-.0067		
National Sample Log-linear	-.0275 (γ=.0369) -.0235 (γ=.07)	-.0151 (γ=.0369) -.0131 (γ=.07)	-.0400 (γ=.0369) -.0359 (γ=.07)	-.0299 (γ=.0369) -.0289 (γ=.07)
National Sample Quadratic	-.0251	-.0093	-.0542	-.0384

References

Baily, Martin N (1981), “Productivity and the Services of Capital and Labor,” Brookings Papers in Economic Activity, pp. 1-50.

Boskin, Micheal J, Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, Dale Jorgenson, (1996) “Final Report of the Advisory Commission to Study the Consumer Price Index,” United States Senate Committee on Finance, Washington, D. C. U. S. Government Printing Office.

Costa, Dora L. (1977), “Less of a Luxury: The Rise of Recreation since 1888,” NBER Working Paper #6054.

Deaton, Angus, and John Muellbauer (1980), “An Almost Ideal Demand System,” American Economic Review, vol 70, pp. 312-326.

Friedman, Milton (1957), A Theory of the Consumption Function Princeton, Princeton University Press.

Gordon, Robert J. (1990) The Measurement of Durable Goods Prices, Chicago, University of Chicago Press.

Greenlees, John S., and Charles C. Mason (1996) “Overview of the 1998 Revision of the Consumer Price Index,” Monthly Labor Review, December, pp. 3-10.

Houthakker, H. S. (1987), “Engel’s Law,” in The New Palgrave Dictionary of Economics (Vol 2), ed. Eatwell, Milgate and Newman, Macmillan Press (London), pp. 143-144

Jorgenson, Dale, and D. T. Slesnick (1997) “Individual and Social Cost of Living Indexes,” in Welfare, Vol 2: Measuring Social Welfare ed Dale Jorgenson, Cambridge, MIT Press.

Kain, John, and John Quigley (1970), “Measuring the Value of Housing Quality,” Journal of the American Statistical Association, vol 65, pp. 532-549.

King, A. Thomas, and Peter Mieszkowski (1973) “Racial Discrimination, Segregation, and the Price of Housing,” Journal of Political Economy vol 81, pp. 590-606.

Krueger, Alan B., and Aaron Siskind (1997), “Assessing Bias in the Consumer Price Index from Survey Data,” Working Paper #392, Industrial Relations Section, Princeton University.

Leser, Conrad E. V. (1963), “Forms of Engel Functions,” Econometrica 31, No. 4, pp. 694-703.

Mills, Edwin S., and Bruce W. Hamilton (1994), Urban Economics, New York (Harper Collins).

Moulton, Brent (1996) “Bias in the Consumer Price Index: What is the Evidence?” Journal of Economic Perspectives, vol. 10 no. 4, pp. pp. 159-77.

Nakamura, Leonard (1996), “Is U. S. Economic Performance Really That Bad?” Working Paper # 95-21/R, Federal Reserve Bank of Philadelphia.

Nordhaus, William (1998), “Quality Changes in Price Indexes,” forthcoming in Journal of Economic Perspectives, vol. 12, no. 1.

Rosen, Sherwin (1979), “Wage Based Indexes of Urban Quality of Life,” in Current Issues in Urban Economics, ed. Peter Mieszkowski and Mahlon Straszheim, Baltimore, Johns Hopkins University Press.

Sexton, Donald E. (1971) "Comparing the Cost of Food to Blacks and to Whites - A Survey," Journal of Marketing, July 1971, Vol 35, pp. 40-46.
Tobin, James (1950), "A Statistical Demand Function for Food in the U. S. A.," Journal of the Royal Statistical Society, Vol 113, pp. 113-141.

Shapiro, Matthew, and David Wilcox (1996) "Mismeasurement in the Consumer Price Index: An Evaluation," in Ben Bernanke and Julio Rotemberg, eds. NBER Macroeconomics Annual Cambridge, MA, MIT Press, pp. 93-142.

Working, H. (1943), Statistical Laws of Family Expenditure," Journal of the American Statistical Association, Vol. 38, pp. 43-56.

Appendix: Cost Functions

e. Within-Group Prices, Weights, and Substitution

The debate over CPI bias properly centers on the question of the accuracy of individual-good prices, their weights in a correct index, and the degree of substitutability among goods as their relative prices change. In my approach, none of these things is observed, except as they are borrowed from the CPI itself. And even when I borrow, the only use I make of CPI data is to estimate the price coefficient.[40] In this section I ask specifically how these crucial components of a cost of living index (prices, weights, and substitutability) manifest themselves in my method, and by extension what assumptions we must adopt in order to place credence in the method.

Clearly, "food" and "nonfood" are not individual goods but aggregates, and therefore, "p_f" and "p_n" are not prices but price indices. In order to write a demand function in the form of Equation (1), I must assume:

i. food and nonfood are additively separable in the utility function, and

ii. subutilities for food and nonfood are both homothetic.

Homotheticity is required so that food (and nonfood) can meaningfully be decomposed into a price index and a quantity index.; without homotheticity this decomposition changes with utility. To see the role of homotheticity consider the suboptimization problem for nonfood:

Equation 13

$$\begin{aligned}&\max \quad u_n = v(z_1,...,z_m)\\ &s.t. \quad x_n = \sum_1^m z_i \rho_i\end{aligned}$$

where z's are individual goods and ρ's are their prices. If preferences are homothetic, the solution can be written as

Equation 14

$$\rho_i z_i = h(u_n) \cdot f_i(\rho_1,...,\rho_m)$$

[40] I deflate money incomes by the CPI, but this is only for convenience. If I left incomes undeflated, my year dummy coefficients would be my estimate of the movement in the true cost of living index, rather than the movement of this index less the CPI.

Now without loss of generality let subutility be homogeneous of degree 1 (not just homothetic):

Equation 15

$$h(u_n) = u_n = n \Rightarrow$$

$$\sum_i f_i(\rho_1,...,\rho_m) = \sum_i \rho_i z_i \mid u_n = 1 = c_n(\rho,1)$$

where $c_n(p,1)$ is the consumer's expenditure function (for nonfood) evaluated at $u_n = 1$.

With this normalization we can treat utility itself as the quantity index, and expenditure decomposes:

Equation 16

$$x_n = P_n \cdot n$$

That is, total expenditure on nonfood can be decomposed into a price index P_n and a utility or "quantity" index. (And of course the same is true for food.) This verifies that under homotheticity within food and nonfood, and additivity between food and nonfood, the form of equation (1) is correct. Next I show how it is possible to infer a great deal about movements in P (the true cost of living) without knowing the values of the individual prices or the form of $f(\rho_1,\ldots,\rho_m)$. Let us suppose that some price in P_n, call it ρ_{n1}, declined (but that the decline was either not observed or not properly interpreted by BLS). The effect of this change on food's share is given by

Equation 17

$$\Delta\omega_f = \Delta \ln \rho_{1n} \cdot \frac{d\omega_f}{d \ln \rho_{1n}} = \Delta \ln \rho_{1n}\left(-\beta \cdot (1-\alpha)\frac{d \ln P_n}{d \ln \rho_{1n}} - \gamma \frac{d \ln P_n}{d \ln \rho_{1n}}\right)$$
$$= \left\{\Delta \ln \rho_{1n}\left(\frac{d \ln P}{d \ln \rho_{1n}}\right)\right\} \cdot \left(-\beta - \frac{\gamma}{1-\alpha}\right)$$

For an unobserved change in the price of a foodstuff, ρ_{1f}, the analogous expression is

Equation 18

$$\Delta\omega_f = \left\{\Delta \ln \rho_{1f}\left(\frac{d \ln P}{d \ln \rho_{1f}}\right)\right\} \cdot \left(-\beta + \frac{\gamma}{1-\alpha}\right)$$

The first term in the second line of (17) (and repeated in equation (18), {•}, is the change in the true cost of living brought about by our hypothetical Δρ. Next, aggregate the effects of all of the price changes within both food and nonfood:

Equation 19

$$\begin{aligned}\Delta\omega_f \quad &= \sum_{k=1}^{m} \left\{\Delta \ln \rho_{kn}\left(\frac{d\ln P}{d\ln \rho_{kn}}\right)\right\}\cdot\left(-\beta-\frac{\gamma}{1-\alpha}\right)\\ &+\sum_{l=1}^{L}\left\{\Delta \ln \rho_{lf}\left(\frac{d\ln P}{d\ln \rho_{lf}}\right)\right\}\cdot\left(-\beta+\frac{\gamma}{1-\alpha}\right)\end{aligned}$$

The change in the true cost of living brought about by all of these price changes is

Equation 20

$$\Delta \ln P \quad = \sum_{k=1}^{m} \left\{\Delta \ln \rho_{kn}\left(\frac{d\ln P}{d\ln \rho_{kn}}\right)\right\} \qquad +\sum_{l=1}^{L}\left\{\Delta \ln \rho_{lf}\left(\frac{d\ln P}{d\ln \rho_{lf}}\right)\right\}$$

In essentially all CPI research, the strategy is to first (attempt to) observe all of the Δρ's and then to make an assumption about (d ln P)/(d ln ρ). In a fixed-weight index the derivative is simply given by the individual goods' budget shares; if substitution is allowed for, some modification is required based on demand parameters.

Systematic direct observation of the components of {•} is extraordinarily difficult; hence the lively debate over bias in the CPI. My approach rests on the proposition that it is much easier to observe the other components of equation (19), and that we can use this insight to solve directly for ΔlnP rather than estimate it piece by piece.

If one is willing to make an assumption about the relative importance of unobserved price changes as between food and nonfood (an assumption on the value of **r**, see equqations 10 and 11), then with data on food's budget share, and the demand parameters β and γ, one can infer movements in the true cost of living without recourse to direct data on within-group prices, budget shares, or substitution elasticities. These prices, budget shares and substitution elasticities will manifest themselves in food's share directly. Unless one has faith in the identification by functional form implicit in the nonlinear specification of Equation (24) and the identification of bias from Equation (27), this is as far as we can go; the level of bias is identified, regardless of the source of bias (missing price change or failure to account for substitution), up to an assumed value of **r**, the unknown ratio of bias in food and nonfood.