

Measurement Error in the Consumer Price Index: Where Do We Stand?

David E. Lebow and Jeremy B. Rudd*
Board of Governors of the Federal Reserve System

December 7, 2001

Abstract

We survey the evidence bearing on measurement error in the CPI and provide our best estimate of the magnitude of CPI bias. We also identify a “weighting” bias in the CPI that has not been previously discussed in the literature. In total, we estimate that the CPI overstates the change in the cost of living by about 0.6 percentage point per year, with a confidence interval that ranges from 0.1 to 1.2 percentage points. Roughly half of this bias is accounted for by the CPI’s inability to fully capture the welfare improvement from quality change and the introduction of new items. Our bias estimate is smaller than that found in several earlier studies, in part because the BLS has recently made a variety of improvements to its procedures; our study highlights several potential areas for further improvement.

* We thank Ana Aizcorbe, Ralph Bradley, Rob Cage, Darrel Cohen, Tim Erickson, Thesia Garner, John Greenlees, Pat Jackman, David Johnson, Mary Kokoski, Mary McCarthy, Frank Ptacek, Matt Shapiro, Dan Sichel, Ken Stewart, David Stockton, Sandy Struckmeyer, Roger Von Haefen, Karl Whelan, and David Wilcox for helpful comments and discussions, and the Bureau of Labor Statistics for providing unpublished data. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the staff, by the Board of Governors, or by the Federal Reserve Banks.

During the 1990s, the accuracy of the consumer price index came under increased scrutiny, with several analysts judging that changes in the CPI tended to significantly overstate increases in the cost of living. Most prominently, the Advisory Commission to Study the Consumer Price Index estimated in 1996 that the CPI was then overstating increases in the cost of living by about 1.1 percentage points per year, with a plausible range around this estimate extending from 0.8 to 1.6 percentage points per year. Other commonly cited estimates were of similar magnitude.¹ The ramifications of bias in the CPI are numerous, as these statistics affect public and private indexation arrangements, the construction of data on real output and productivity, and the formulation of government policy, including monetary policy.

This paper derives a new estimate of CPI bias, one that differs from earlier estimates for several reasons. First, the BLS has made a number of improvements to its procedures in recent years; according to a recent General Accounting Office report (2000), these changes led the members of the Advisory Commission by 1999 to revise down their estimates of bias to center around 0.8 percentage point per year. Second, we incorporate new research that has become available since the time of the earlier studies. Third, in areas where no new research is available, we sometimes apply different judgment than earlier researchers regarding the interpretation of existing information. Finally, we identify and quantify a previously unrecognized source of bias, which we label “weighting bias”; specifically, we argue that the CPI’s weights, which are derived from expenditure estimates from the consumer expenditure survey, may be inaccurate in a manner that systematically overstates the true rate of change in the cost of living.²

Table 1 summarizes our estimates of the various sources of bias and compares them with previous estimates. We conclude that the CPI is currently overstating the true rate of change in the cost of living by about 0.6 percentage point per year, with a confidence interval for our

¹ Lebow, Roberts, and Stockton (1994) computed bias estimates that ranged from 0.4 to 1.5 percentage points per year. Shapiro and Wilcox (1996) estimated a midpoint of 1 percentage point per year, with an 80 percent confidence bound of 0.6 to 1.5 percentage points. Of course, these studies were informed by each other, so these are not truly independent estimates.

² We follow the BLS in taking the objective of the CPI to be the measurement of the cost of living--an assumption that commands widespread but not universal assent (see the discussions by Triplett, 2001, Greenlees, 2001, and Deaton, 1998). We also do not question the CPI’s use of “plutocratic” rather than “democratic” weighting of expenditures (see Deaton, 1998).

estimate ranging from 0.1 to 1.2 percentage points. As in previous studies, we judge the largest single source of bias to be the CPI's inadequate accounting for quality improvements and the introduction of new items--the component of bias whose magnitude is most uncertain. However, this is also the component of bias for which our estimates differ most notably from earlier estimates.³

The remainder of this introduction summarizes our findings on each type of bias, and the five sections that follow consider these sources of bias in detail (with section 5 on quality-change/new-items bias constituting the bulk of the paper). Section 6 discusses our aggregation of each type of bias to obtain a confidence interval around our overall bias estimate, and section 7 concludes.

Upper-level substitution bias. Because the CPI is a fixed-weight Laspeyres index, it is subject to substitution bias--that is, it tends to overstate increases in the "true" cost of living because it ignores the substitutions that consumers make in response to changes in relative prices. Estimates of the magnitude of the bias from ignoring substitution *across* the CPI's roughly 8,000 item-area strata--upper-level substitution bias--are typically made by comparing the CPI with an alternative measure that does take substitution into account. Such estimates are relatively uncontroversial, and now center around 0.1 percentage point per year.

Lower-level substitution bias. A similar substitution bias can occur *within* the item-area strata, and bias from failing to capture such substitution--lower-level substitution bias--had previously been estimated to be larger than upper-level substitution bias. However, since 1999 the CPI has utilized within most of the strata a geometric means aggregation formula that does assume a

³ Other studies have attempted to shed light on CPI measurement questions using different methodologies than those employed here. Bils and Klenow (2000) use cross-sectional evidence on durable goods spending to predict which items will display rapid quality change over time; on the assumption that the quality-adjusted price increases should not be especially large for these goods, they estimate that the CPI overstated price increases for durable goods by 2.2 percentage points per year between 1980 and 1996. Nordhaus (1998) and Krueger and Siskind (1998) used data on households' perceptions of improvement in living standards to assess the accuracy of the CPI, each coming to different conclusions regarding the CPI's accuracy in capturing changes in the cost of living. Finally, Hamilton (2001) argues that observed changes in the budget share of food are consistent with real-income mismeasurement (that is, CPI bias) of slightly less than 1 percentage point per year from 1980 to 1991.

certain degree of substitution. Thus, our estimate of lower-level substitution bias is smaller than that from studies made prior to 1999, and centers at just 0.05 percentage point per year. But the degree of within-stratum substitution cannot be measured directly, and we are therefore less certain about the magnitude of this bias than we are in the case of upper-level substitution bias.

New-outlet bias. When new retail outlets are rotated into the CPI sample, the BLS implicitly assumes that any difference in price between the old and new outlets is fully explained by differences in quality. However, the fact that the new outlets reflect shifts in buying patterns suggests that this is not so. Thus, the CPI likely fails to capture the quality-adjusted declines in price that occur as buying patterns shift. We put the magnitude of this bias at 0.05 percentage point per year, an estimate that is sketchy because we have to rely on a single study of food and energy prices in the late 1980s.

Weighting bias. The weights in the CPI are derived from the BLS's consumer expenditure survey and may be measured inaccurately, thereby leading to a "weighting bias" in the CPI--a topic that has not been addressed previously in the literature. The sign of such a weighting bias is not clear *a priori*; it depends on whether items with weights that are too large happen to display above- or below-average price increases. We argue that consumer expenditures as measured in the NIPAs may be more accurate than those used in the CPI; based on those data, we estimate that weighting bias pushed up the rate of change of the CPI by 0.1 percentage point per year on average from 1987 through 1997 (table 3). We find reason to believe that the bias may be smaller going forward, and our point estimate is 0.05 percentage point per year.

Quality-change and new-items bias. The final source of bias in the CPI arises because it is difficult to measure the effect on welfare of changes in the quality of existing items or from the introduction of new items. This is easily the most controversial area of CPI measurement, both because this component of bias is often viewed as being large and because our knowledge is so incomplete that any such estimates must involve a large subjective component. Estimating the magnitude of this bias requires detailed judgments about each category of prices in the index; these judgments are based on a comprehensive review of the available literature on price

measurement. In several instances, updated BLS procedures, new research, or differing judgment led us to make a smaller bias estimate than was chosen by the Advisory Commission--the only study to have addressed the topic in as much detail as we do here. In particular, our estimates of the bias arising from incomplete quality adjustment of transportation, apparel, and computers and other electronic equipment are smaller than the Advisory Commission's estimates (table 4). In all, we judge quality-change and new-items bias to center a little below 0.4 percentage point per year. We place a substantial confidence bound around this figure; this reflects our assessment that our bias estimates for items comprising less than 10 percent of the CPI are based on at least a moderate degree of hard evidence, while our estimates for about 40 percent of the CPI are based on a small or inadequate degree of evidence and our estimates for more than half of the CPI are almost entirely subjective (table 5).

1. Upper-level substitution bias

Because the CPI is a fixed-weight Laspeyres index, it is subject to substitution bias--that is, it tends to overstate increases in the "true" cost of living because it ignores the substitutions that consumers make in response to changes in relative prices. (In the CPI's context, the term "upper-level" substitution bias refers to substitution *across* the item-area strata; the bias that results from substitution among the specific items *within* these strata--"lower-level" substitution--is discussed below.) Estimates of upper-level substitution bias have been reported in a number of studies that compare a Laspeyres CPI with an alternative CPI based on a so-called superlative aggregation formula that does take substitution into account. The most recent and complete study (Cage and Jackman, 1999) yields an estimated bias that averages 0.15 percentage point per year from 1987 to 1997.⁴

⁴ This figure is obtained from the standard practice of comparing a "chained" superlative index--one that updates the expenditure weights annually (as BLS would do were it to produce such an index)--with a BLS-style fixed-base Laspeyres index. Strictly speaking, however, this is not a correct measure of upper-level substitution bias inasmuch as the former index involves different reference indifference curves in each period, while the latter index does not. Unfortunately, it is not known how to construct a point estimate of upper-level substitution bias that deals adequately with this problem. (Blow and Crawford, 2001, develop a method of estimating a *range* of values for this component of bias; however, as their study covers the U.K. retail price index, its results are not directly applicable here.)

Greenlees (2001) has argued, however, that these estimates of substitution bias are too large as a result of random sampling error in the underlying price data; he shows that increases in superlative indexes are biased downward when such sampling error is present.⁵ Greenlees proposes correcting for this problem with a composite estimation procedure that mitigates the effect of the error by averaging the item-area price data with U.S.-level item indexes. This procedure reduces his estimate of upper-level substitution bias (which he computed over the period 1987 to 1995) from 0.12 percentage point to 0.08 percentage point per year. Applying a proportional reduction to the Cage-Jackman estimate of 0.15 percentage point yields our preferred estimate of 0.1 percentage point per year for upper-level substitution bias. We convey our uncertainty about this estimate by assuming a confidence interval around this estimate that ranges from zero to 0.25 percentage point. Note that this range is not symmetric about the midpoint, as we view it as highly unlikely that upper-level substitution bias could be negative.

Is this estimate reasonable going forward? Given that many categories of goods are characterized by persistent changes in relative prices (for example, durable goods *versus* services), one would expect the degree of substitution bias to increase as one moves further from the base period. Surprisingly, though, there is little evidence that upper-level substitution bias in the CPI changes in this manner (Greenlees, 1998).⁶ Furthermore, in 2002 the BLS intends to begin updating the weights and shifting forward the base period in the CPI every two years, as opposed to continuing its past procedure of updating the weights only at the time of major revisions (approximately once every ten years). This more frequent updating should tend to

⁵ Intuitively, one can think of a superlative index as measuring substitution by determining how expenditure shares change in response to changes in relative prices; random error in the prices biases the estimated elasticity of substitution toward unity in much the same way that a regression coefficient is biased toward zero by an errors-in-variables problem. In a private communication, Ralph Bradley argues that aside from this issue of *variance* in the price relatives, small sample effects lead to a *bias* in these price relatives that serves to exaggerate the difference between increases in Laspeyres and superlative indexes; if so, some (perhaps most) of the difference between these indexes--while still a genuine source of bias in the CPI--should properly be attributed to finite-sample effects rather than upper-level substitution bias.

⁶ By contrast, substitution bias in a fixed-weight price index for personal consumption expenditures increases notably as one moves further away from the base period.

ameliorate any increase in the degree of substitution bias that might otherwise have begun to occur. Thus, we expect our estimate of a 0.1 percentage point bias to hold prospectively.⁷

2. Lower-level substitution bias

Substitution occurs within the CPI item-area strata as well, but in this case the expenditure data are not available (even with a lag) to measure the degree of substitution. Accordingly, the magnitude of lower-level substitution bias is known with much less certainty than is the case for upper-level substitution bias. The Advisory Commission and other analysts have generally estimated lower-level substitution bias as the difference between the published CPI (that originally used a modified Laspeyres weighting within strata) and an alternative that employs geometric means within strata. Unlike the Laspeyres formula, which assumes a zero elasticity of substitution, the geometric-means formula assumes a unit elasticity.⁸ This calculation was the basis for the Advisory Commission's estimate that lower-level substitution bias raised measured CPI inflation by 0.25 percentage point per year.

The BLS moved in January 1999 to employ the geometric-means aggregation formula within most of the CPI's item-area strata. The BLS retained the Laspeyres formula in items for which an elasticity of zero was deemed more likely, including renter- and owner-occupied

⁷ Beginning in 2002, the BLS intends to publish a superlative version of the CPI. Historical values of this series would be essentially free from upper-level substitution bias (provided the series were to adequately avoid the Greenlees criticism); more recent values, covering the period for which the requisite expenditure data are not yet available, would be based on an estimate of the substitution effect. Of course, the BLS's goal should be to reduce this bias in the official CPI as well. For this reason, Shapiro and Wilcox (1997) have suggested that the BLS replace the Laspeyres formula in the official CPI with a constant-elasticity-of-substitution formula in which a non-zero elasticity is assumed based on recent years' data. This proposal was seconded by the Conference Board's Study Group on the Consumer Price Index (Conference Board, 1999).

⁸ Actually, in small samples the geometric means formula assumes an elasticity of slightly *less* than unity--see McClelland and Reinsdorf (1999). In particular, under certain assumptions, if all items in a sample of size n have equal weight, the geometric means index is an exact cost of living index for a CES utility function with an elasticity of substitution equal to $1-1/n$. Hence, if elasticities of substitution were truly equal to unity, the geometric means indexes would be biased upward in small samples. McClelland and Reinsdorf report that the average item-area stratum in the CPI includes only nine price quotes per month.

housing, public utilities, and most medical care services; in all, geometric-means aggregation was used for items that constitute roughly three-fifths the weight of the CPI. BLS estimated this revision to have reduced the rate of increase in the CPI by about 0.2 percentage point per year on average. (This is a bit smaller than their previous estimate of 0.25 percentage point, which was based on a calculation that used the geometric-means formula within *all* of the CPI's strata.)

Has the BLS's move toward using geometric means eliminated lower-level substitution bias? The answer depends on whether the true elasticities of substitution within strata tend to be larger or smaller than the BLS assumptions of (slightly less than) one for the geometric-means strata and zero for the Laspeyres strata. As noted above, there exists little evidence on this question. Shapiro and Wilcox (1997) found that the estimated amount of upper-level substitution bias is consistent with a *cross-stratum* elasticity of 0.7, and this might be taken as a lower bound on the typical elasticity among the comparatively homogeneous items *within* strata (although examples certainly can be found of heterogenous strata, such as prescription drugs, for which the BLS did decide to adopt geometric means). By contrast, Tellis (1988) analyzed the results from a large number of papers in the marketing literature that estimate cross-brand elasticities and found a mean elasticity (after adjusting for certain biases in the results) of 2.5. Because the items considered in Tellis's study are more homogenous than most of the CPI strata, this estimate probably represents an upper bound on the typical within-stratum elasticity. Similarly, the handful of studies based on scanner data are of limited use because these studies typically consider products that are more homogeneous than the strata in which they are found.⁹

Although we have very little to go on, our sense is that typical elasticities within the geometric-means strata are probably a bit larger than unity. Accordingly, because the geometric-means formula in small samples is consistent with an elasticity slightly *less* than unity, we suspect that a small amount of lower-level substitution bias remains in the CPI. We therefore

⁹ For example, Bradley *et al.* (1997) study scanner data on milk, canned tuna, ketchup, and toilet tissue. Only the first is itself a CPI item stratum, and furthermore, the authors do not examine these items independently but combine them all into a single measure. Thus, it is hard to interpret their result that a geometric-means index overstates increases in the cost of living as measured by a superlative index (though by less than a Laspeyres index). Reinsdorf's (1999) study of coffee prices is somewhat more relevant, because coffee is an item stratum in the CPI. Reinsdorf finds high substitutability within roasted coffee and within instant coffee, but low substitutability between those two categories; overall, he finds that the geometric-means index rises slightly faster than a superlative index for this stratum.

pencil in a relatively small number--0.05 percentage point--to convey our suspicions. Our subjective confidence interval around this estimate is symmetric and ranges from -0.15 to 0.25 percentage point per year; this range is somewhat wider than that assumed for upper-level substitution bias.

3. New-outlet bias

A third potential source of bias in the CPI involves the rotation of retail outlets into and out of the CPI sample. At the time of rotation, any difference in price between items in the old outlet and items in the new outlet is implicitly assumed to reflect a difference in quality (broadly construed to include not just the quality of the product itself, but also the convenience of the outlet, the helpfulness of the service, and so on). This is an extreme assumption inasmuch as the rotation of outlets in the CPI reflects shifts in households' buying patterns. The very fact that buying patterns change suggests that people believe quality-adjusted prices to be lower at the new outlets; if so, then the CPI fails to capture these quality-adjusted declines in price.^{10,11} (In our discussion of the BLS's use of hedonics in section 5, we discuss one approach that BLS might consider to help alleviate both new-outlet bias and quality-adjustment bias, namely, applying hedonic regression techniques not only for item substitutions at a given outlet but also during sample rotations.)

There are no solid estimates of new-outlet bias. All estimates to date are based on Reinsdorf's (1993) study that compares the prices of certain food and energy items in incoming and outgoing outlets between 1987 and 1989. He found that prices were lower on average at

¹⁰ One exception would be if the price at the old outlet is reduced to match the lower quality-adjusted price at the new outlet; the CPI would correctly capture the price decline in this case. See Shapiro and Wilcox (1996) for a careful discussion of this and other possible scenarios.

¹¹ A related issue involves the speed with which new outlets are brought into the CPI. In particular, "e-commerce" internet sites are sufficiently different from traditional brick-and-mortar outlets that their introduction into the CPI may be occurring with a longer lag than usual. Of course, as with any new outlet, price differences between existing outlets and new internet-based outlets are ascribed to quality differences by BLS. (In addition, to the extent that the rise of e-commerce has led to a slower rise in quality-adjusted prices, new outlet bias may have picked up in recent years. We discuss the internet more generally below.)

incoming outlets by an amount that translated to a difference of about 0.25 percentage point per year. Thus, the bias for these items would be between zero and 0.25 percentage point per year, depending on the degree to which the lower prices reflect lower quality. Lebow, Roberts, and Stockton (1994) judged that new-outlet bias would be relevant for about 40 percent of the CPI, and applied Reinsdorf's 0.25 percentage point figure to yield a bias of between zero and 0.1 percentage point per year for the overall CPI. The Advisory Commission picked the upper end of the range and assumed a new-outlet bias of 0.1 percentage point, and Shapiro and Wilcox (1996) also judged the mean bias to be 0.1 percentage point (though the mode of their subjective distribution was around 0.05 percentage point).

Because no new information has come to bear on this question, our judgment is that the midpoint of the original Lebow, Roberts, and Stockton range is reasonable, and we put our point estimate of new-outlet bias at 0.05 percentage point per year. We are fairly uncertain about this estimate, but we also view the bias as unlikely to be negative; accordingly, we specify our subjective distribution as being skewed to the right, with a confidence interval ranging from zero to 0.20 percentage point per year.

4. Weighting bias

The weights in the CPI are derived from the BLS's consumer expenditure (CE) survey. If these weights are measured inaccurately, then the CPI could suffer from a "weighting bias"--a possibility that has not, to our knowledge, been addressed by previous studies. In contrast to the substitution biases discussed above, there is no *a priori* presumption as to the sign of this bias; it depends on whether items with weights that are too large happen to display above- or below-average price increases. We present evidence that weighting bias pushes up the rate of change of the CPI.

As a means of assessing the accuracy of the CE-based weights that underlie the CPI, we compare them with an alternative set of weights for personal consumption expenditures (PCE) from the national income and product accounts.¹² Neither measure of weights is perfect, but we

¹² Triplett (1997) discusses many of the advantages and disadvantages of these two expenditure measures.

see advantages to the PCE data on balance. In benchmark years, the PCE data are derived in large part from businesses' responses to the economic censuses, which provide a reasonably comprehensive record of expenditures.¹³ The CE survey, by contrast, relies in large part on respondents' memory of their own expenditures as well as their knowledge about the expenditures of other household members, and these may be suspect in many cases. The CE survey also relies on respondents' willingness to report expenditures that may be viewed as private, such as purchases of alcohol or tobacco.¹⁴ Moreover, for the rental value of owner-occupied housing--an extremely important category owing to its large weight--the CE survey estimates are based on homeowners' estimates of what their homes would rent for, and these estimates may be quite inaccurate. (In the PCE data, the equivalent estimates are imputed by applying rent/value ratios for tenant-occupied units to the stock of owner-occupied housing.) Finally, the CE survey's relatively small size has led to additional questions about its accuracy--though the size has increased recently--and it has been suggested as well that sample attrition may lead to biases in aggregate reported expenditures (Triplett, 1997).

To construct weights based on the PCE data that are comparable to the CPI, however, several types of adjustments must be made. (Details of these adjustments are presented in Appendix A.) First, and most important, the CPI is intended to cover only out-of-pocket expenditures by households, whereas PCE is considerably broader in scope, representing all goods and services purchased by both individuals and the nonprofit institutions that serve them. For example, PCE includes all expenditures on medical care whether paid for by households, employers, or governments, whereas the CPI only covers the portion of expenditures paid by households out of their own pockets. In all, roughly one-quarter of PCE consists of expenditures that are outside the scope of the CPI. Thus, we first adjust the PCE data to cover approximately

¹³ The main difficulty with the PCE data in this context lies in the need to subtract the purchases of businesses and governments from total expenditure data in order to obtain spending by households and nonprofit institutions. In the Census of Retail Trade, for example, estimates of expenditures by class of customer is available for each establishment on average, but not by the specific line of merchandise sold.

¹⁴ The probable underreporting of alcohol and tobacco expenditures in the CE survey appears to have been first noted by Houthakker and Taylor (1970, p. 252). More recently, Gieseman (1987) discusses this problem, and also cites evidence of recall bias in the CE survey. (These types of mismeasurement are not unique to the CE survey; see Deaton, 1997, pp. 24-28 for a discussion of recall bias and the underreporting of such purchases in other countries' household expenditure surveys.)

the same scope as the CPI by eliminating these out-of-scope expenditures; that is, we exclude the portion of expenditures on medical care and education that we estimate was made by governments and employers, as well as expenditures by nonprofit institutions, the consumption of items that are provided without explicit charge (such as many banking services), and so on.¹⁵ Second, we adjust for the fact that the PCE definitions of tenants' rent and owners' equivalent rent include space rent only, while the CPI concepts are somewhat broader, including some utility costs in tenants' rent and some service flows of major appliances in owners' equivalent rent. Finally, we adjust the PCE data for the fact that the CPI covers urban households only (for example, urban households spend a larger share of income on housing and a smaller share on motor vehicles than do rural households). Although the conceptual differences between PCE and the expenditures that underlie the CPI weights are important and these adjustments cannot be made perfectly--the first set of adjustments described above is especially problematic--we believe that our adjustments capture the most important factors needed to make the PCE data roughly comparable to the expenditure data used in the CPI.

After making these adjustments, we use the resulting modified PCE data to construct an alternative set of relative importance weights.¹⁶ Table 2 shows the December 1997 relative importance weights for a 24-item decomposition of the CPI along with alternative, PCE-based weights. The differences between the two sets of weights are substantial, and the pattern of differences largely corresponds to our expectations given the potential problems in the CE survey that we discussed above.¹⁷ Specifically, the CPI weights are smaller for many items like apparel, audio and video equipment, and broad categories of other nondurable goods, where a

¹⁵ Our adjustments benefitted from Fixler and Jaditz's (1997) concordance between the detailed line items in the CPI and PCE.

¹⁶ Given the adjusted PCE expenditures, we constructed the weights using procedures analogous to those used for the CPI. Specifically, the weights are derived from average PCE expenditures during 1993-95; these expenditures were then multiplied by the corresponding relative CPI price change between the 1993-95 average and December 1997 to obtain the December 1997 relative importance weights. The 1993-95 based weights are used to aggregate the CPI since 1998; from 1987 through 1997, an equivalent set of weights based on 1982-84 PCE data is used. (For that earlier period, the relative importance weights are constructed by multiplying the 1982-84 average PCE expenditures by the relative CPI price change from the midpoint of the period, June 1983, rather than from the average of the period.)

¹⁷ See Branch (1994) or Gieseeman (1987) for a related comparison of CE and PCE expenditures.

household head (the usual respondent) may be least knowledgeable about the expenditures of other members of the household. Conversely, the CPI weights are larger for tenants' rent, utilities, and motor vehicles, where a household head probably is more knowledgeable about the overall household's expenditures and where we would therefore expect the relative weight to be boosted by the undercounting of other expenditures. (Going the other way, though, the CPI weight is smaller for other durable goods, where we might also have expected the respondent to be relatively knowledgeable about overall household expenditures.) The CPI weights also are notably larger for owners' equivalent rent, where, in addition to being boosted by the undercounting of other expenditures, the CE survey's expenditure figures may be especially inaccurate--though the direction of bias is not clear *a priori*--because they are based on homeowners' estimates of how much their homes would rent for.¹⁸ Finally, the CPI weights are smaller for tobacco and especially for alcohol, where respondents may be reluctant to report their expenditures accurately.

To investigate the importance of these differences in expenditure weights, table 3 shows an alternative aggregation of the CPI that replaces the CE survey weights with the PCE-based weights from table 2. As can be seen, the alternative index tends to run a little bit lower than the published CPI: The average difference from 1987 through 2000 is about 0.1 percentage point. Much of this gap can be attributed to the substantially smaller weight of shelter in the PCE-based weights than in the CPI, which, combined with the above-average increase in prices in that sector, leads to a smaller increase in the alternative index.

Does this difference constitute a "bias" in the CPI? Because the pattern of differences between the weights corresponds to what one would have expected given the known shortcomings of the CE survey, we suspect that it does, and we have included it as one component of our estimate of overall bias in table 1. Whether this bias should be expected to

¹⁸ The CPI and PCE housing categories are defined slightly differently. Vacation homes--roughly 1.5 percent of the CPI--are included in the lodging away from home category of the CPI, but are included in owner- or tenant-occupied housing in PCE. In fact, the CPI prices very few such vacation homes, and so the lodging away from home index is driven by hotel and motel prices. Thus, our procedure of applying the CPI price to the PCE-based hotel and motel weight is probably reasonable, especially if the user cost of vacation homes tends to behave like the cost of regular owner- or tenant-occupied housing.

persist in the future is a more complicated question. For reasons that are not clear, the 1982-84 CE expenditures that were used to construct the CPI from 1987 to 1997 tend to generate higher rates of price increase than do CE expenditures from other years (Greenlees, 1998; Shapiro and Wilcox, 1997). Our result that the published CPI rose more rapidly than our PCE-based alternative through 1997 could simply be a reflection of that fact; if so, we might expect the magnitude of this bias to diminish now that the 1982-84 CE weights are no longer in use. Circumstantial evidence suggests that this might be the case. The right portion of table 3 compares the CPI based on 1986 CE expenditures (from Cage and Jackman, 1999) with a PCE-based alternative that also uses 1986 expenditures. As can be seen, the difference between increases in these series averages close to zero. Nevertheless, some bias apparently remained in the CPI beyond 1997: For 1998 through 2000, the gap between the published CPI--which was by then based on CE weights from 1993 to 1995--and our PCE-based alternative was actually a bit larger than the average over the earlier period. In all, then, we believe that it is reasonable to assume some positive bias going forward, and we assume a weighting bias of 0.05 percentage point per year--smaller than the historical average but still greater than zero. We set our confidence interval for weighting bias to range from -0.05 to 0.15 percentage point per year.

The BLS is investigating ideas for improving the accuracy of the CE survey, including expanding the diary portion of the survey so that it is filled out by all members of the household and not just the household head. In addition, we believe it would be useful for the BLS to consider augmenting the CE data with additional information on certain components that are known to be poorly measured--such as alcohol and tobacco, and perhaps owners' equivalent rent as well--to bring their estimates more into line with PCE data or other outside information. Such actions could help to improve the accuracy of the CPI.

5. Quality-change and new-items bias

A true cost-of-living index attempts to measure the expenditure needed to maintain a given level of utility. If consumption goods are changing in quality, or if new products are being introduced, then consumer utility will change even if the new items sell for the same price as the items they replace. A cost-of-living index must therefore attempt to value these quality changes. Although

the BLS devotes considerable effort to this task (see Greenlees, 2000, and Moulton and Moses, 1997), it is a daunting one--Shapiro and Wilcox (1996) refer to quality change estimation as the "house-to-house combat of price measurement"--and many analysts believe that unmeasured quality improvement is a source of significant upward bias in the CPI.

Of the several issues surrounding the topic of CPI bias, measuring quality change is easily the most controversial. First, estimates of quality-change bias are often large; for example, the Advisory Commission concluded that unmeasured quality change accounts for some 0.6 percentage point per year of bias in the CPI. Second, the issue is controversial because estimates of bias frequently involve a large judgmental component and are inherently highly uncertain.

Our approach to estimating the magnitude of bias is the same as that taken by the Advisory Commission: We review the research on quality-adjustment bias for each category of expenditure. Although more research is available now than when the Commission wrote its original report, in many cases we still are left with little guidance, and our estimates, like previous ones, are often judgmental.

Table 4 presents our estimates of quality-change bias for items in the various expenditure categories of the CPI. Using consumer expenditures from 1998 as weights, we obtain an overall estimate of quality-change bias of a little less than 0.4 percentage point per year; we view this as an estimate of current and prospective bias. Owing to the substantial uncertainty about this estimate, we place a confidence interval ranging from -0.1 to 0.8 percentage point around our point estimate. For reference, table 4 also presents comparable bias estimates from the Advisory Commission's report, which imply aggregate quality-change bias of about 0.7 percentage point per year.¹⁹ Our smaller estimate stems from lower estimates of bias in a number of categories, especially transportation, apparel, and computers and other electronic equipment.

Table 5 provides some information that helps shed light on the uncertainty around these estimates. We divide the expenditure components into three groups: items for which our bias estimate is based on at least a moderate degree of hard evidence, items for which our estimate is based on a small or inadequate degree of evidence, and items for which our estimate is almost

¹⁹ To a second decimal place, this figure is 0.69 percentage point; this is a little higher than the Commission's 0.61 percentage point estimate, reflecting differences in weights used in the aggregation.

entirely subjective. The implications of this table are sobering. The first category--the items for which we are most confident about our estimates--accounts for only about 7 percent of the CPI; the second category comprises a little less than 40 percent; and the third category is the largest, comprising more than half of the index. Of our 0.37 percentage point estimate of quality-adjustment/new-items bias, essentially none comes from the first category, about two-thirds comes from items in the second category and the remainder comes from items in the third category.

Before proceeding to an item-by-item dissection of quality-adjustment bias, we highlight a few general issues that pertain to our analysis. First, in constructing our estimates of quality-change bias, we must be careful to ensure that there is no double-counting of other types of bias. If we obtain a “true” outside estimate of price change for a specific good that properly measures quality improvements, we must be careful to compare this series to a CPI that has already been adjusted for other sources of bias, such as lower-level substitution bias. In many cases, we can avoid these problems by comparing the “true” estimate with the current-methods CPI (Stewart and Reed, 1999), in which many of the other sources of bias have already been minimized.

Second, in some cases we take an outside estimate of price change to be superior to the CPI when that estimate is based on the sort of detailed and comprehensive data (such as scanner data) that are increasingly becoming available with improved information technology. The fact that these data are more comprehensive and include far more price quotes than the CPI raises the possibility that they may be more accurate, and such data also may allow more rapid introduction of new items into index, thereby minimizing new-items bias. The BLS is investigating how to integrate such data into the CPI, an undertaking with considerable promise in our view.²⁰

Third, quality-change bias probably varies over time (perhaps even more so than other sources of bias). However, estimates of bias are frequently based on examination of a relatively short period of time that might well be atypical. In some cases, we will speculate that past quality changes are unlikely to be repeated in the future, or likely to continue at a slower pace. Similarly, since we are concerned with deriving a prospective measure of quality-change bias,

²⁰ See Richardson (2000) for a description of BLS efforts in this area. Also see Feenstra and Shapiro (2001) for a discussion of some potential pitfalls in using such data.

we do not correct for the effect of new goods that are now fully incorporated in the CPI (such as cellular telephones and VCRs) except to the extent that we view them as being representative of new goods that will continue to be introduced. In addition, we note that BLS has procedures under way to bring new items into the CPI more rapidly (see Lane, 2000), and this could help to mitigate the amount of new-goods bias that is currently present in the index.²¹

Finally, we need some discussion of the weights that we use to aggregate our estimates of quality-change bias for each of the detailed expenditure categories. As discussed in Appendix B, we would ideally like to use weights from a superlative aggregation formula that are as accurate as possible (if we do not, then we run the risk of confounding quality-change bias with the other categories of bias in table 1). We therefore use recent weights from the adjusted PCE data, which, as we discussed in section 4, we suspect to be more accurate than the weights derived from the CE survey. These weights differ from the current CPI weights (the December 2000 relative importance weights) in several ways. First, even after adjusting the PCE data to the CPI's out-of-pocket scope, those PCE data place a higher weight on medical care services, an expenditure category in which we judge there to be a very large quality-change bias. Second, the fact that the CPI weights are based on CE expenditures from 1993-95 implies that the more recent PCE weights place a notably higher weight on computers (whose relative importance declines over time as its relative price falls, only to increase again when the expenditure weights are updated) and internet services (whose expenditure share has increased rapidly in recent years).²²

Because we recognize that not all readers will agree with our assessment about weighting bias in the CPI, we also calculate an aggregation using expenditure shares from the 1998 CE survey, the most up-to-date CE data currently available.²³ Doing this yields a quality-adjustment

²¹ As noted by Pakes (2001), the more rapid introduction of new items into the CPI need not reduce new-items bias, though we expect that in most cases more rapid introduction would indeed help.

²² For similar reasons, the Advisory Commission augmented the weights they applied to computers and certain categories of electrical appliances in their aggregation of quality bias.

²³ These shares are derived from a single year of CE data (adjusted to the CPI item structure and rental equivalence concept) and therefore may be less reliable than either the three-year averages that have gone into the CPI's weights thus far or the two-year averages that will be used beginning in 2002. In addition, these data have not been through all of the processing that go into production of the weights

bias of 0.31 percentage point per year, 0.06 percentage point smaller than the estimates presented in table 4, with the difference driven by the higher weight of medical care services in the adjusted PCE data. (Of course, this smaller estimate is well within the substantial confidence interval around our point estimate.) Thus, readers who prefer the CE survey data to the adjusted PCE data should reduce our overall estimate of CPI bias by about 0.1 percentage point--0.05 percentage point for weighting bias, and another 0.06 percentage point to reflect a smaller estimate of quality-change/new-items bias.

Item-by-item estimates of quality change

Shelter: If quality adjustment is the “house-to-house combat” of CPI bias estimation, then obtaining an estimate for the quality bias in shelter is much of the battle: Owners’ equivalent rent and tenants’ rent account for more than one-quarter of the CPI. The Advisory Commission was forced to rely largely on informal judgments in estimating that quality change biased the CPI for shelter upward by 0.25 percentage point per year between 1976 and 1996. Since the Commission’s work, two additional papers have become available that shed additional light on this question. First, Moulton (1997) used data on housing characteristics to consider quality-change bias for the rent of tenant-occupied housing; he concludes that the CPI for tenants’ rent *understates* the true quality-adjusted price increase by 0.15 to 0.25 percentage point per year. Second, Crone, Nakamura, and Voith (2001) use hedonic techniques to estimate constant-quality price indexes for tenants’ rent based on data from the American Housing Survey. Their estimate is close to Moulton’s, rising 0.3 percentage point per year more rapidly than the current-methods CPI (this is perhaps not remarkable, given that both studies’ methodologies and data sources are similar).

Crone *et al.* also attempted to measure quality-change bias in the service flow obtained from owner-occupied housing. Under the assumption that trait prices for owner- and tenant-occupied housing are the same, Crone *et al.* are able to estimate a hedonic model that allows them to impute constant-quality rents for owner-occupied housing. According to their results,

in the official CPI. We are grateful to the BLS for allowing use of these unpublished data.

the true cost of owner-occupied housing increased by about 0.6 percentage point per year less than the CPI for owners' equivalent rent (based on current methods) between 1985 and 1993.²⁴

It is not clear, however, that the difference between the CPI for owners' equivalent rent (OER) and Crone *et al.*'s estimates completely reflect the effect of quality-change bias. Crone *et al.*'s maintained hypothesis--that the trait prices for renter- and owner-occupied housing are the same--is identical to the rental equivalence concept that the BLS invokes in computing OER; namely, that OER seeks to capture what an owner-occupied unit would receive were it rented out.²⁵ In principle, therefore, the Crone *et al.* estimates of constant-quality OER should only differ from the CPI for owners' equivalent rent for the following three reasons. First, the sample of housing units employed by Crone *et al.* in their study differs from that employed by the BLS in computing OER. Second, the specific procedure for imputing rents to owner-occupied units differs: Over the period considered by Crone *et al.*, the CPI imputed rents to an owner-occupied unit based on the rents of comparable renter-occupied units, while Crone *et al.*'s imputation is based on estimates from a hedonic regression. Finally, as with any item in the CPI, BLS introduces new housing units into the existing sample during a sample rotation by "linking in" the new price quotes--thereby assuming that there is no difference in quality-adjusted price between old and new units. By contrast, new units in the Crone *et al.* sample are handled by their hedonic regressions. Sample rotation in the CPI occurs roughly every ten years; notably, the period considered by Crone *et al.* saw one instance of sample rotation, in 1987.

Of these three possible sources of difference between the Crone *et al.* OER measure and the corresponding CPI measure, only the third can unambiguously be considered a failing of the CPI that Crone *et al.*'s procedure would remedy. Hence, we are not inclined to view the entire 0.6 percentage point difference between the CPI for owners' equivalent rent and Crone *et al.*'s estimate as reflecting quality-change bias. We therefore scale down the bias estimate

²⁴ The current-methods rent indexes incorporate estimates of the effects in earlier years of the 1995 adjustments to the formulas used to compute rent changes, as well as the introduction (in 1988) of quality adjustments to control for depreciation (Stewart and Reed, 1999).

²⁵ We are indebted to Timothy Erickson for clarifying this point. Interestingly, the hypothesis that trait prices are equal across owner- and renter-occupied housing is rejected in Crone *et al.*'s data; this calls into question both Crone *et al.*'s methodology as well as the BLS's rental equivalence methodology.

judgmentally, and assume that prospective quality-change bias for owner-occupied housing averages 0.3 percentage point per year.

The preceding discussion regarding the Crone *et al.* bias estimates for owners' equivalent rent also applies to their estimate of quality-change bias in tenants' rents, as well as to Moulton's estimates. Both studies use data from the American Housing Survey to compute estimates of tenants' rents that control for changes in observable housing characteristics (although Moulton's menu of characteristics is somewhat less detailed than that employed by Crone *et al.*). We therefore assume a downward bias of 0.2 percentage point per year for tenants' rent, which is a little smaller than both Crone *et al.*'s point estimate and the upper bound of Moulton's range of estimates. (Note that this is one of the few estimates in this paper that we classify in table 5 as being based on at least a moderate degree of hard evidence.)

Using expenditure weights from the 1998 PCE data, we find that tenants' rent and owners' equivalent rent contribute about 0.03 percentage point to the bias in the overall CPI. As for the other components of shelter costs--that is, lodging away from home and household insurance--we know of no analyses of these prices nor do we have any special reason to doubt the accuracy of the CPI for these items. We therefore assume no quality-change bias for these categories.

Medical care services: Evaluating quality changes for medical care is perhaps the most daunting challenge BLS faces. There are at least two quality-related problems with the CPI's estimates of medical care prices. The first is methodological: Prior to 1997, the CPI priced hospital services by pricing a fixed set of inputs (such as one night's stay in a hospital room), rather than the costs incurred in treating a given disease. Because technological change can alter the mix of inputs used to treat a given condition (for example, some procedures no longer even require a hospital stay), such a procedure will likely mismeasure the "true" price of medical services. As a result, the CPI changed its procedure for measuring hospital services prices beginning in 1997, and now attempts, in principle, to price a course of treatment for a given disease. In practice, this is implemented by repeated pricing of a given set of services specified in a random selection of bills for a specific hospital visit, combined with an attempt to identify changes in treatment practices. While this should bring the CPI closer to the goal of being able to price the treatment

of a specific disease, most such changes in treatment are apparently unlikely to be captured by the CPI's procedures.²⁶

The results from two studies--Shapiro, Shapiro, and Wilcox's (2001) analysis of cataract surgery and Cutler, McClellan, Newhouse, and Remler's (2001) study of heart attack treatments--suggest that failing to control for changes in treatment can represent a significant source of bias. Shapiro *et al.* construct a cost index for cataract treatment that accounts for the changing mix of inputs employed; they find that the difference in growth rates between an index in which input weights change every ten years (as in the CPI) and one for which weights are allowed to change more frequently averages about 5.5 percentage points per year from 1969 to 1994.²⁷ Similarly, Cutler *et al.* estimate that from 1983 to 1994, the actual cost of treating a heart attack rose approximately 2.1 percentage points per year more slowly than the average change in the cost of a fixed input bundle for heart-attack treatment (see their table 8.4).

However, we are aware of one study that reaches a different conclusion regarding the sign of this type of bias. Berndt, Busch and Frank (2001) present estimated price indexes for treatment bundles for clinical depression, albeit only over the five-year period 1991-95. They do not make a clean comparison between an index that captures changes in treatment paths with a CPI-like index that does not, but according to our reading of the information they present, such an index appears to increase about 1.4 percentage points per year *more* rapidly than a CPI-like index over this period, reflecting a switch toward more expensive treatments.²⁸

²⁶ To use an example cited by Shapiro *et al.* (2001, p. 433), sutures are no longer needed after cataract surgery in some cases. Even under the new CPI procedures, small changes such as this might not be picked up by BLS enumerators (and even major changes such as the greater use of outpatient surgery for cataracts might not be captured). Cutler *et al.* (2001, note 24) are similarly skeptical that the new CPI methodology will be able to adjust for significant changes in treatment inasmuch as these changes are too gradual to be identified.

²⁷ However, the difference between the two indexes changes significantly over the sample period, averaging 11 percentage points per year from 1969 to 1985 but less than 1 percentage point per year from 1985 to 1994. (The large difference in growth rates over the earlier period is largely the result of cataract surgery's having moved toward being an outpatient procedure.)

²⁸ We are comparing the average supply price from their table 12.4 with the fixed-weight Laspeyres index from their table 12.7. Of course, this calculation ignores any improvement in quality, which presumably is the reason for the change in treatments in the first place--though the authors describe clinical results suggesting that each of the examined treatments lead to comparable outcomes. We note as

The second quality-related problem with the CPI's estimates of medical care prices involves *pure* quality change--for many (though certainly not all) illnesses, advances in treatment have significantly improved treatment outcomes. The issue of pure quality adjustment is a much more difficult one since it involves assigning a value to a medical outcome; indeed, we have found only one study--the paper on heart attacks by Cutler *et al.*--that provides useable estimates of a quality-adjusted index for a specific medical treatment.²⁹ They find that such an index declines by 1.7 percent per year--a marked contrast from their fixed-weight index, which rises 4.5 percentage points per year more rapidly. However, this estimate is extremely sensitive to the assumed value of additional life (as Cutler *et al.* point out, different assumptions yield indexes that decline by 0.3 to 16.8 percent per year); valuing medical outcomes that involve well-being or quality of life rather than length of life is likely to be even more difficult. In addition, Cutler *et al.*'s cost-of-living index only includes the cost of treatment in the year following a heart attack, even though evidence suggests that treatment received more than one year after a heart attack contributes significantly to lower mortality.³⁰

Altogether, there is an immense degree of uncertainty surrounding the issue of quality adjustment for medical services. That said, we would guess that estimates derived from the Shapiro *et al.* and Cutler *et al.* studies represent upper bounds, because treatments for cataracts and heart attacks likely have enjoyed unusually rapid improvement (indeed, Shapiro *et al.* chose to examine cataract treatments in part for that very reason). In contrast, progress in the treatment of other diseases--such as Alzheimer's or the common cold--has been virtually nonexistent.³¹

well that the results of this study differ sharply from earlier research by the same authors (Frank, Berndt, and Busch, 1999), reflecting changes in their sample.

²⁹ Berndt, Busch and Frank's (2001) estimated price indexes for treatment bundles for clinical depression covered treatments that were found to be comparable. As noted above, however, this finding raises the question as to why the treatments changed over time; perhaps "comparability" ignores some quality issues that are important to consumers, such as time saved in replacing some therapy with drugs.

³⁰ Fixler (1999, pp. 106-7) discusses some of the conceptual difficulties involved in measuring total treatment costs.

³¹ Berndt *et al.* (2001, p. 192) emphasize this point. Note that the discussion above ignores other, deeper issues related to the measurement of medical services prices in a true cost-of-living index, such as the negative impact that reduced mortality could have on lifetime utility (with fixed resources, longer life reduces an individual's sustainable level of lifetime consumption). See Shapiro *et al.*, p. 412, and

(Treatment of depression may be in the latter category as well, at least over the limited period covered by the Berndt, Busch, and Frank study.) To estimate the amount of quality-change bias in medical services (ignoring for the moment the CPI's 1997 changes in procedures), therefore, we assume that rapidly improving treatments such as those for cataracts and heart attacks are representative of two-thirds of all hospital services, and that the amount of quality bias in these rapidly improving treatments is 4.5 percentage points per year (the value reported in Cutler *et al.* for heart-attack treatments).

Obtaining an estimate for the period following the 1997 change in CPI procedures is even harder. In principle, pricing the items from a given bill for a specific hospital visit--as is now done--should allow BLS analysts to identify changes in treatment for a specific disease; this then would permit the BLS to factor these treatment changes into the price index. Thus, the new procedures have the potential to enhance the CPI's ability to price a specified treatment path. In addition, the CPI now measures transaction prices, rather than list prices, to a larger extent than had previously been the case. Unfortunately, very little information is available as to how these changes affected the CPI. The CPI's procedures apparently remain unable to capture most changes in treatment; indeed, discussion with BLS analysts suggests that almost no such changes have been detected since 1997. However, the use of list (or "chargemaster") prices is widely believed to have biased upward the CPI during the 1980s, as rapidly rising list prices were met with increased discounting for certain classes of purchasers (though the Cutler *et. al* and Shapiro *et. al* studies on which we base our bias estimates should not have been affected by this problem). In the end, we are prepared to believe that the change in CPI procedures had at least some effect on the index. Thus, following the CPI's revision in procedure in 1997, we use an estimate of quality-change bias for the rapidly improving components of medical services that is lower by 0.7 percentage point per year; this is intended to reflect the CPI's improved ability to capture changes in the "mix" of inputs and is based on the difference between the rate of growth of a fixed-input price index for heart-attack treatment and an index with input weights that are updated every five years (see Cutler *et al.*, 2001, table 8.4). In all, then, we assume a bias in the CPI for medical services of 2.5 percentage points ($= 2/3 \times (4.5-0.7)$) since 1997. Given the

Cutler *et al.*, pp. 343.

weight for medical care services, this component contributes about 0.16 percentage point to quality-change bias in the overall CPI.

One additional issue involving the CPI for medical care that we did not factor into our bias estimate (because we have no presumption about which direction it goes) is the difficulty that arises in handling health insurance.³² The CPI does not currently price health insurance directly; rather, the weight of health insurance is distributed among the other medical strata (with a small weight remaining in the health insurance stratum itself to capture insurers' retained earnings) and its price is assumed to move with those components. Thus, any changes in health insurance provisions (such as higher copayment requirements) will be reflected in the CPI only indirectly and to the extent that these changes result from cost pressures on insurers (such as higher physicians' charges) that are themselves captured by the CPI. But those changes may be hard to capture by such indirect means. BLS is currently investigating the direct pricing of health insurance, which may help to address this issue.

Medical care commodities (pharmaceuticals): A number of papers have documented problems with official measures of pharmaceutical prices; in particular, the CPI's previous treatment of generic drugs--in which price differences between branded and generic drugs were treated as quality differences and purged from the index--probably induced significant upward bias in years past.

In January of 1995, the CPI changed its procedures regarding the entry of generic drugs into the index. Under the new procedure, a generic drug is permitted to replace the corresponding branded drug six months after the branded drug's patent has expired; the probability that this substitution occurs is proportional to the generic's relative expenditure share. The introduction of this new procedure reduces measured inflation for prescription pharmaceuticals by 0.4 percentage point per year over the period 1993 to 1997 (Stewart and Reed, 1999, p. 32).

³² See Bradley, 2000, for an analysis of some related issues involved in measuring the cost of living in the presence of insurance.

However, because generic drugs continue to gain market share six months after their introduction, they may not receive sufficient weight in the CPI even under the new procedure. For the seven antidepressants considered by Berndt, Cockburn, and Griliches (1996), generic drugs' market shares after one year ranged from 25 to 100 percent of their eventual (three-year) shares.³³ Similarly, in a study of two antibiotics, Griliches and Cockburn (1994) found that revenue shares for generic variants reached only 30 to 70 percent of their three-year shares after six months.

For our estimates of bias, we assume that the CPI's current methodology assigns a weight to generic drugs that is only one-quarter as large as the drugs' true final expenditure share (this figure is taken from the lower end of the range of one-year market-share estimates from Berndt *et al.*'s 1996 paper). Because the BLS estimated that the new procedures for generic drugs reduced the increase in prescription pharmaceutical prices by 0.4 percentage point per year (Stewart and Reed, 1999), we assume that the effect of fully incorporating generic drugs into the CPI would be four times larger than BLS estimated, or 1.6 percentage points per year, yielding a current bias estimate of 1.2 percentage point per year. We emphasize that this is a highly speculative estimate since it relies on the experience of the few drugs that have been studied, which may not be representative.

We have found only a modest amount of useable research that bears on the issue of pure quality bias in the CPI for prescription drugs. An attempt to apply hedonic techniques to anti-arthritis drugs (Cockburn and Anis, 2001) finds little correlation between these drugs' characteristics and their price, although characteristics do affect quantities. In their study of antidepressants, Berndt *et al.* (1996) estimate that the incremental effect of using hedonic adjustments to control for the introduction of new drugs reduces the average annual rate of price increase by anywhere from zero to 1.2 percentage points, depending on the subperiod that they consider. In addition, Suslow (1996) finds that a hedonically adjusted price index for three anti-ulcer drugs rises about 0.8 percentage point per year more slowly than a fixed-quantity price index for these drugs.

³³ The fact that the market share of most of the generic drugs in the Berndt *et al.* sample remains significantly below 100 percent even three years after their introduction suggests that consumers and/or physicians do not consider them to be perfect substitutes for branded drugs.

Finally, some internal Federal Reserve data are available that provide further information on quality-adjustment issues in pharmaceuticals. Federal Reserve analysts constructed a matched-model price index based on extremely detailed data across a very wide range of drugs.³⁴ The data therefore are far more complete than the data collected for use in the CPI, and new drugs are brought into the index very rapidly, when their share of the market is very small; this latter feature likely mitigates any new-items bias and obviates the need for hedonic adjustment (see Griliches, 1990, and Aizcorbe, Corrado, and Doms, 2000). The Federal Reserve index does not include any adjustments to treat generic drugs as directly comparable to their branded counterparts; nevertheless, a Laspeyres version of this price index rises about 1 percentage point per year less rapidly than the current-methods CPI for prescription drugs between 1995 and 1998.

We interpret these studies as pointing to some pure quality bias in the CPI for pharmaceuticals. We therefore add 1 percentage point of pure quality-change bias to our estimate of the bias that results from failing to fully incorporate generic drugs; this figure is roughly in line with the values found in the studies described above.³⁵ In total, then, we obtain a bias estimate of 2.2 percentage points per year for prescription drugs.

Quality-change bias likely exists for *non*prescription drugs as well. For example, the variety of drugs available over-the-counter appears to have risen notably over time as prescription drugs switch to OTC status and as new products and varieties become available, and the welfare benefit of this increased availability goes unmeasured.³⁶ Indeed, between 1995 and

³⁴ The data were purchased from IMS HEALTH, a private consulting firm. Although these data cover drugs purchased in the United States, they represent prices received by the drugs' distributors rather than by pharmacists, and in this sense do not correspond precisely to the concept measured by the CPI.

³⁵ One factor preventing us from choosing a larger bias estimate is the fact that it may not always be conceptually appropriate to include the benefit of the introduction of new drugs into the CPI. Consider, for example, the development of new antibiotics to treat drug-resistant infections that were previously treatable by older drugs; because the reduction in welfare owing to those drug-resistant infections is not included in the CPI, it could be misleading for the CPI to include the benefit of the new drugs. (As noted by the Advisory Commission, 1996, and Baker, 1998, a similar argument could be made regarding drugs developed to treat new conditions such as AIDS.)

³⁶ Any price decline that happens to be associated with a drug's moving from prescription to OTC status, however, would be captured by the CPI.

1998 the current-methods CPI for nonprescription drugs rose about 1/2 percentage point per year more rapidly than the internal Federal Reserve data mentioned above, in which new drugs are brought into the sample quickly (when their expenditure share is very small). We therefore assume 1/2 percentage point of bias per year (half the size of our estimate for pure quality improvement in prescription drugs) for this category of medical care commodities.

Electronic equipment: The CPI recently began extending hedonic quality adjustments to a number of types of electronic goods. In 1998, the CPI began using hedonic models to adjust computer prices, and in 1999, the CPI incorporated hedonic adjustments for television prices. More recently, hedonics are being used to adjust camcorders and 12 types of audio equipment (starting January 2000), and VCRs and DVD players (starting April 2000). Many analysts had advocated such a move, and the change overall has been a welcome development. Nevertheless, hedonics are no panacea, and we find reason to believe that measurement problems remain even in some of the areas in which these techniques are now employed.

In many cases, the new, hedonically adjusted indexes increase as rapidly as--or even more rapidly than--their unadjusted counterparts. (Computer prices are an important exception.) There are several possible explanations for this seemingly counterintuitive result. First, if manufacturers tend to link normal price increases to the introduction of new, higher-quality models, then a hedonically adjusted index may appropriately rise more rapidly than its unadjusted counterpart. This is because the hedonic adjustments allow the normal price increase to be identified and retained; under the old overlap procedure, the increase would have been purged from the index along with the portion of the price increase that truly reflects improved quality.

A second possibility, however, is that the hedonic procedures could be inadequate. Construction of a workable hedonic model is difficult, for it requires very complete information on product characteristics. Furthermore, even with a good hedonic model, BLS's implementation of hedonics may leave some fraction of quality-change bias intact; this is because BLS uses hedonic models to adjust prices only during item substitutions--that is, when an item is no longer available to be priced and a substitute is therefore chosen at the same outlet. But during routine sample rotation, when a new item is chosen at a new outlet, the usual overlap

method of quality adjustment is employed. For this reason, we refer to the BLS's procedure as "partial" hedonic adjustment. The evidence that we consider below suggests that for at least one good (televisions), confining the hedonic adjustments to item substitutions appears to miss a significant fraction of quality change.³⁷

BLS is experimenting with new procedures for item substitutions--called "directed" substitutions--that may help to reduce a portion of the remaining quality-change bias in such cases. Under directed substitution, item substitution (and hedonic adjustment) would occur after fixed intervals even if the old item remains available to be priced. Directed substitution should allow BLS to maintain a more up-to-date sample; it therefore ought to diminish the differences between existing items and those that replace them during sample rotation, and so might reduce the bias associated with the lack of hedonic quality adjustment at such times. Directed substitution is currently being implemented for personal computers, and related procedures are in train for prescription drugs (Lane, 2000).

To address this problem more directly, BLS could consider applying hedonic adjustment during sample rotation. A schematic description of this alternative, and its interaction with new-outlet bias, is outlined in table 6. Under current procedures, new items are linked into the CPI using the "overlap method" during sample rotations; this leaves the CPI susceptible both to quality-adjustment bias (because differences in item characteristics are not valued) and to new-outlet bias (because the entire price differential between outlets is attributed to differences in outlet quality).³⁸ Applying hedonics during sample rotation ought to help alleviate quality-adjustment bias--presuming, of course, that the hedonic estimates are accurate. Whether it also will help alleviate new-outlet bias will depend on whether outlet characteristics also can be priced in the hedonic regressions. If they cannot, then BLS would implicitly be making the

³⁷ Another problem arises because the goods that are substituted into the sample when item substitutions occur are typically chosen so as to be as similar as possible to the goods that are replaced. To the extent that pure price declines tend to be larger over the initial portion of a good's life cycle, this will cause the CPI to be biased upward. (Of course, this problem exists under both the new and old BLS procedures for quality adjustment.)

³⁸ With the "overlap" method, the price change between months $t-1$ and t is given by the old item at the old outlet, and the price change between months t and $t+1$ is given by the new item at the new outlet. No direct comparison is made between prices of the old and new items.

opposite assumption to the one currently made in allowing the full price differential across outlets to show through to the CPI; this might or might not be an improvement over current practice. But if outlet characteristics can be priced, then applying these hedonic estimates should help alleviate new-outlet bias as well as quality-adjustment bias.

Unfortunately, with the exception of Moulton, LaFleur, and Moses's (1999) work on televisions, BLS's methodology papers are of limited use in assessing the importance of the absence of complete hedonic adjustment: Comparable full-hedonic models usually cannot be estimated for these goods, and the studies typically cover a relatively short period of time (often less than one year). For want of better estimates, we assume that the bias for televisions applies to many other electronic goods as well, and take comfort in the fact that these items receive a very small weight in the CPI (so our estimates of quality-change bias for these goods contribute little to our overall bias estimate). Below, we discuss our estimates in more detail.

Televisions: In recent work, Moulton, LaFleur, and Moses (1999) find that constant-quality television prices rise about 2.1 percent more slowly than the published CPI for televisions from August of 1993 to August of 1997. (If the published television CPI is adjusted for the effects of geometric-mean aggregation, the difference over this period is 1.6 percentage points.) However, when the BLS's current procedures for applying "partial" hedonics to the television CPI are used, the resulting index grows a scant 0.1 percentage point more slowly than an index without hedonics. As discussed above, this appears to reflect the fact that BLS continues to employ the overlap method of quality adjustment during sample rotations; evidently, this method misses most of the improvement in the quality of newer televisions.³⁹ We therefore assume that complete quality adjustment would reduce the rate of change of the CPI for televisions by an additional 1.5 percentage points per year on average.

³⁹ It is also possible that the small impact of hedonics under the BLS's new procedures arises because the hedonic regressions do not control for changes in outlet (although this seems unlikely inasmuch as there was probably not a significant shift in outlets over the period considered here). Alternatively, the BLS's enumerators might be missing changes in quality when they deem item substitutions to be "comparable." However, Moulton *et al.* characterize the sample-rotation explanation as being most plausible.

Audio equipment: In January of 2000, the BLS began applying hedonic adjustments for audio equipment prices when there are item substitutions. Estimates presented by Kokoski, Waehrer, and Rozaklis (2000) for the preceding two years indicate that the BLS's new procedures yield a price index for audio equipment that rises about 0.1 percentage point slower than the unadjusted CPI.

Kokoski *et al.* also estimate a fully adjusted hedonic price index and find that it increases 1.5 to 10.2 percentage points more slowly than the new index, with the larger difference obtained by including vintage effects in the hedonic regression.⁴⁰ It is difficult, though, to know how to utilize these estimates. First, the data on prices and characteristics that are used to construct the hedonic models are taken from an industry source, as opposed to the CPI sample, and there is evidence that the two samples differ significantly. In particular, the measured price declines in the industry data appear to be much larger than those for the CPI price quotes, and it is therefore not clear whether the full hedonic indexes are even comparable to the CPI measures.⁴¹ Second, the significant influence of the vintage effects--which Kokoski *et al.* argue reflects unmeasured characteristics that are related to quality improvement--is disturbing, and suggests that the hedonic models are misspecified. A case can therefore be made that the estimates obtained from the two variants of the full hedonic-adjustment model should be heavily discounted.

On balance, we view the full hedonic indexes as providing a bit of evidence that there remains an upward bias in the CPI for audio equipment, but not a usable quantitative estimate of such a bias. We therefore fall back on assuming that the use of only "partial" hedonics leads to the same bias for audio equipment as for televisions, namely, 1.5 percentage points per year.

VCRs, camcorders, and DVD players: BLS's procedure for applying hedonic adjustments to the prices of these video products is also confined to instances of item substitution and not sample

⁴⁰ The dependent variable in the hedonic regressions is log price, which makes the full-hedonic index comparable to a geometric-means index. When we compare the full-hedonic and partial-hedonic indexes, therefore, we adjust the 1998 change in the partial-hedonic measure to reflect the estimated impact of geometric-means aggregation.

⁴¹ Several studies compare the CPI to hedonic price indexes that are estimated using independently gathered price data (such as prices taken from store catalogs). The results of Kokoski *et al.* underscore the importance of validating these alternative data by comparing them to the CPI.

rotation; hence, we might suspect that quality-change bias will persist as a result of employing overlap methods when samples are refreshed. However, it is difficult to determine the extent of the problem because BLS's methodology papers do not compare the published indexes for these goods to fully adjusted hedonic price indexes; in addition, the effects of the partial use of hedonics are themselves especially uncertain in these cases, because they are calculated over only a six- or seven-month period in 1999.⁴² Militating for a positive bias is the fact that the hedonic adjustments are specific to each type of equipment, so any benefits from the *introduction* of VCRs, camcorders, and DVD players have gone unmeasured in the CPI; to the extent that this rate of new-goods introduction is representative, the prospective degree of bias in the CPI will be greater for this category of goods. In all, we assume again that measured price changes for this category of the CPI will suffer from 1.5 percentage points of bias per year, the same as for televisions.

Computers: BLS began to apply (partial) hedonic adjustment to computers in 1998, and found that this change led to a 6.5 percentage points larger decline in computer prices that year than would have occurred under the old procedures. To determine whether these partial hedonics are adequate, we can compare them to the estimates in Aizcorbe, Corrado, and Doms (2000), which use extremely comprehensive, disaggregated, and timely data to construct price indexes for desktop and notebook computers. As these authors demonstrate, the data's timeliness is such that new items enter into their index with extremely small weight, reducing any new-items bias and obviating the need for direct hedonic adjustment. Comparison with the CPI is not completely straightforward because the CPI stratum includes peripheral equipment as well as computers, but a rough calculation suggests that the Aizcorbe *et al.* estimates would generate an

⁴² Thompson (2000) studies hedonic adjustments for VCR prices over a seven-month period in 1999 and finds that, using the CPI's procedures for partial hedonic adjustment, quality-adjusted prices fell 1.9 percentage points (annual rate) more slowly than the unadjusted CPI. (Liegey and Shepler, 1999, conduct a similar study in which the regression is based on data from Consumer Reports, and find a smaller effect for VCRs over 1997.) Shepler (2000a) considers camcorder substitutions over a six-month period and finds that a quality-adjusted price index falls about 0.4 percentage point faster (at an annual rate) than an unadjusted index. Finally, Liegey (2000a) finds that his hedonic model for DVD players would have had no effect on the CPI over the six-month period in 1999 that he studied, because there were almost no DVD item substitutions made during that time.

index that declined roughly 4 percentage points per year more rapidly than even the post-1998 CPI stratum.⁴³ This, together with the fact that the CPI does not do any hedonic adjustments to the peripheral equipment portion of the index (mostly printers and monitors) leads us to believe that an upward bias remains in this stratum on the order of 4 percentage points per year.

Transportation: The BLS makes a host of direct quality adjustments in constructing its price indexes for new motor vehicles, based on manufacturers' estimates of the cost of new and improved auto characteristics. Surprisingly, we know of no detailed studies evaluating the adequacy of BLS's procedures more recent than Gordon's (1990) careful study, which covered only the period through 1983. After factoring in the cost of pollution-control-related improvements as price increases, Gordon found that the CPI *understated* automobile price increases through 1983.⁴⁴ Nevertheless, the Advisory Commission argued that this downward bias was unlikely to persist after 1983 because there were fewer safety- and fuel-efficiency-related improvements since then, and furthermore, that additional quality-change bias was present because automobiles have become more durable over time. However, the direct quality adjustments made by BLS includes items related to durability, such as corrosion protection and longer-lived parts. This suggests that the Commission's estimate of quality-change bias may be too large (see Baker, 1998, pages 114 and 152, or Moulton and Moses, 1997, page 319 for additional discussion). Furthermore, as noted by Triplett, automobile manufacturers' estimates of the cost of new features may be in excess of their true value.⁴⁵ On balance, therefore, we assume no quality adjustment bias in either direction for new vehicles.

⁴³ The Aizcorbe *et al.* preferred (geometric-mean) index of desktop PC prices declined 44 percent over 1998 and the index of notebooks declined 34 percent. By contrast, the CPI stratum--which is roughly 80-90 percent desktops and 5 percent laptops--declined only 36 percent that year. Assuming zero price change for peripheral equipment yields the 4 percentage point figure cited in the text.

⁴⁴ See page 43 of the Advisory Commission report. The Advisory Commission also argued that mandated pollution-control devices for automobiles should be treated as an indirect tax (that is, as a price increase) rather than as a quality improvement. In 1999, the BLS adopted this view as well.

⁴⁵ Triplett's claim appears in a comment to Moulton and Moses (1997, p. 363), and is based on his experience at the Bureau of Labor Statistics.

Since late 1999, the CPI has included prices of leased vehicles as well as new vehicles (though the stratum will be published only beginning in 2002). This is a welcome development, both because leasing has become more common and because leases might some day form the basis of a rental-equivalence index for vehicles that are purchased. For present purposes, because BLS applies the same quality-adjustment factors to leased vehicles as to purchased vehicles, we assume no quality-adjustment bias for this index either. Similarly, the BLS now controls for used-car quality by applying the new-car quality adjustments with a lag of three years. (Prior to 1987, no such adjustments were made.) This strikes us as reasonable, and so we assume no quality bias for used cars as well.

Citing the convenience of built-in credit-card readers in gasoline pumps, the Advisory Commission estimated quality-change bias for motor fuels equal to 0.25 percentage point per year. Moulton and Moses (1997, pp. 319-320) use data on the diffusion of “pay-at-the-pump” technology in order to better assess the improvement in quality that has obtained from being able to purchase gasoline more conveniently, and obtain an estimate of quality-change bias for motor fuel that equals 0.1 percentage point per year. Even through pay-at-the-pump technology by now likely has diffused to most gasoline stations, some further retail innovations appear to be continuing, and we retain this small bias estimate going forward.

Good, Sickles, and Weiher (2001) attempt to compute a hedonic index for airfares, and find that their index rises 3.5 to 5.1 percentage points per year less rapidly than the CPI for airfares from 1979 to 1992. The difference appears to largely reflect BLS’s difficulties in capturing a representative mix of airfares in the sample rather than quality-change bias *per se*. However, the BLS significantly revised its methodology in 1991 so as to increase the share of discount fares in the index, and this suggests that Good *et al.*’s bias estimate is not applicable going forward. We therefore assume no quality-change bias in airfares (as did the Advisory Commission).⁴⁶ Similarly, for the other components of the CPI for transportation--largely motor

⁴⁶ Gordon (2000a, p. 25) argues that the CPI for air travel remains overstated because of its failure to measure frequent-flyer discounts (though he does not provide an estimate of the magnitude of the bias). However, note that going forward, frequent-flyer discounts will lead to bias in the rate of change of the CPI only if the share of air-travel expenditures accounted for by frequent-flyer fares rises.

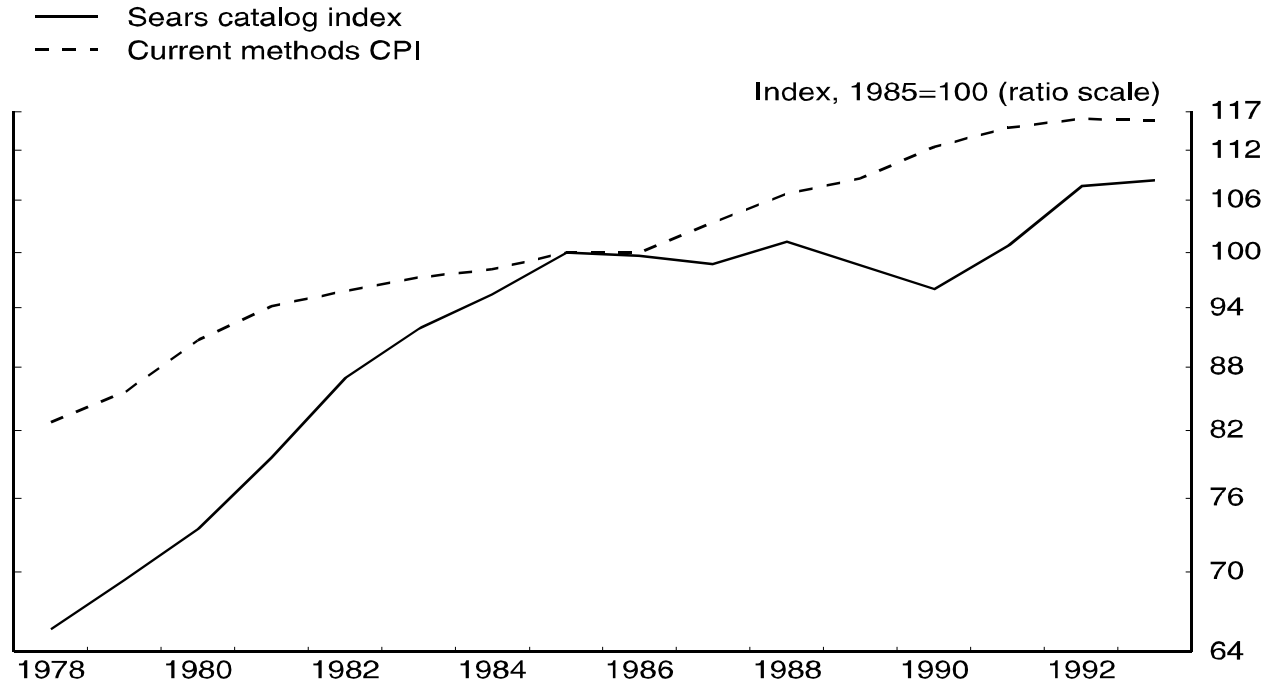
vehicle insurance and repair services--we have neither any evidence nor intuition that the CPI is biased in either direction.

Apparel: Analyses of the CPI for apparel are limited to two studies by Robert Gordon, both based on comparisons with apparel prices as listed in Sears catalogs. We are uncomfortable in drawing conclusions about overall apparel prices from prices that are derived from a single outlet, which are unlikely to be representative of the nation as a whole: Indeed, as Gordon (1996, p. 15) notes, the Sears catalog is likely to be more heavily weighted to standard utilitarian items, whose prices may behave differently from other items. Nevertheless, Gordon's results are the most comprehensive set that we have, and it is well worth examining them.

As shown in figure 1, over the period from 1978 to 1993--the longest period for which both the current-methods CPI and Gordon's (1996) index is available--Gordon's catalog-based measure rises more rapidly than the current-methods CPI.⁴⁷ (Extrapolating the current-methods CPI backward prior to 1978 to generate an even longer comparison would yield a similar conclusion.) The Advisory Commission put more weight on the years since 1985, during which the catalog-based index rose more slowly than the CPI for apparel (though somewhat less so relative to the current-methods index). This led them to assume quality-change bias of 1.0 percentage point per year in the CPI for apparel. Whether the recent period is likely to be a better indicator of the future is hard to say, but we see some merit in the longer comparison, especially given that the existence of the current-methods CPI allows that comparison to be made relatively cleanly. Furthermore, Gordon's (2000b) more recent work on apparel prices suggests that the Sears catalog index, which is a matched-model index that "links in" new items, may itself importantly understate inflation by failing to capture price increases that occur when new items are introduced. (The CPI began using hedonic methods in 1991 to correct for precisely this problem.)

⁴⁷ In addition to the implementation (noted below) of hedonic adjustments, the current-methods CPI utilizes a geometric-means aggregation formula, which also significantly affects the rate of change of the apparel index. Note that Gordon's catalog-price index aggregates price quotes across items using geometric means with unit weights (Gordon, 1996, p. 21).

Figure 1
Indexes of Apparel Prices



Thus, to the extent that the catalog-based index provides an independent check on the CPI for apparel, the CPI may more likely be understating the rate of price change for these goods rather than overstating it. At the same time, though, Gordon argues that the Sears catalog price series might be an overestimate of economy-wide price changes in that Sears was losing market share during the latter years of his study. Putting this all together, and given our doubts about the representativeness of the catalog-based index, we are inclined to assume no quality-change bias for this component of the CPI under current procedures.

Other durable goods: The BLS recently introduced hedonic methods to adjust the prices of a variety of household appliances, including microwave ovens (starting July 2000), refrigerators/freezers (July 2000), and clothes washers and dryers (October 2000).⁴⁸ These items

⁴⁸ These are described in Liegey (2000b) for microwaves, Shepler (2000b) for refrigerators, and Liegey (2000c) for clothes dryers. These papers fit the hedonic regressions to an expanded set of CPI data (that is, a set of additional quotes that are obtained using the same sampling techniques that are employed in the published CPI). They therefore mitigate the potential comparability problem that could arise from using price data from an outside source.

make up about 75 percent of the “major appliances” stratum. Introducing these hedonic estimates into the CPI--using the BLS’s “partial” hedonic approach of utilizing the hedonics to measure changes in characteristics during item substitutions but not sample rotations--reduced the rate of increase of the index for this stratum by 0.8 percentage point per year in total. Unfortunately, the estimates for each of these items were based on an analysis of only several months of CPI data, so they provide at best a very rough idea of the effect of using these hedonics over longer periods of time. Moreover, none of these studies computes a full hedonic price index. Hence, while we suspect that BLS’s new procedures are not fully correcting for the presence of quality-change bias, we have no way of assessing the magnitude of any remaining bias. We therefore use the same estimate for these goods as for televisions (1.5 percentage points), based admittedly on nothing more than the idea that the rate of technological improvement for these goods may be similar.

For housefurnishings other than appliances, which includes items such as furniture and cookware, we maintain the Advisory Commission’s assumption that quality-change bias averages about 0.3 percentage point per year. We also assume a small amount of quality-change bias for personal care products in order to reflect quality improvements for small electrical appliances such as hair dryers (which make up about half of the personal care products category in “other goods and services”); and we assume that the quality-change bias for these goods is the same as for televisions without any hedonic adjustment, or 1.6 percentage points per year. We make a similar estimate for the toys category (which appears under the recreation heading) to capture any such effects in electronic games. These estimates are almost entirely subjective.

Food: Food constitutes more than 15 percent of the CPI. Nevertheless, the only careful analysis of the adequacy of the CPI for food that we could find is the Moulton and Moses (1997, pp. 312-14) response to the Advisory Commission’s estimates--and this analysis was limited to the index for fresh fruits and vegetables. The Advisory Commission identified increased seasonal availability of certain varieties of fruits and vegetables as a source of new-items bias that they guessed to be worth about 0.6 percentage point per year. Moulton and Moses pointed out two potential problems with this value. First, for fruits, expenditure data suggest that there has not been a sufficiently large increase in the consumption of seasonal varieties to support the

Advisory Commission's estimate.⁴⁹ In addition, Moulton and Moses report that virtually all of the increase in the seasonal availability of vegetables appears to have taken place prior to 1985. In light of this evidence, we find it most reasonable to assume no new-items bias in the fresh fruits and vegetables category of the CPI going forward.

For the remainder of the food items in the CPI, we are left in the realm of purely subjective judgments. The Advisory Commission cited the emergence of specialty and mid-market restaurants as a source of increased quality for food away from home. In addition, they argued that the introduction of supermarkets that include services that were previously available only in separate shops (such as delicatessens and butcher shops), as well as the increased availability of imported alcoholic beverages, should be considered quality improvements. We agree that these improvements were real and probably were not captured by the CPI. Nevertheless, the Commission's quantitative estimates of quality-change bias (namely, 0.3 percentage point per year) strike us as a little bit too generous, and we assume that quality-change bias in these categories (and alcoholic beverages as well) is 0.2 percentage point per year.

Finally, for packaged foods we note Hausman's (1997b) well-known study that ascribes a large consumer surplus to the introduction of new brands of breakfast cereal. While Hausman's study is an exemplar of how such estimates should be constructed, we are persuaded by Bresnahan's (1997) view that one of Hausman's key identifying assumptions--namely, that there are no brand-specific demand shocks (induced, say, by advertising)--is probably not defensible. As a result, we are somewhat skeptical of Hausman's estimate of consumer surplus, and are inclined to assume that any such new-items bias for foodstuffs is subsumed in our 0.2 percentage point estimate.

Education: The most important component of the CPI for education involves out-of-pocket expenditures for college tuition and for child care and nursery school. Regarding the former, a

⁴⁹ Moulton and Moses cite an approximation formula for the consumer surplus that results from new-good introduction that was derived by Hausman (1997a); this formula indicates that the expenditure share of fresh fruits and vegetables would have to have increased by 40 percent over the period considered by the Advisory Commission for their bias estimate to be correct.

recent paper by Schwartz and Scafidi (2000) employs hedonic techniques in order to construct a constant-quality price index for four-year colleges. The authors find a slight deterioration in the quality of four-year colleges over the period 1991 to 1995. However, their most interesting result is that accounting for financial aid leads to a pattern of college-price growth that is somewhat different from the CPI; this results mainly from a single year's observation (1994), which saw the introduction of HOPE scholarships. (With the exception of this one year, changes in the CPI for tuition and fees track Schwartz and Scafidi's index quite well.) We are inclined to agree that the CPI probably should measure college tuition net of financial aid, and note that the BLS has tentative plans to make this change to their procedures. However, we do not know by how much the average amount of financial aid has varied over time, and cannot, therefore, compute an adjustment to the CPI that controls for this.⁵⁰ Thus, we make no quality-bias adjustment to this component of the CPI. Regarding child care and nursery school, we know of no analyses of the adequacy of the CPI, and again we assume no quality bias for this component.

The CPI for education also includes a small weight on textbooks. This is another area where BLS recently began using hedonic adjustment to capture quality changes, in this case those associated with changes in the number of pages, changes toward or away from use of soft covers, and so on (Reese, 2000). We have no reason to believe there is any bias in this index.

Telephones: An oft-cited example of the CPI's failure to include new goods involves the belated introduction of cellular telephones into the index in 1998. Hausman (1997a) argues that as a result, the rate of change in the CPI for telecommunications services was biased upward by about 2.3 percentage points per year from 1988 to 1996. However, as Moulton and Moses (1997, pp. 321-322) pointed out, this estimate is likely to be too large because Hausman's analysis includes business as well as consumer use of cell phones. Using data from the Consumer Expenditure Survey, Moulton and Moses compute that failing to initially include this good in the CPI led to a cumulative new-items bias that is only about 40 percent as large as Hausman's estimate, or about 1 percentage point per year.

⁵⁰ In practice, the problem of measuring financial aid is likely to be extremely difficult; in addition, some types of financial aid take the form of government subsidies, which are outside the scope of the CPI.

Is this estimate of the bias from failure to introduce cell phones in a timely manner indicative of an ongoing bias in this category of the CPI? On the one hand, the spread of cell phones was probably the most important innovation in communications consumption in the past decade, making us somewhat hesitant to assume that a bias of this magnitude should be expected to continue in the future. On the other hand, we recognize that telecommunications is an area of especially rapid product innovation. In the end, we arbitrarily scale back a little the Moulton-Moses cell-phone estimate, and assume a bias of 0.8 percentage point per year in telecommunications services going forward. This estimate would take into account any unmeasured improvements in sound clarity, convenience associated with the spread of pay phones that accept credit cards, and so on, that the Advisory Commission used to justify their estimate of a 1 percentage point per year bias in telecommunications services.

Hall (1993) argues that the CPI for long-distance telephone calls overstates the true price of long distance because it fails to properly measure discount plans. As with Hausman's cell phone estimates, Hall's estimate appears to suffer from the inclusion of business expenditures on long distance; more importantly, however, the CPI changed its methodology for measuring long-distance services in the early 1990s, and now prices telephone calls in a way that better captures discounting plans.⁵¹ While BLS's new methodology is not perfect--respondents in some markets refuse to provide necessary data, citing confidentiality concerns--it probably mitigates much of the problem. Thus, we make no additional adjustments to incorporate Hall's estimates.

The internet: One of the most important new consumption items in recent years--and one whose expansion appears likely to continue for some time--has been the ability to connect to the internet. Prior to 1998, when the CPI began to include the fees of internet service providers, the omission of this service contributed to new-items bias. And, as the internet has expanded since then, these ISP fees can be viewed as having purchased a wider and wider variety of services, suggesting that the quality-adjusted monthly fees for internet use in fact may have been

⁵¹ In many cases, the BLS now uses an average revenue concept to price long-distance telephone calls. This change in methodology was introduced in response to the increased prevalence of discounting plans; note that it is also used in some cases to price cellular telephone usage.

declining sharply. BLS makes no adjustment to capture the quality improvement associated with this expansion of services.

The types of services provided by the internet can be grouped into two broad categories. First, some uses of the internet may themselves be new goods; for instance, surfing the web to obtain information on a hobby could be considered a recreational pursuit on its own. Second, existing goods or services may now be provided more efficiently or at lower cost by the internet; for example, one can now peruse the online version of a newspaper rather than purchasing a printed copy, or make a purchase online rather than by telephone or from a brick-and-mortar outlet.

The unmeasured welfare benefit that obtains from this latter set of internet-provided goods spans several categories of CPI bias, including outlet bias, upper-level substitution bias, and “pure” new-item or quality-change bias. Consider, for example, an online version of a newspaper. In this case, the internet can serve as a type of new outlet (purchasing the newspaper with an online subscription may be more convenient than going to a newsagent); as a higher-quality version of an existing good (hypertext links may improve the usefulness of the newspaper); and as a substitute for an existing good (one consumes the online paper--the subscription to which might be cheaper--in lieu of a paper copy).⁵² Similarly, one can now use e-mail in place of letters or telephone calls; in some cases, this reflects an improvement in the quality of these goods (for example, e-mail is faster and possibly cheaper than regular mail), while in other cases, it can be considered an entirely new good.

We have not found any studies that estimate the effects of the introduction and expansion of the internet on consumer welfare. But as a way of assessing its approximate impact we use the Hausman (1997a) formula for consumer surplus together with expenditure weights for internet services (which constitute most of “computer information processing services”) from the consumer expenditure survey. The approximate consumer surplus that obtains from a new good

⁵² We emphasize that this last source of bias should properly be considered substitution bias, even if the change in consumption involves using a “free” online news service in place of purchasing a newspaper. Adding an online news service to the internet acts to shift the demand curve for paper newspapers or magazines inward; ideally, the resulting reduction in the consumer’s expenditure on news services will be captured by a superlative index, and thus in our estimate of upper-level substitution bias. (The situation is analogous to the impact of videocassette rentals on purchases of film tickets.)

is equal to one-half of the current expenditure share for this good divided by the absolute value of the good's price elasticity of demand. In 1998, expenditure on these services amounted to about 0.1 percent of total expenditures. If we assume that the price elasticity for internet services is -0.5 and that internet services first became available five years earlier, then the new-goods bias from the introduction of the internet contributed 0.02 percentage point per year to bias in the overall CPI during those years.⁵³ We have every reason to believe that the internet will continue to expand, and we have built such a figure into our bias estimates going forward. Note that, given the small share of expenditures in this category, the magnitude of this contribution to overall bias implies that bias for the computer information processing services stratum itself is on the order of 19 percentage points per year.

Financial services: Personal financial services (included in the CPI's "other goods and services" expenditure category) is yet another area where we have found no detailed studies of the adequacy of the CPI. The Advisory Commission assumed that such innovations in the retail banking industry as ATMs and cash-management accounts induce 2 percentage points per year of quality-change bias in the CPI for personal financial services. While we have no way of assessing how reasonable this estimate is, we suspect that it should be lower to the extent that the rate of expansion of the specific innovations identified by the Commission may have slowed considerably (for example, ATM use is now ubiquitous). Of course, these developments may be indicative of further innovations yet to come. We arbitrarily assume a 1 percentage point bias for this category of expenditure going forward; in any event, these services receive a tiny weight in the CPI.

⁵³ Our assumed elasticity is similar to the elasticity for cellular-telephone usage estimated by Hausman (1997a) and is also similar to the elasticities for travel and communication (-0.5) and entertainment (-0.6) reported in Deaton and Muellbauer (1980a, p. 71) for a Rotterdam model of consumer demand. However, it is somewhat lower (in absolute value) than the price elasticities for transport and communication (-0.9 to -1.2) and other services (-0.7 to -0.9) estimated by Deaton and Muellbauer (1980b, p. 320) using an "almost-ideal" demand system.

6. An Estimate of Aggregate CPI Bias

For each source of bias listed in table 1, we present not only our point estimate but also a confidence interval that summarizes our view of the likely distribution around each estimate. To obtain such a confidence interval around the total bias, we follow Shapiro and Wilcox (1996) and specify explicit functional forms for the distributions of each source of bias; this allows us to aggregate them numerically. These distributions convey nothing more than our subjective degree of confidence about our estimates; we present them at some risk of creating a false sense of precision about the distributions.

To perform these calculations, we associate our confidence intervals with a 90 percent range. For lower-level substitution bias, weighting bias, and quality-adjustment/new-items bias, we view our confidence intervals as symmetric, and we formalize our distributions as being normal with means equal to our point estimates and with standard deviations such that we obtain 90 percent confidence intervals equal to the intervals listed in table 1. For upper-level substitution bias and new-outlet bias, our confidence intervals are skewed to the right. In these cases, we formalize our distributions as being the concatenation of the left and right halves of two normal distributions in which the two halves have different standard deviations set so as to generate the desired mean, 5 percent tail, and 95 percent tail. This assumption, though unusual, is transparent and it avoids the implication of some better-known skewed distributions (such as the lognormal) that the distribution be strictly positive. We construct a distribution for the overall CPI bias by computing the sum of the draws from each of the component distributions, and obtain a 90 percent confidence interval that ranges from 0.1 to 1.2 percentage points.⁵⁴ Note that, although our confidence intervals extend below zero for three of the five components of bias, it does not do so for our total bias estimate.

⁵⁴ In making this calculation, we assume that the biases are uncorrelated with one another. Shapiro and Wilcox (1996) assumed that the distributions for lower-level substitution bias, new-outlets bias, and the new-items portion of quality-adjustment bias--each of which relates to the propensity to substitute one item for another--are positively correlated with a correlation coefficient of 0.25. Making such an assumption here leads to only a minuscule effect on the confidence interval around our estimate of total bias. Similarly, our discussion of the use of PCE weights to aggregate quality-adjustment bias suggests that this bias may be positively correlated with weighting bias, but imposing such a correlation leads to essentially no change in the confidence interval.

Finally, we mention a few words about the confidence interval around our point estimate of quality-change/new-items bias. We specified this confidence interval as being wide--ranging from -0.1 to 0.8 percentage point per year--to convey the subjective nature of our bias estimates for a large portion of the CPI. But it is worth noting one factor that argues for a narrower confidence interval (and that led us to temper the size of our confidence interval somewhat). Because we build up this bias estimate by forming judgments about the bias for each expenditure item, one can imagine constructing a distribution for overall quality-change bias as the weighted average of bias distributions for each of those items. If one believes that these bias distributions are essentially independent--that is, that biases in the CPIs for televisions, hospital services, owners' equivalent rent, new vehicles, and so on, have little to do with one another--then the aggregate distribution will be much less dispersed than the distributions of the pieces.⁵⁵ One should not take this argument too far, as the biases for many of these items may not in fact be independent--for example, there may be a common element to the BLS's problems in measuring the prices of many different items. But neither should one get carried away in translating the substantial uncertainty about the degree of quality-change bias in many components of the CPI into uncertainty about overall quality-change bias.

7. Summary and Conclusions

In this paper, we survey the evidence bearing on measurement error in the CPI and provide our best estimate of the magnitude of CPI bias, which we find to be about 0.6 percentage point per year with a confidence interval that ranges from 0.1 to 1.2 percentage points. We judge the inability of the CPI to fully capture the welfare improvement from quality change and the introduction of new items as accounting for about half of this bias. Nevertheless, our overall bias estimate, and our estimate of quality-change and new-items bias in particular, is somewhat lower than that of many earlier studies. At the same time, we identify a "weighting" bias that

⁵⁵ This reflects the fact that the variance of uncorrelated random variables will be larger than the variance of their sample average. For example, suppose we assign to each of the 22 items in table 5 a normal distribution with a standard deviation that generates a 90 percent confidence interval that is 1 percentage point wide. Assuming these distributions to be independent generates an aggregate distribution with a confidence interval that is only 0.3 percentage point wide.

has not been previously discussed in the literature. We fully recognize the uncertainty surrounding our estimates, and in particular we have tried to be clear about the uncomfortably small degree of evidence supporting our estimates of quality-change and new-items bias for many categories of expenditure.

One reason our bias estimate is smaller than in several earlier studies is that the BLS has recently made a variety of improvements to its procedures. Although much of the low-hanging fruit has already been picked, further progress remains possible, and our study has highlighted several potential areas for additional improvement. These include utilization of alternative aggregation formulas in addressing substitution bias; improvement of the consumer expenditure survey and the possible augmentation of these data with other expenditure data in forming the weights in the CPI; utilization of outside sources of comprehensive, high-frequency price data (from check-out scanners or other sources); attempting to bring new items into the CPI more rapidly; expanding the use of directed substitution; applying hedonic adjustment during sample rotation as well as during item substitutions; and devoting continued attention to capturing quality changes in myriad areas including medical care. BLS and academic research is underway in most of these areas and should result in continued improvement in the accuracy of the CPI in the years to come.

Table 1
Estimates of CPI Bias

Category of Bias	Lebow-Roberts- Stockton (1994) ¹	Shapiro-Wilcox (1996) ²	Advisory Commission		This paper ⁵
			Report (1996) ³	GAO update (1999) ⁴	
Upper-level substitution	.1 - .2	.2 (.0 - .4)	.15	.1	.1 (.0 - .25)
Lower-level substitution	.3 - .4	.25 (.0 - .5)	.25	.05	.05 (-.15 - .25)
New outlets	.0 - .1	.1 (.0 - .2)	.1	.1	.05 (.0 - .20)
Weighting	--	--	--	--	.05 (-.05 - .15)
Quality change	.0 - .3	.25 (-.05 - .5)	.6	.55	.37 (-.08 - .82)
New items	.0 - .5	.2 (.0 - .4)			
TOTAL BIAS	0.4 - 1.5	1.0 (.6 - 1.5)	1.1 (.8 - 1.6)	.8	.62 (.1 - 1.2)

1. Lebow, Roberts, and Stockton did not specify a point estimate. Implicitly, one may consider the midpoint of their ranges to be their point estimates, with the possible exception of new items bias, for which 0.5 percent was called “surely an upper limit on this effect.”

2. Ninety percent confidence intervals are in parentheses, with the exception of the total bias, which is an eighty percent confidence interval.

3. Range on total bias is in parentheses.

4. Total bias is the mean of the Advisory Commission members’ estimates. Figures for the categories of bias are approximate.

5. Confidence intervals are in parentheses.

Table 2
Weights in the CPI and PCE (CPI scope)

	Relative importance weights, Dec. 1997		Difference	Ratio
	CPI	PCE-based		
Nondurable goods				
Meats, poultry, fish, eggs	2.6	2.4	0.2	1.08
Fruits and vegetables	1.4	1.3	0.1	1.04
Other food at home	5.6	6.7	-1.0	0.84
Food away from home	5.7	6.6	-0.9	0.86
Motor fuel	3.0	2.8	0.2	1.08
Heating oil	0.3	0.3	0.0	0.94
Apparel	4.9	7.2	-2.2	0.69
Tobacco	0.9	1.2	-0.3	0.73
Alcoholic beverages	1.0	2.5	-1.5	0.39
Medical commodities	1.2	0.7	0.5	1.68
Other nondurables	4.4	6.8	-2.4	0.65
Durable goods				
Motor vehicles	7.9	5.6	2.3	1.40
Computers	0.2	0.2	0.0	1.02
Audio/video equipment	0.9	1.2	-0.3	0.76
Other durables	3.4	5.1	-1.7	0.66
Services				
Natural gas	1.1	0.9	0.2	1.22
Electricity	2.6	2.0	0.6	1.30
Owners' equivalent rent	20.2	14.8	5.4	1.37
Tenants' rent	6.9	5.4	1.5	1.28
Lodging away from home	2.3	0.8	1.5	2.77
Medical services	4.4	5.8	-1.4	0.76
Tuition & school fees	2.4	2.4	0.0	0.99
Airfares	0.8	1.2	-0.4	0.68
Other services	15.7	16.0	-0.2	0.98
TOTAL	100.0	100.0		

Table 3
Alternative CPI using PCE-Based Weights
 (Percent changes/percentage points, annual rate)

	Published CPI (CE weights)	Alternative CPI (PCE-based weights)	Difference	Memo: Using 1986 Expenditures		
				CE weights ¹	PCE-based weights	Difference
1988	4.14	4.00	.14	3.98	3.92	.06
1989	4.82	4.85	-.03	4.69	4.71	-.02
1990	5.40	5.42	-.01	5.16	5.26	-.10
1991	4.21	4.16	.05	3.93	4.07	-.14
1992	3.01	2.91	.10	2.84	2.87	-.03
1993	2.99	2.87	.13	2.80	2.80	.00
1994	2.56	2.41	.15	2.58	2.39	.18
1995	2.83	2.62	.22	2.75	2.57	.18
1996	2.95	2.84	.11	2.85	2.76	.09
1997	2.29	2.25	.04	2.11	2.16	-.05
1987-1997	3.52	3.43	.09	3.36	3.35	.02
1998	1.56	1.45	.11			
1999	2.21	2.11	.10			
2000	3.36	3.13	.24			
1997-2000	2.37	2.23	.15			
1987-2000	3.25	3.15	.10			

1. From Cage and Jackman (1999), Table A-1

Table 4
Estimates of Quality-Change and New-Items Bias in the CPI, 2001
(Percentage points per year)

Adjusted PCE shares, 1998	Expenditure category	Our estimates		Memo: Advisory Commission	
		Estimated bias	Cont. to total	Estimated bias	Cont. to total
18.15	Food	.2	.03	.3	.05
1.16	Fresh fruits and vegetables	.0		.6	
8.16	Other food at home	.2		.3	
6.40	Food away from home	.2		.3	
2.43	Alcohol	.2		.15	
30.33	Housing	.2	.05	.2	.07
4.90	Tenants' rent	-.2		.25	
14.29	Owners' equivalent rent	.3		.25	
1.66	Other shelter	.0		.25	
.39	Appliances	1.5		3.0	
2.14	Housefurnishings	.3		.33	
6.95	Fuels and utilities, other housing	.0		.0	
7.48	Apparel	.0	.00	1.0	.07
14.36	Transportation	.0	.00	.3	.04
6.10	New and used vehicles	.0		.59	
2.38	Motor fuel	.1		.25	
5.87	Airfares, mv parts and repair, insurance	.0		.0	
7.52	Medical care	2.4	.18	2.8	.21
.65	Prescription drugs	2.2		2.0	
.35	Nonprescription drugs	.5		1.0	
6.53	Medical care services	2.5		3.0	
9.38	Recreation	.3	.03	.9	.09
.45	Televisions	1.5		4.0	
.27	Other video equipment	1.5		4.0	
.34	Audio equipment	1.5		4.0	
.89	Toys	1.6		2.0	
7.42	Other recreation	.0		.2	
5.76	Education and communication	1.0	.06	1.8	.10
2.33	Education	.0		.0	
2.51	Telephone (incl. cellular)	.8		1.5	
.47	Personal computers & peripherals	4.0		15.0	
.10	Personal computer services (internet)	19.0		--	
.35	Postage, other info. processing	.0		.0	
7.01	Other goods and services	.3	.02	.7	.05
1.21	Personal care products	.8		1.6	
1.22	Personal financial services	1.0		2.0	
.31	Apparel services	.0		1.0	
4.27	Other	.0		.0	
100.000	TOTAL	.37		.69	

Table 5
Quality-Change/New-Items Bias
 Items for which our bias estimate is . . .

. . . Based on at least a moderate degree of hard evidence			. . . Based on a small or inadequate degree of evidence			. . . Almost entirely subjective		
Item	Weight	Cont. to total bias ¹	Item	Weight	Cont. to total bias ¹	Item	Weight	Cont. to total bias ¹
Fresh fruits and vegetables	1.16	.000				Other food and alcohol	16.99	.034
Tenants' rent	4.90	-.010	Owners' equivalent rent	14.29	.043	Other shelter	1.66	.000
						Appliances, housefurnishings	2.53	.012
						Fuels and utilities, other housing	6.95	.000
			Apparel	7.48	.000			
			New and used vehicles	6.10	.000	Other transportation	5.87	.000
			Motor fuel	2.38	.002			
			Medical care services	6.53	.163			
			Pharmaceuticals	1.00	.016			
Televisions	.45	.007	Audio & video equip.	.61	.009	Other recreation (toys)	8.31	.014
Personal computers	.47	.019	Computer services (internet)	.10	.020	Education	2.33	.000
						Telephone	2.51	.020
						Postage, other info. proc.	.35	.000
						Other goods and services	7.01	.023
SUM	6.98	.02	SUM	38.49	.25	SUM	54.51	.10

1. Percentage points per year.

Table 6
Applying Hedonics during Sample Rotation
 (Methods for handling differences in item and outlet characteristics)

	Item substitution	Sample rotation		
	Current BLS Practice	Current BLS Practice	Apply Hedonics during Sample Rotation	
			Hedonics do not include outlet characteristics	Hedonics include outlet characteristics
Item differences	Hedonics ¹	Overlap method ² → quality bias	Hedonics ¹	Hedonics ¹
Outlet differences	N.A.	Overlap method ² → outlet bias	Direct comparison ³ → opposite outlet bias	Hedonics ¹

1. Prices of old and new items are compared using estimates of the value of different item (and, possibly, outlet) characteristics from hedonic regressions.

2. The price change between months t-1 and t is given by the old item at the old outlet, and the price change between months t and t+1 is given by the new item at the new outlet. No direct comparison is made between prices of the old and new items.

3. Prices of old and new items are compared directly.

References

Advisory Commission to Study the CPI, "Final Report of the Advisory Commission to Study the Consumer Price Index" (Washington: U.S. Government Printing Office), December 1996.

Aizcorbe, Ana, Carol Corrado, and Mark Doms, "Constructing Price and Quantity Indexes for High Technology Goods," mimeo, Federal Reserve Board, July 2000.

Baker, Dean, ed., *Getting Prices Right: The Debate over the Consumer Price Index* (Armonk, NY: M. E. Sharpe), 1998.

Berndt, Ernst R., Susan H. Busch, and Richard G. Frank, "Treatment Price Indexes for Acute Phase Major Depression," in Cutler and Berndt, eds., *Medical Care Output and Productivity* (Chicago: University of Chicago Press), 2001, pp. 463-508.

Berndt, Ernst, Ian Cockburn, and Zvi Griliches, "Pharmaceutical Innovations and Market Dynamics: Tracking Effects on Price Indexes for Antidepressant Drugs," *Brookings Papers on Economic Activity (Microeconomics)*, 1996, 133-199.

Berndt, Ernst, David Cutler, Richard Frank, Zvi Griliches, Joseph Newhouse, and Jack Triplett, "Price Indexes for Medical Care Goods and Services: An Overview of Measurement Issues," in Cutler and Berndt, eds., *Medical Care Output and Productivity* (Chicago: University of Chicago Press), 2001, pp. 141-200.

Bils, Mark, and Peter J. Klenow, "Quantifying Quality Growth," mimeo, University of Rochester, October 2000.

Blow, Laura, and Ian Crawford, "The Cost of Living with the RPI: Substitution Bias in the UK Retail Prices Index," *Economic Journal* 111, June 2001, F357-F382.

Bradley, Ralph, Bill Cook, Sylvia E. Leaver, and Brent R. Moulton, "An Overview of Research on Potential Uses of Scanner Data in the U.S. CPI," paper presented at the Third Conference of the International Working Group on Price Indexes, Statistics Netherlands, Voorburg, April 16-18, 1997.

Branch, E. Raphael, "The Consumer Expenditure Survey: A Comparative Analysis," *Monthly Labor Review*, December 1994, 47-55.

Bresnahan, Timothy, "Comment" to Hausman, in Bresnahan and Gordon, eds., *The Economics of New Goods* (Chicago: University of Chicago Press), 1997, 237-247.

Cage, Robert A., and Patrick C. Jackman, "Estimated Upper Level Substitution in the Consumer Price Index, 1987-97," mimeo, U.S. Bureau of Labor Statistics, July 1999.

Cockburn, Iain M., and Aslam H. Anis, "Hedonic Analysis of Arthritis Drugs," in Cutler and Berndt, eds., *Medical Care Output and Productivity* (Chicago: University of Chicago Press), 2001, pp. 439-61.

Conference Board, "Measuring Prices in a Dynamic Economy: Re-Examining the CPI," Study Group on the Consumer Price Index, 1999.

Crone, Theodore, Leonard Nakamura, and Richard Voith, "Measuring Housing Services Inflation," *Journal of Economic and Social Measurement*, forthcoming, 2001.

Cutler, David, Mark McClellan, Joseph Newhouse, and Dahlia Remler, "Pricing Heart Attack Treatments," in Cutler and Berndt, eds., *Medical Care Output and Productivity* (Chicago: University of Chicago Press), 2001, pp. 305-62.

Deaton, Angus, "Getting Prices Right: What Should Be Done?" *Journal of Economic Perspectives* 12(1), Winter 1998, pp. 37-46.

Deaton, Angus, *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy* (Baltimore: Johns Hopkins University Press), 1997.

Deaton, Angus, and John Muellbauer, *Economics and Consumer Behavior* (New York: Cambridge University Press), 1980a.

Deaton, Angus, and John Muellbauer, "An Almost Ideal Demand System," *American Economic Review* 70, June 1980b, 312-326.

Feenstra, Robert C., and Matthew D. Shapiro, "High-Frequency Substitution and the Measurement of Price Indexes," NBER working paper #8176, March 2001.

Fixler, Dennis, "Comments on Chapters 2 and 3," pp. 103-108 in Jack Triplett, ed., *Measuring the Prices of Medical Treatments* (Washington: Brookings Institution), 1999.

Fixler, Dennis, and Ted Jaditz, "An Examination of the Difference Between the CPI and the PCE Deflator," mimeo, Bureau of Labor Statistics, December 29, 1997.

Frank, Richard, Ernst Berndt, and Susan Busch, "Price Indexes for the Treatment of Depression," Chapter 3 in Jack Triplett, ed., *Measuring the Prices of Medical Treatments* (Washington: Brookings Institution), 1999.

General Accounting Office, "Consumer Price Index: Update of Boskin Commission's Estimate of Bias" (Washington: Government Printing Office), 2000.

Gieseeman, Raymond, "The Consumer Expenditure Survey: Quality Control by Comparative Analysis," *Monthly Labor Review*, March 1987, 8-14.

Good, David, Robin Sickles, and Jesse Weiher, "A Hedonic Price Index for Airline Travel," mimeo, May 2001.

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

Gordon, Robert, "The Sears Catalog Revisited: Apparel and Durable Goods," mimeo, Northwestern University, April 1996.

Gordon, Robert, "The Boskin Commission Report and Its Aftermath," mimeo, Northwestern University, 2000a.

Gordon, Robert, "Apparel Prices and the Hulton/Bruegel Paradox," mimeo, Northwestern University, August 1, 2000b.

Greenlees, John S., "The U.S. CPI and the Cost-of-Living Objective," Bureau of Labor Statistics mimeo, November 2001.

Greenlees, John S., "Random Errors and Superlative Indexes," Bureau of Labor Statistics Working Paper 343, March 2001.

Greenlees, John S., "Consumer Price Indexes: Methods for Quality and Variety Change," *Statistical Journal of the United Nations ECE*, 17 (2000), pp. 59-74.

Greenlees, John S., "Expenditure Weight Updates and Measured Inflation," in B.M. Balk, ed., *Proceedings of the Third Meeting of the International Working Group on Price Indices* (Statistics Netherlands), 1998, pp. 57-64.

Griliches, Zvi, "Hedonic Price Indexes and the Measurement of Capital and Productivity," in Ernst R. Berndt and Jack E. Triplett, eds., *Fifty Years of Economic Measurement* (Chicago: University of Chicago Press), 1990.

Griliches, Zvi, and Ian Cockburn, "Generics and New Goods in Pharmaceutical Price Indexes," *American Economic Review* 84, December 1994, 1213-1232.

Hall, Robert, "Long Distance: Public Benefits from Increased Competition," mimeo, October 1993.

Hamilton, Bruce W., "Using Engel's Law to Estimate CPI Bias," *American Economic Review* 91, June 2001, 619-630.

Hausman, Jerry, "Cellular Telephone, New Products, and the CPI," National Bureau of Economic Research Working Paper 5982, 1997a.

Hausman, Jerry, "Valuation of New Goods under Perfect and Imperfect Competition," Chapter 5 in Bresnahan and Gordon, eds., *The Economics of New Goods* (Chicago: University of Chicago Press), 1997b.

Houthakker, H. S., and Lester D. Taylor, *Consumer Demand in the United States: analyses and Projections* (Cambridge, MA: Harvard University Press), 1970.

Kokoski, Mary, Keith Waehrer, and Patricia Rozaklis, "Using Hedonic Methods for Quality Adjustment in the CPI: The Consumer Audio Products Component," mimeo, Bureau of Labor Statistics, April 2000.

Krueger, Alan B., and Aaron Siskind, "Using Survey Data to Assess Bias in the Consumer Price Index," *Monthly Labor Review*, April 1998, 24-33.

Lane, Walter, "Addressing the New Goods Problem in the Consumer Price Index," mimeo, Bureau of Labor Statistics, June 2000.

Lebow, David E., John M. Roberts, and David J. Stockton, "Monetary Policy and 'The Price Level'," mimeo, Board of Governors of the Federal Reserve System, 1994.

Liegey, Paul, "Developing a Hedonic Regression Model for DVD Players in the U.S. CPI," mimeo, Bureau of Labor Statistics, 2000a.

Liegey, Paul, "Hedonic Quality Adjustment Methods for Microwave Ovens in the U.S. CPI," mimeo, Bureau of Labor Statistics, 2000b.

Liegey, Paul, "Hedonic Quality Adjustment Methods for Clothes Dryers in the U.S. CPI," mimeo, Bureau of Labor Statistics, 2000c.

Liegey, Paul, and Nicole Shepler, "Adjusting VCR Prices for Quality Change: A Study Using Hedonic Methods," *Monthly Labor Review*, September 1999, 22-37.

McClelland, Robert, and Marshall Reinsdorf, "Small Sample Bias in Geometric Mean and Seasoned CPI Component Indexes," U.S. Bureau of Labor Statistics Working Paper 324, August 1999.

Moulton, Brent, "Issues in Measuring Price Changes for Rent of Shelter," mimeo, U.S. Bureau of Labor Statistics, April 1997.

Moulton, Brent, Timothy LaFleur, and Karin Moses, "Research on Improved Quality Adjustment in the CPI: The Case of Televisions," mimeo, Bureau of Labor Statistics, 1999.

Moulton, Brent, and Karin Moses, "Addressing the Quality Change Issue in the Consumer Price Index," *Brookings Papers on Economic Activity*, 1997:1, 305-366.

Nordhaus, William, "Quality Change in Price Indexes," *Journal of Economic Perspectives*, Winter 1998, 59-68.

Pakes, Ariel, "New Goods, Hedonics, and Price Indices, with an Application to PCs," mimeo, May 2001.

Reese, Mike, "Hedonic Quality Adjustment Methods for College Textbooks in the U.S. CPI," mimeo, Bureau of Labor Statistics, 2000.

Reinsdorf, Marshall B., "Using Scanner Data to Construct CPI Basic Component Indexes," *Journal of Business and Economic Statistics*, April 1999, 152-60.

Reinsdorf, Marshall B., "The Effect of Outlet Price Differentials on the U.S. Consumer Price Index," in M. Foss, M. Manser, and A. Young, eds., *Price Measurements and Their Uses*, NBER Studies in Income and Wealth, Vol. 57 (Chicago: University of Chicago Press), 1993, 227-60.

Richardson, David, "Scanner Indexes for the CPI," mimeo, Bureau of Labor Statistics, 2000.

Schwartz, Amy Ellen, and Benjamin Scafidi, "Quality Adjusted Price Indexes for Four Year Colleges," mimeo, June 2000.

Shapiro, Irving, Matthew D. Shapiro, and David W. Wilcox, "Measuring the Value of Cataract Surgery," in Cutler and Berndt, eds., *Medical Care Output and Productivity* (Chicago: University of Chicago Press), 2001, pp. 411-37.

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

Shapiro, Matthew D., and David W. Wilcox, "Alternative Strategies for Aggregating Prices in the CPI," *Federal Reserve Bank of St. Louis Review*, May/June 1997, 113-25.

Shepler, Nicole, "Developing a Hedonic Regression Model for Camcorders in the U.S. CPI," mimeo, Bureau of Labor Statistics, 2000a.

Shepler, Nicole, "Developing a Hedonic Regression Model for Refrigerators in the U.S. CPI," mimeo, Bureau of Labor Statistics, 2000b.

Stewart, Kenneth, and Stephen Reed, "Consumer Price Index Research Series Using Current Methods," *Monthly Labor Review*, June 1999.

Suslow, Valerie, "Measuring Quality Change in the Market for Anti-Ulcer Drugs," Chapter 4 in Richard Helms, ed., *Competitive Strategies in the Pharmaceutical Industry* (Washington: AEI Press), 1996.

Tellis, Gerard J., "The Price Elasticity of Selective Demand: A Meta-Analysis of Econometric Models of Sales," *Journal of Marketing Research* 25, November 1988, 167-77.

Thompson, William, "Developing a Hedonic Regression Model for VCRs in the U.S. CPI," mimeo, Bureau of Labor Statistics, 2000.

Triplett, Jack E., "Should the Cost-of-Living Index Provide the Conceptual Framework for a Consumer Price Index?" *Economic Journal* 111 (472), June 2001, F311-34.

Triplett, Jack E., "Measuring Consumption: The Post-1973 Slowdown and the Research Issues," *Federal Reserve Bank of St. Louis Review*, May/June 1997, 9-42.

Appendix A: Adjusting the PCE data to match the coverage of the CPI

This appendix details three adjustments we make to PCE to put it on a conceptual basis similar to that of the CPI. Each is described in turn.

1) We first adjusted PCE by excluding items that are outside the scope of the CPI, mainly expenditures by nonprofit institutions and the portion of medical and educational expenses that are not made out of pocket. The latter adjustments in particular required making some rough assumptions. Specifically;

Goods: Exclude food produced and consumed on farms and food furnished to employees (including military), fuel produced and consumed on farms, and apparel provided to military personnel.

Medical care services: Subtract government transfers to persons for medical care, employer contributions for employees' health insurance and workers' compensation, and expenditures by foreigners. These subtractions are made proportionally across medical goods and services.

Education services: Exclude foundations and nonprofit research organizations. Subtract government transfers for education, and then subtract 50 percent of the remainder, from private higher and lower education.

Other services: Exclude rental value of farm housing, religion and welfare, net foreign travel (except for Americans' passenger fares), domestic services paid in-kind, imputed financial service charges, brokerage fees, expenses of handling life insurance, casino gambling, parimutuel net receipts, lotteries, expenditures by labor unions, professional associations, and clubs and fraternal organizations.

2) We then made two adjustments to specific PCE expenditure categories.

Tenants' rent and utilities. In roughly 30 percent of rental units in the CPI, the contract rent includes at least some utility costs (fuel and natural gas for heating or hot water, electricity, water, or sewer charges). But the tenants' rent category in PCE is based on a pure rent concept. To make the PCE data comparable to the CPI, we boost PCE expenditures on tenants' rent by 6 percent, with an equivalent amount subtracted proportionally from the aforementioned utility categories. This adjustment is based on conversations with BLS analysts, and can loosely be understood as 30 percent of renters having 20 percent of rent covering utilities.

Major appliances and owners' equivalent rent. Because services provided by owner-occupied housing in the CPI are based on a rental-equivalence concept, BLS reduces the CPI weight for major household appliances (that typically are included with the house) to exclude purchases by owners above those purchases that they estimate would have been made by renters. In addition, the rental values on owner-occupied housing that are estimated by respondents to the CE survey presumably include the service flow from these appliances as well, while the rental equivalence

concept in PCE covers the pure rent only. In constructing our PCE weights, we therefore adjusted both the “other durable goods” and owners’ equivalent rent categories to estimate the effect of putting the PCE data on the CPI concept. Specifically, we (a) subtracted 39 percent of PCE for major household appliances and 38 percent of PCE for floor coverings from other durable goods expenditures, and (b) added into owners’ equivalent rent expenditures an imputed service flow from these items equal to 2.3 percent of the pure rent. These two adjustments, which are of similar magnitude in dollar terms, are calculated as follows.

(a) The 39 percent figure is based on unpublished detail from the CE survey on expenditures on major appliances made by owners and renters in 1992 (the only year we had available), along with adjustments for the share of owners’ expenditures on five major appliances that BLS discounts in constructing the CPI weights. (We are grateful to BLS for providing these adjustment factors.) A similar calculation implies that 82 percent of expenditures on wall-to-wall carpeting should be subtracted, but because this item is not shown separately in PCE (it is included with other floor coverings), we scaled that estimate by 46 percent, the share of wall-to-wall carpeting in overall floor coverings in the 1992 CE survey, to obtain our 38 percent adjustment. (BLS also makes adjustments in the CPI weight for household maintenance and supplies other than wall-to-wall carpeting, but these items are largely omitted from PCE so they necessitated no adjustment to our estimates.)

(b) The adjustment to owners’ equivalent rent is based on BEA’s estimates of the stock of “kitchen and household appliances.” We calculated the service flow from this stock using a user-cost formula (assuming a 10 percent depreciation rate, a 3 percent real interest rate, and no capital gains or losses on the asset); this service flow is then scaled down by 39 percent to account for the share of appliances owned by renters (described above) and is scaled down by another 8 percent to account for the fact that this category includes small household appliances as well as major appliances. The resulting service flow averages 1.9 percent of PCE for owners’ equivalent rent in recent years. BEA does not have capital stock estimates for floor coverings, so we scaled the service flow upward proportionally given expenditures on these items to obtain the 2.3 percent figure that we employ.

3) Finally, we adjusted the data to account for the fact that the CPI-U covers urban households only (about 87 percent of the total), while PCE is national. Specifically, we multiplied the PCE data for the various expenditure categories in table 2 by the ratios of urban/total expenditures per household, calculated from the 1994 CE survey (the midpoint of the 1993-95 weights currently underlying the CPI). These are shown in table A-1.

Table A-1
Ratio of Urban/Total Expenditures per Household, 1994 CE Survey

Expenditure category	Urban/Total	Comments
Meats, poultry, fish, eggs	1.001	
Fruits and vegetables	1.011	
Other food at home	0.990	
Food away from home	1.019	
Motor fuel	0.959	
Heating oil	0.816	
Apparel	1.044	
Tobacco	0.954	
Alcoholic beverages	1.043	
Medical commodities	0.952	
Other nondurables	0.986	Household equipment, pets & toys, other entertainment goods, reading, 1/2 of personal care
Motor vehicles	0.947	
Computers	1.026	Ratio for TV/radio/sound equipment used
Audio/video equipment	1.026	TV, radios, sound equipment
Other durables	1.035	Household furnishings, less adjustments for major appliances and floor coverings
Natural gas	1.099	
Electricity	0.967	
Owners' equivalent rent	1.032	
Tenants' rent	1.098	
Lodging away from home	1.058	
Medical services	0.976	
Tuition & school fees	1.076	Education
Airfares	1.076	Public transportation
Other services	1.031	Water and telephone, household operations, m.v. expenses (excl. finance charges), 1/2 of personal care, fees and admissions, miscellaneous

Appendix B: Decomposition of the CPI Bias

In this appendix we present formulas showing how the bias in the CPI can be decomposed into the components listed in table 1. These decompositions will help to clarify some of the calculations that we employ.

Overall CPI bias can be defined as

$$\text{Bias} = \frac{w_l p}{\text{Published CPI}} - \frac{w_s^* p^*}{\text{True COLI}} \quad (1)$$

where w and p are the weights and price changes used in constructing the CPI, w^* and p^* are their true values, the subscript l (for Laspeyres) represents the aggregation formulas used in the CPI, and the subscript s (for superlative) represents the aggregation formulas that would ideally be used in constructing a true cost-of-living index.⁵⁶

The bias can be decomposed into components reflecting differences between p and p^* (quality-adjustment/new-items bias and new-outlet bias), between w and w^* (weighting bias), and between w_l and w_s (upper- and lower-level substitution biases). Such decompositions are not unique; for example, each of the following decompositions are correct:

$$\text{Bias} = w_s^* (p - p^*) + (w_s - w_s^*) p + (w_l - w_s) p \quad (2a)$$

$$= w_s (p - p^*) + (w_s - w_s^*) p^* + (w_l - w_s) p \quad (2b)$$

$$= w_s^* (p - p^*) + (w_l - w_l^*) p + (w_l^* - w_s^*) p \quad (2c)$$

Given this notation, the calculations we use in the text can be expressed as follows.

Upper-level substitution bias is calculated as the difference between a modified Laspeyres and a superlative aggregate of CPI prices, based on weights derived from the CE survey; lower-level substitution bias is estimated judgmentally, but again the thought experiment involves taking CPI-based weights and prices and modifying only the aggregation formula. Thus, our estimates of substitution biases correspond to $(w_l - w_s) p$ in the notation above. Weighting bias is based on the supposition that PCE provides a more accurate set of expenditures than the CE survey, and we base our calculations on an aggregate of CPI prices using PCE-based weights, or $(w_l - w_l^*) p$. Finally, we aggregate our detailed estimates of quality-change/new-items bias using up-to-date values of our preferred PCE weights to best approximate the weights that would be used in a superlative aggregation formula, that is, $w_s^* (p - p^*)$.⁵⁷

⁵⁶ More precisely, the subscript l represents the combination of all aggregation formulas that are used in the CPI, including geometric means for aggregating prices within many strata and modified Laspeyres across strata.

⁵⁷ Our estimates of new-outlet bias were similarly constructed by applying a bias estimate of 0.25 percent per year to the proportion of the CPI deemed likely to contain such a bias, and using these PCE weights yields similar results.

Summing our estimated bias components yields:

$$\begin{aligned} \text{Our Bias} &= w_s^* (p - p^*) + (w_l - w_l^*) p + (w_l - w_s) p \\ &= \text{Bias} + [(w_l - w_l^*) - (w_s - w_s^*)] p \end{aligned} \quad (3)$$

The difference between our bias estimate and the true bias, given by the bracketed term in equation (3), is likely to be very small on average. (The bracketed term represents different ways of constructing weighting bias; for practical reasons it would have been difficult for us to construct such estimates.)

Some readers may not agree with our assessment that the adjusted PCE data are more accurate than the CE survey expenditures. In that case, there is no difference between w and w^* and the bias estimate becomes:

$$\text{Bias} = w_s (p - p^*) + (w_l - w_s) p . \quad (4)$$

For this reason, we also presented an alternative aggregation of our detailed quality-change/new-items bias estimates based on the most up-to-date CE weights that are available (from 1998), corresponding to $w_s (p - p^*)$. This calculation generated an estimate that is 0.05 percentage point per year smaller than 0.36 percentage point figure in table 1. Thus, readers who prefer the CE survey data to the adjusted PCE data should reduce our overall estimate of CPI bias by 0.1 percentage point--0.05 percentage point for weighting bias, and another 0.06 percentage point to reflect a smaller estimate of quality-change/new-items bias.