BLS WORKING PAPERS

U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Prices and Living Conditions

Does Quality Adjustment Matter for Technologically Stable Products? An Application to the CPI for Food

John S. Greenlees, U.S. Bureau of Labor Statistics
Robert McClelland, U.S. Bureau of Labor Statistics

Working Paper 444
January 2011

Does Quality Adjustment Matter for Technologically Stable Products? An Application to the CPI for Food

by

John S. Greenlees
Robert McClelland

U. S. Bureau of Labor Statistics

December 23, 2010

Paper prepared for presentation at the Annual Meeting of the American Economic Association in Denver, Colorado, January 9, 2011. An earlier version of this paper was presented at the NBER Summer Institute Joint Session with the Conference on Research in Income and Wealth in Cambridge, Massachusetts, July 20, 2010. The authors thank Taylor Blackburn for outstanding research assistance. Valuable comments have been received from Ana Aizcorbe, Ralph Bradley, Ronald Johnson, Aviv Nevo, and Marshall Reinsdorf. All views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Bureau of Labor Statistics.

Abstract

For most indexes in the Consumer Price Index (CPI), the Bureau of Labor Statistics (BLS) uses a form of the "matched-model" approach. It has long been recognized that the matched-model method can underestimate quality improvement, and therefore overestimate price inflation, for products exhibiting rapid technological improvement. In contrast, it is frequently taken for granted that the matched model approach accurately reflects price inflation for items that have no major trend in quality. In this paper we investigate that hypothesis using CPI data for retail food items. Our results provide evidence on the accuracy of the CPI, on the validity of models of retail firm pricing behavior, and on recent work on the stickiness of prices.

The BLS handling of product replacements differs from a pure matched-model method. When an item disappears and is replaced in the CPI sample by another item, a CPI analyst in Washington decides whether the replacement product's characteristics are "comparable" or "non-comparable" to those of the old product. If the two products are judged non-comparable, they are treated as wholly different models, and their quality-adjusted price difference is imputed from the price movements of other items in the sample. If the two products are judged comparable, however, their prices are used in the index without adjustment, in the same way as if no model change had taken place. As typically discussed, a matched-model index is one that imputes price change in all cases of product replacement.

Our paper examines monthly price changes of continuing goods, comparable substitutes and non-comparable substitutes for a set of food items during a six-year period from 2003 through 2009. For substitutions, we compute both the actual price changes and estimated quality-adjusted changes. In previous work we selected 14 relatively homogeneous food categories in order to focus on price differences across stores. In this paper, we examine an expanded set of 39 categories, including many that comprise products with widely varying characteristics and more frequent turnover: for example, lunchmeats and snacks.

We find, first, that price changes for comparable substitutions are, on average, sharply greater than price changes for continuing quotes. Second, comparing our quality-adjusted and unadjusted price changes for comparable substitutes, we cannot reject the hypothesis that they are equal on average. This indicates that the CPI analysts may be roughly correct in their decisions that no significant quality changes have taken place in those cases. It also supports the idea described most recently in Nakamura and Steinsson, (as opposed to Klenow and Willis) that firms take the opportunity of product replacement to make changes in the quality-adjusted price. Finally, for substitutes judged non-comparable, we find that on average the hedonically-adjusted price changes are significantly higher than the price changes imputed using the CPI method. This suggests that the current CPI procedure for non-comparable substitutions may underestimate price change, thereby overestimating quality change. Research on this issue should continue, and conclusions about the CPI treatment of product changes will, in turn, have potentially important implications for the measurement of real output and the response of prices to monetary policy.

I. Introduction

For most indexes in the Consumer Price Index (CPI), the Bureau of Labor Statistics (BLS) estimates price inflation using a form of the "matched-model" approach. The intent of that approach is to form a price index from a sample of related goods or services, with the objective of following the price of each specific model or item for the entire life of the sample. Because the same model is followed over time, the quality-adjusted change in price is simply the observed change in price. If, as may happen, an item is no longer sold, another model is found and the price of the new item is followed. In the purest form of the matched-model approach, no explicit comparison is made between the old and new items, so that no account is made of differences in price or quality between the old and new items.

Consequently, if technological improvements lead to a persistent upward trend in quality, so that new items are superior to old items, quality improvement is underestimated, and therefore price inflation will be overestimated. This has frequently led to recommendations that the BLS employ the hedonic method of quality adjustment, and since the late 1990s the CPI has adopted hedonic regression models for several consumer durable goods indexes, such as televisions.[1] As a corollary, it is frequently taken for granted that the matched model approach accurately reflects price inflation for product areas that have no ongoing change in quality.

The BLS handling of product replacements, however, has one crucial aspect that researchers and commentators sometimes fail to take into account. When a product disappears and is replaced in the CPI sample by another similar item, an analyst in Washington first decides whether the new product's characteristics are "comparable" or "non-comparable" to those of the old product. If the two products are judged non-comparable, they are treated as wholly different models, and their quality-adjusted price difference is imputed from the price movements of other items in the sample. But if the two products are judged comparable, their prices are used in the index without adjustment, in the same way as if no model change had taken place. As typically discussed, a matched-model index is one that imputes price change in all cases of product replacement.

If the BLS method for imputing the prices of non-comparable replacements is, on average, incorrect, price inflation will be underestimated or overestimated. Price inflation also will be underestimated or overestimated if, on average, products judged comparable actually are of lower or higher quality than the items they replace in the CPI sample. This problem potentially exists in any product, regardless of the extent of technological change.

In this paper we investigate both issues by comparing the BLS variant of the matched model approach to quality adjusted indexes for retail food items. We examine monthly price changes of continuing goods, comparable substitutes and non-comparable substitutes for a set of 39 food item categories during a six-year period from 2003 through 2009. Using the CPI Research Database developed by the BLS, our work

[1] Earlier the BLS had implemented hedonic quality adjustment in CPI shelter and apparel indexes. See Johnson *et al.* (2006) for a discussion of CPI hedonic models.

observes differences in product characteristics, or what we will loosely refer to as "quality," using detailed information on the individual items priced in the CPI. For substitutions, we compute both the actual price changes and estimated quality-adjusted changes. In previous work we selected 14 relatively homogeneous food categories in order to focus on price differences across stores.[2] In this paper, we examine a much larger and broader set of categories, including many that comprise products with widely varying characteristics and more frequent turnover: for example, lunchmeats and snacks.

Section II describes the CPI pricing procedures and the CPI data that will be used in our analysis, and section III lays out a simple price change model. We present our index simulation results in section IV and our major conclusions and implications in section V.

II. CPI Procedures and Data

CPI pricing begins with the sampling of geographic areas, then of outlets (stores, service providers, websites, and other sales establishments), followed by the sampling of one or more individual products within each sample outlet. Each of the 39 item categories we study here represents either an "Entry-Level Item" (ELI), the level of item definition from which data collectors begin item sampling, or a "cluster," a major division within the ELI. An example of an ELI is Coffee, which comprises two clusters: roasted coffee and instant or freeze dried coffee. The area, outlet, and product selection processes all are based on decentralized probability sampling, in contrast to some other countries that select the CPI item sample at the national level to ensure widespread and continuing product availability. As a result, the US CPI contains an unusually great variety of sampled items within any category, and it also encounters a relatively large number of forced product replacements when sampled items disappear from outlets.

Once selected, CPI items are "repriced" on a regular basis. In most areas of the country, collection of most prices takes place only in odd or even months. The BLS, however, prices food at home, along with at energy and selected other items, on a monthly basis in all areas. Thus, the empirical price data we study in this paper are almost exclusively observed on a monthly frequency.[3]

For our empirical analysis we constructed a sample of all item prices—what the BLS calls "quotes"— for each month from September 2003 through September 2009 and in each of the 39 food categories. Note that the same individual item in a given store will be observed in multiple months until the outlet rotates out of the sample after four years or the item can no longer be found and the BLS data collector must substitute a similar item. When the latter situation occurs, CPI terminology refers to the substituted item as a new "version." The CPI analyst in Washington must then decide whether the new version's characteristics are "comparable" or "non-comparable" to those of the old version. If the two versions are judged comparable, their prices are used in the index without

[2] Greenlees and McClelland (2007, 2008).

[3] The exception is wine at home, which is not included under Food in the CPI item structure and is therefore priced on a bimonthly basis in most areas.

adjustment, in the same way as if no substitution had taken place. If the versions are non-comparable, however, they are, in effect, treated as different products, and the difference in their prices is implicitly attributed to a difference in item quality (except for an inflation factor between the two periods, which is imputed from the movements of other items in the sample).

We will refer to the sequence of observations on a version as a "version string", and the sequence of observations on comparable versions of an individual product as an "item string." Thus, an item string may comprise more than one version string, and it may extend over a period from one month to several years, depending on when or if a non-comparable substitution takes place for that product. Taken together, the version strings for a particular sampled item comprise what Klenow and Malin (2009) refer to as a "quote line"; the item strings together comprise a quote line as well. In the simplest cases, when there are no substitutions of the item, the quote line, item string, and version string will be identical.

Because it is based primarily on the month-to-month changes in the prices of identical items, the CPI approach described in the previous paragraph yields a form of "matched model" index, although the latter term is not always precisely defined. The international CPI manual defines "matched models" merely as the practice of pricing exactly the same product in two or more consecutive periods,[4] and it stresses that when items are replaced with new ones of different quality a quality-adjusted price is required. Other studies of the efficacy of matched-model indexes, however, sometimes use the term to refer to the stylized case in which the index is based only on price changes for continuing items.[5] The CPI approach clearly differs from this pure matched-model index because of the use of comparable substitutions, the same practice that forces us to distinguish in this paper between version strings and item strings.[6] If all item replacements were judged non-comparable, an item string would be equivalent to a version string, and *only* prices of identical items would be compared for use in the CPI. That would follow the pure matched-model concept. At the other extreme, if CPI analysts judged that all item replacements are comparable, each item string would last until the sample is rotated. In that case the CPI would be computed with no recognition of any changes in item quality from period to period.[7]

Because these two extremes should produce substantially different estimates of price inflation, CPI staff have developed written criteria for each ELI to help guide the analysts' comparability decisions. As an example, the appendix shows the characteristics that are recorded by the CPI for the breakfast sausage cluster and that are used in comparability decisions. For breakfast sausage, the criteria indicate that a substitution

---

[4] ILO (2004, p. 47).

[5] See, for example, Silver and Heravi (2005, p. 269), Aizcorbe *et al.* (2000, p.4), Aizcorbe *et al.* (2003, p.6), and Aizcorbe and Pho (2005, p.5).

[6] This is recognized by Aizcorbe *et al.* (2003, footnote 6), who distinguish between "matched model indexes" and the methods used by statistical agencies. Triplett (2006, p. 15) notes the two meanings of "matched model" and suggests that "matched models only" might be a better term for the pure matched model approach.

[7] Exceptions to the substitution-handling process described here, such as the use of hedonic regressions for quality adjustment, occur elsewhere in the CPI but are very rare in the food categories.

from links to patties, for example, should always be judged non-comparable. The same would be true for a substitution from a national brand of sausage to a store brand. On the

A small research literature exists on the difference between matched model and hedonic indexes; among the most widely-cited studies are other hand, a substitution from mild to hot sausage would normally be considered comparable. those by Silver and Heravi (2005) and Aizcorbe *et al.* (2000, 2003). Triplett (2006) devotes chapter IV to reviewing some of that literature and examining the criteria necessary for a matched model index to closely approximate an index using hedonic quality adjustments. Triplett assumes that all replacements are treated as non-comparable, so that the difference in price between an old item and its replacement is implicitly attributed to a difference in item quality. Markets with both rapid turnover of models and improvement in quality for items would seem to cause the matched model index to rise faster (or fall more slowly) than a hedonic index. However, this depends on the structure of the market: if prices of existing items fall in response to the introduction of higher quality items, the matched model index may be quite close to the hedonic index. Aizcorbe and Pho (2005) further point out that part of the difference between hedonic and matched model indexes may be caused by differences in how weights are used in the two indexes.[8]

While Triplett examines the case of items undergoing rapid technological improvement, such as personal computers, Silver and Heravi (2005) also examine durable goods, such as washing machines, that are not experiencing the same degree of improvement. While the markets they study are subject to rapid market turnover, the difference between matched model and hedonic indexes will depend on a number of possible strategies by firms. For example, retailers could sharply drop the prices of items at the end of their life, or they may price discriminate by charging a premium at the beginning of a model's life. Further, manufacturers may introduce items at a discount in order in penetrate a new market.

III. A Model of Substitution Handling Methods in the CPI

In this section we present a simple model for price change within a given item-area stratum, or "cell," of the CPI. We categorize price changes in two ways: first, by whether or not there is a product substitution (a change in product characteristics or "quality") and second, by whether the CPI commodity analyst judges the substitute item to be "comparable" or "non-comparable" to the older version. For each of these situations we compare the treatment in the food categories of the CPI to the quality adjustments yielded by a hedonic model.

Our stylized discussion of price change is designed not to model firm decision-making but rather to express, in statistical terms, the contrasting treatment of substitutions in the CPI and hedonic methods. Both can be thought of as estimating the "true" or "pure" quality-adjusted price change for each item in the sample, then aggregating up those estimated values to generate an index for the full sample. Therefore, we focus on the

[8] In particular, they consider hedonic indexes calculated by running unweighted regressions of variables for item characteristics and dummy variables for time.

measurement of price change for a specific item *i* at a specific point in time *t* under the two approaches.

All variables are measured as log-changes. We suppress the time subscript for brevity in notation. The variable π_i represents the observed log-change in price for the specific item, whereas τ_i represents the underlying variable of interest: the price change for that item net of the value of any change in quality. The quality change is indicated by q_i, and most other symbols we introduce will represent random variation terms. For convenience, we ignore all issues of sample weighting.

For convenience of exposition, we proceed by working within the unweighted stochastic approach to the price index problem. As discussed in, for example, the international CPI manual,[9] the stochastic approach is usually contrasted with the more familiar economic approach embedded in cost-of-living index theory and adopted by the BLS as a framework for the CPI. By relying on the stochastic approach we avoid dealing with such factors as utility functions, substitution elasticities, and spending weights, while losing nothing of importance for present purposes. In algebraic terms, our assumption is that the price of a specific sampled and tracked product changes according to:

$$\pi_i = q_i + \tau_i \qquad (1)$$

We now turn to a description of how the CPI and hedonic methods handle continuing and substitute items. In formulas for estimators, we will use superscripts or subscripts to represent the pricing situation and method being used. Substitutions are denoted by the superscript S, or by SC or SN when it is necessary to indicate whether they are comparable or not; unchanged, continuing observations have superscript U. The H superscript indicates a hedonic, as opposed to CPI-method, estimate.

<u>Price measurement for continuing items.</u> In this case, for an item we have by assumption no quality change, and consequently:

$$\pi_i = \tau_i \qquad (2)$$

The CPI will measure the price change by direct observation of the unadjusted price change, and thus its estimate of "pure" price change will be correct for continuing items:

$$\hat{\tau}_i^U = \pi_i = \tau_i \qquad (3)$$

Now, assume that the hedonic model being employed is specified appropriately and that it yields an unbiased estimate of quality in each period and therefore an unbiased estimate of quality log-change q for each item. When a product version changes, the hedonic estimate of quality change is given by:

$$q_i = \hat{q}_i^H + \hat{\varphi}_i \qquad (4)$$

In equation (4), $\hat{\varphi}_i$ is a mean-zero term representing the difference between the true quality change q_i and the estimated change $\hat{q}_i^H$. When there is no version change, however, the hedonic estimate of quality change will necessarily be zero under our assumption that no product characteristic regressor variables have changed. As a result,

[9] ILO (2004), paragraphs 16.75-16.78.

as in the CPI case, the hedonic model also yields the correct measure of pure price change τ_i:

$$\hat{\tau}_i^{UH} = \pi_i = \tau_i \tag{5}$$

Price measurement for substitute items. As mentioned earlier, the CPI has two methods for handling price changes of substitute food items, depending on whether the versions are judged comparable or non-comparable. (We note again that for some items outside the food area, the CPI employs direct quality adjustments, including hedonic adjustments.) For comparable substitutions, the CPI directly compares the product price in the current and previous period. The CPI index change then over- or underestimates the target τ_i by the size and direction of the true quality change q_i.

$$\hat{\tau}_i^{SC} = \pi_i = q_i + \tau_i \tag{6}$$

We emphasize that the classification of a version change as a comparable substitution is based on a judgment by the trained CPI commodity analyst that any quality change is minimal. If q_i is always zero for comparable substitutions, there is no error, and if the conditional expectation of q_i is zero, there is no bias. Otherwise, both the bias and variance of the CPI estimate will depend on the distribution of true quality changes, conditional on the comparability decision.

In the case of non-comparable substitutions, the CPI does not use the observed price difference between the old and new version to generate its estimate of pure price change. Instead, the CPI estimate is the cell-average of the log-price changes measured for continuing items and comparable substitutes in that period.[10] Let w_U and w_{SC} indicate the proportions of those groups, respectively, in the population.[11] Further, let $\overline{\pi^U}$ and $\overline{\pi^{SC}}$ represent the means of observed price changes in the corresponding subsamples. Then, using (3) and (5), we obtain:

$$\hat{\tau}_i^{SN} = \frac{w_u}{w_u + w_{SC}} \overline{\pi^U} + \frac{w_{SC}}{w_u + w_{SC}} \overline{\pi^{SC}} \tag{7}$$

In a substitution case, whether comparable or non-comparable, the hedonic index estimates pure price change by subtracting the estimated change in quality:

$$\hat{\tau}_i^{SH} = \pi_i - \hat{q}_i^H = \tau_i + (q_i - \hat{q}_i^H) = \tau_i + \hat{\varphi}_i \tag{8}$$

The second equality in (8) is obtained by substituting from equation (1) and the third equality comes from (4). Under our assumptions, the hedonic estimate is unbiased overall. However, in principle the expectation and/or variance of $\hat{\varphi}_i$ are conditional on the information that the analyst has judged the versions comparable.

---

[10] As a technical matter, the substitute item in question is excluded from the calculation of the cell estimate of price change, and that cell estimate is then employed to impute the pure price change for the non-comparable substitute. The cell estimate is the mean log-price change because basic CPI food indexes are computed using a geometric mean formula.

[11] As detailed in section IV below, in reality the population of items over which the weighted-average prices in (8) are calculated is actually a subset of the item-area stratum for which the basic index is defined. This distinction is not germane to the fundamental objectives of this paper.

Basic index measurement. We now introduce an asterisk superscript to indicate τ^*, the underlying population mean price change, which is the target aggregate statistic to be measured for the item-area stratum. We now assume that

$$\tau_i = \tau^* + \varepsilon_i \tag{9}$$

The ε_i term is unique to the item and reflects the difference between its pure price change and the underlying population mean. There may be some intertemporal correlation in the ε_i if, for example, individual item price levels exhibit a reversion to trend. Moreover, we permit the possibility that, although the population mean of ε_i is zero, its expected value conditional on a substitution may be nonzero. That is, we want to allow for the possibility that the expected value of τ_i, the change in the item's "pure" price, is larger or smaller for changing products than for continuing identical products. This might occur if, for example, there is an operational or public-relations cost associated with increasing prices, making it more advantageous to increase the pure price at the same time that a product modification is being introduced.[12] Alternatively, in some product markets it is commonly observed that new, improved products tend to appear at lower quality-adjusted prices, suggesting that ε would be negative and q positive at the time of substitution.

Assuming that the sample is drawn and weighted appropriately, both the CPI and hedonic approaches will yield unbiased estimates of τ^* if they are based on unbiased estimates of τ_i for each item.

Using the notation $\overline{\hat{\tau}_i^U}$ to represent the mean of $\hat{\tau}_i^U$ averaged over the sample of unchanged items, and similarly for the other two item groups, we can write the CPI index change estimate I as:

$$I = w_U \overline{\hat{\tau}_i^U} + w_{SC} \overline{\hat{\tau}_i^{SC}} + (1 - w_U - w_{SC}) \overline{\hat{\tau}_i^{SN}} \tag{10}$$

Based on equations (3), (6), and (7) in the previous subsection, however, this naturally collapses to the same value used to impute the price changes for non-comparable substitutions:

$$I = \frac{w_u}{w_u + w_{SC}} \overline{\pi^U} + \frac{w_{SC}}{w_u + w_{SC}} \overline{\pi^{SC}} \tag{11}$$

We can now combine (11) with (9) to write the expected value of I as:

$$\begin{aligned} E(I) &= \frac{w_u}{w_u + w_{SC}} [\tau^* + E(\varepsilon_i | U)] + \frac{w_{SC}}{w_u + w_{SC}} [\tau^* + E(\varepsilon_i | SC) + E(q_i | SC)] \\ &= \tau^* + \frac{w_u}{w_u + w_{SC}} [E(\varepsilon_i | U)] + \frac{w_{SC}}{w_u + w_{SC}} [E(\varepsilon_i | SC) + E(q_i | SC)] \end{aligned} \tag{12}$$

Then, since the expected value of ε_i over the entire sample will be zero,[13] simple algebra yields:

$$E(I) = \tau^* - \frac{w_{SN}}{w_u + w_{SC}} E(\varepsilon_i | SN) + \frac{w_{SC}}{w_u + w_{SC}} [E(q_i | SC)] \tag{13}$$

---

[12] This hypothesis was raised by Armknecht *et al.* (1997) and most recently by Nakamura and Steinsson (2008).

[13] Here for convenience we ignore the complication that, in finite samples, the proportions of continuing items and of comparable and non-comparable substitute products may not equal their population proportions.

Equation (13) shows that E(*I*) can fail to equal τ^* due to two important sources of potential bias: first, the expected value of *q* for substitutions judged comparable may be nonzero (i.e., the decision to call them comparable was incorrect); and second, the expected value of ε_i for non-comparable substitutions may differ from zero (i.e., the implicit valuation of quality difference was incorrect).

At first it may seem surprising that the true quality change for non-comparable substitutes plays no role in (13) and the equations that precede it. This occurs because, conditional on the other variables, the true change in the non-comparable substitute's quality q_i does not affect the amount by which the item's true price change τ_i is mis-estimated. One can, however, think of the CPI process as making an implicit estimate of the quality change for a non-comparable substitution, equal to the observed price change minus the CPI estimate of τ:

$$\hat{q}_i^{SN} \equiv \pi_i - \hat{\tau}_i^{SN} \tag{14}$$

From equation (1), however, we know that the true quality change for the item is

$$q_i = \pi_i - \tau_i \tag{15}$$

Thus, the error in the implicit estimate of q_i for non-comparable substitutes is just the negative of the error in estimating τ_i.

The hedonic index estimate is given by:

$$I^H = w_U \overline{\hat{\tau}_i^{UH}} + (w_{SC} + w_{SN})\overline{\hat{\tau}_i^{SH}} \tag{16}$$

Using equations (5), (8), and (9) this has an expected value of

$$\begin{aligned} E(I^H) &= w_U[\tau^* + E(\varepsilon_i|U)] + w_{SC}[\tau^* + E(\varepsilon_i|SC) + E(\hat{\varphi}_i|SC)] \\ &\qquad + w_{SN}[\tau^* + E(\varepsilon_i|SN) + E(\hat{\varphi}_i|SN)] \\ &= \tau^* + w_{SC}[\tau^* + E(\hat{\varphi}_i|SC)] + w_{SN}[\tau^* + E(\hat{\varphi}_i|SN)] \\ &= \tau^* \end{aligned} \tag{17}$$

The second equality in (17) uses the fact that the expectation of the sample mean of ε_i is zero, and the third relies on our assumption that the hedonic regression is unbiased in all cases. Under these assumptions, the hedonic index estimate is unbiased.

IV. Simulation Results

In this section, we examine the behavior of the CPI-method index *I*, defined in equations (10) and (11), and the hedonic index I^H, defined in (16). The two indexes are simulated and compared for each of our 39 sample item categories. In building our simulation model, we make several simplifying assumptions. First, we deal only with price changes at the time of product substitution; that is, within CPI quote lines. In taking this approach we are assuming that the hedonic model is used in the same way as other models employed elsewhere in the CPI. When a product substitution occurs, the price difference is adjusted by the inner product of the estimated hedonic coefficient vector and the vector of changes in item characteristic regressors. That is, the change in the hedonic index is not constructed from a time dummy coefficient or by evaluating regression predictions estimated for two points in time.[14]

A second simplification in our model is that the CPI sampling weights attached to individual items are ignored, at little or no loss in relevance. Third, we assume that the characteristics used in the hedonic regression only change at the time of version changes (i.e., substitutions). There may be occasions where small product size changes, for example, can affect the hedonic estimate of quality without generating a version change. These possibilities are ignored here for convenience in order to limit the number of situations to be analyzed. Fourth, we ignore the possibility that direct quality adjustments may be used by the CPI in some food cases. Finally, it is important to recognize that the CPI is not routinely revised when new analytical information becomes available. As a consequence, the BLS can only apply a hedonic model that has been estimated using data from a prior period.[15] Our purpose here is not to develop a new real-world procedure but to use the results of hedonic regressions to cast light on the processes of quality and price change.

To implement our approach, we began by estimating a semi-logarithmic hedonic regression for each of our item categories, pooling data over our entire six-year sample period. The dependent variable in each case is (the logarithm of) price per ounce, the measure used in the CPI. As described in Greenlees and McClelland (2008), the explanatory variable vectors include a set of product characteristics specific to the item, dummy variables for each month, and dummy variables for each outlet in the sample. The outlet dummies are included to reflect the CPI maintained assumption that differences in prices across outlets, holding item characteristics constant, reflect differences in locational convenience, quality of service, and other unmeasured outlet-specific characteristics. In the case of breakfast sausage, for example, the product-specific regression variables are dummies corresponding to the characteristics identified in the checklist shown in the Appendix: whether or not the sausage is country or breakfast style, its type of seasoning, and so on. Wherever applicable, the logarithm of package size in ounces, and the square of that variable, are also included in the regressions. Our previous research has shown a strong negative relationship between package size and the price per ounce.

Before reviewing the regression results, it is important to make two observations. First, because they are based on the entire study period, the hedonic coefficients will reflect information and relationships not available to the commodity analysts when they made their substitution decisions. As noted above, the value of these regressions here is for retrospective evaluation, not as a potential tool for analyst decision-making. Second, we do not assume that our hedonic indexes are perfect or that they necessarily are more accurate than the CPI-method indexes in all cases. That said, the hedonic indexes employ

[14] National Research Council (2002, chapter 4) refers to the CPI hedonic procedure as the indirect hedonic approach, as opposed to the direct time dummy method and the direct characteristics method involving estimation of a separate regression for each time period. Greenlees and McClelland (2008) employed the direct time dummy method to generate both hedonic indexes and indexes reflecting matched-model assumptions.

[15] Because the CPI is used widely in contracts, international practice is to revise published CPI data only when major errors in data collection or processing are discovered.

all the checklist variables for our sample item categories, along with package size and some additional variables we added from text fields. Moreover, they use our entire sample to estimate marginal valuations of changes in these checklist characteristics, rather than relying on a binary comparable-non-comparable classification. We therefore believe that support is given to the CPI method when the two approaches yield similar indexes, and that, conversely, a divergence is evidence that the CPI method should be further evaluated.

Although the results of these hedonic regressions are critical for our index comparisons, space permits only an limited discussion. Again using the example of breakfast sausage, that regression employs a relatively small sample size of 13,599 price observations in the 73 months and 458 outlets. Nevertheless, most coefficients on the variables of interest have large, significant t-statistics. For example, the dummy variable for "Country Style" sausage has a t-statistic of 7.02. That style of sausage is estimated to carry approximately an eight-percent higher price than "Breakfast Style," and a 13-percent higher price than the residual Other category. Linked stuffed sausage is priced roughly ten percent higher than patties, ceteris paribus, and about 11 percent higher than formed, unstuffed links. Both package size variables are highly significant, as are the dummy variable vectors for seasoning and for several specific brands of sausage. The vectors of outlet and time dummies are also highly significant. Finally, the time-dummy coefficients indicate an approximately 10.1 percent growth in quality-adjusted price over the six-year period of study.

The breakfast sausage regression has an R^2 statistic of 0.83, which is representative of our regression results. The 39 R^2s ranged from 0.46 to 0.94, but the median was 0.86. Only one item (round steak) yielded an R^2 under 0.64, and only two others were below 0.74.

Using these regression results we simulated a hedonically-adjusted breakfast sausage index, employing the estimated parameters to make quality adjustments whenever there was an item substitution (version change) in the CPI data. We made these adjustments regardless of whether the change in item characteristics between the old and substitute item was judged comparable or non-comparable by the CPI analyst. The resulting index yields a log-increase of 0.105 over six years, similar but not identical to the 0.096 obtained directly from the regression coefficients. The two estimates will necessarily differ because the simulated index makes quality adjustments only in cases of substitutions, whereas the regression parameters can reflect changes over time in the mix of product characteristics occurring through the sample rotation process.

Our simulated CPI-method index for breakfast sausage used the actual analyst comparability decisions to impute pure price change according to the model of the previous section. That log-change in that simulated index was 0.111, very slightly higher than the hedonic index. In this case, therefore, the hedonic method estimates slightly more "quality improvement" from changes in item characteristics than does the CPI method.

Table 1 shows the results of simulating indexes in that way over the period from September 2003 through September 2009 for each of our 39 item categories. The table demonstrates the wide range of categories included in our sample; nevertheless, certain important food items are not represented. The most notable of these are fruits and vegetables, which we excluded because they tend to have few substitutions and/or few measurable characteristics. Table 1 also shows that the total sample sizes – the numbers of month-to-month CPI price comparisons over the six-year period – vary from only 27 per month on average for Frozen whole chicken and chicken roasts to 634 per month for Fresh or frozen chicken parts, another cluster in the same ELI. These national sample sizes are correlated with, but not perfectly proportional to, the expenditure weights of the corresponding item categories in the CPI.

We believe that the number and variety of our sample item categories provide evidence on the adequacy of quality adjustment methods for food. The fact that the sample of items is non-random implies, however, that one should not view our numerical results as representative of the entire CPI or even the CPI for food.

In addition, the computational details of our simulated indexes differ in several ways from those in the actual CPI. Perhaps most importantly, our indexes are based on equally-weighted averages of logarithms of prices, whereas the CPI computes geometric mean indexes for each of 38 geographic areas using individual sampling weights for observations, then aggregates the area indexes using an arithmetic mean formula and expenditure weights taken from the Consumer Expenditure Survey. Our use of the geometric mean at all levels can be expected to produce slightly lower index results, on average.[16] Our imputation methods differ as well. For non-comparable substitutions within each of our 39 categories, we define our "CPI" estimate of price change as the mean (logarithmic) price change in that period and item category. The official CPI, on the other hand, uses the mean change in that period within the area, ELI, and half-sample (replicate subsample used for variance estimation).[17]

White bread, a category for which the BLS publishes both indexes and average-prices, offers another reasonably illustrative example of our simulation results. The second row of Table 1 shows that our CPI-method simulation yields a log-increase of 0.283 over six years, similar although not identical to the 0.287 increase in the published CPI for white bread. The CPI also publishes a national-average per-pound price series for white, pan bread. That series exhibited a log-change of 0.299 over the same six years. CPI average price series are computed and published separately from the index itself, and employ an arithmetic mean formula to aggregate observations. We conclude that, given the limitations discussed in the previous paragraph, our simulations are satisfactorily representative of the CPI methodology.

Table 1 also shows a six-year log-change of 0.295 in our simulated hedonic index for white bread. This is 0.012 greater than our simulated CPI, a difference of about 0.2 percent per year.[18] The greater increase for the hedonic index indicates that it is estimating a lower rate of quality improvement – actually greater quality decline in the case of white bread – than the CPI method implicitly estimates.

---

[16] Geometric mean indexes will generally yield lower index changes, based both on mathematical reasoning and on empirical evidence. See, for example, Johnson *et al.* (2006), Table 2.

[17] BLS (2007), pp. 22-23.

[18] As with breakfast sausage, this simulated hedonic index will be different from the index read directly from the time-dummy regression coefficients, because the former makes quality adjustments only in the cases of item substitutions. The time-dummies in the white bread regression imply a log-change of 0.268 over our study period.

By decomposing the estimated price changes at the time of product substitution we can compute the estimated impacts of individual regressor variables on the hedonic index. In the case of breakfast sausage, a growth in the share of one national brand accounts for almost all of the net quality adjustments. Other item characteristics either had small regression coefficients or exhibited little change during the study period. For white bread the crucial item characteristics in determining quality-adjusted price per ounce were brand and package size. Changes in package size along with substitutions from store brands to other brands accounted for almost all of the hedonically-estimated net quality decline of about 2.1 percent over six years. Had all substitutions been judged comparable, in other words, the log-change in the white bread index would have been about 0.27, less than the increase in the CPI-method index.

The index results in Table 1 are summarized in Figure 1 and Table 2. Of the 39 categories studied, 27 yield a higher index change using the hedonic method than using the CPI method, as we found for white bread. The figure displays the range of index differences, which range from a negative 0.10 for wine at home to a positive 0.10 for miscellaneous prepared foods. In a two-tailed test using the binomial distribution, at the usual significance levels we can reject the hypothesis that this greater number of items with higher hedonic indexes occurred purely by chance. As shown in Table 2, there is only about a two-percent chance that one of two equally-likely outcomes will occur 27 or more times in 39 trials. The table also demonstrates that we can reject the hypothesis that the expected difference between the indexes is zero. The unweighted mean difference across the 39 item categories is -0.0108 with a standard error of 0.0061, and a two-tailed t-test rejects at about the 8 percent level the hypothesis that these differences are drawn from a common distribution with mean zero. Weighting the differences by their shares of the total sample size generates qualitatively similar results: the weighted mean difference is -0.0122 and the null hypothesis is rejected at about the 0.7 percent level. The similarity of the weighted and unweighted results here avoids the risk of drawing conclusions on the basis of a small number of either high-variance or highly-weighted item categories.

We next explore the important question of which types of price comparisons are most influential in determining overall index change. Table 3 shows the contributions to the CPI-method and hedonic-method indexes of price change between continuing items, comparable substitutes, and non-comparable substitutes. The first line of that table repeats Table 2 in showing that the overall unweighted averages of the 39 index log-changes using the two approaches were 0.1478 and 0.1586, respectively, a difference of -0.0108. Price changes for continuing items are much more frequent than substitutions, accounting for almost 97 percent of the approximately 756,000 price change observations overall, and their average contribution to overall price change is 0.1386 of the total 0.1478 in the CPI simulations. Only 36 percent of substitutions are judged comparable. Interestingly, however, these observations, in which the prices of the new and old item are directly compared, account for almost twice as much total price change in the CPI method as do non-comparable substitutions, in which the price changes are imputed from the rest of the sample.

For the hedonic indexes, the Table 3 results are quite different. The 0.1386 contribution from continuing items is necessarily identical to that in the CPI method, since those observations are treated identically in the two indexes. The average contribution from comparable substitutions is also roughly similar. However, non-comparable substitutions account for a total of 0.0143 in the overall average hedonic index log-change, almost five times as much as in the CPI-method indexes. As shown in the last column of the table, of the total -0.0108 difference in the two index averages, the different handling of non-comparable substitutions explains -0.0112, or more than 100 percent.

The implication of Table 3 is that the hedonic method's estimates of quality differences between the old and non-comparable replacements in turn yield estimates of pure price change that are very different from, and higher on average than, the CPI method's approach of estimating pure price change from continuing items and comparable substitutions. For comparable substitutes, the differences between the two approaches do not appear so striking. We look more closely at these phenomena in the next three tables.

Table 4 focuses on comparable substitutions. Across the 39 item categories we have a total of 8,085 comparable substitutions, each with a monthly price-change estimate computed using both the CPI and hedonic methods. In the hedonic method, the estimated log-price change is adjusted for the values of any differences in characteristics or "quality" between the new and replacement items. The CPI method implicitly judges those quality differences to be zero in all cases. The numbers in the first rows of the table show that average hedonic monthly log-change is greater than the CPI-method change in 22 of 38 categories.[19] For each item category we can construct a paired t-test on the differences between the two sets of estimated price changes, with the number of degrees of freedom determined by the number of comparable substitutions in the category. The 38 t-tests indicate that the mean differences are negative (i.e., the hedonic estimates higher) and significant in only one case and positive and significant in one, using a two-tailed test and a five percent significance level criterion. (That is, approximately five percent of the items were found significant at the five percent level, just as would be expected if the differences were random.)

These category-by-category results provide no evidence that the hedonic and CPI methods for handling comparable substitutions have a consistent direction of difference across categories. To examine this we first perform a simple t-test on the 8,085 differences against the null hypothesis that the overall mean difference is zero. The overall average monthly log-change is 0.00977 in the CPI method, which simply compares the prices of the new and replacement item. The overall average price change using the hedonic method is 0.00948, indicating that on average the value of the differences in item characteristics is a mere 0.00029.[20] The standard error of the mean difference is 0.00071, producing a t-statistic of 0.40, which is insignificant.

The above t-test is based on the assumption that the 8,085 differences are independent draws with constant mean and also constant variance regardless of category. Moreover,

[19] One category, canned ham, had only 15 comparable substitutions, and the hedonic and CPI estimates were identical in all those cases.

[20] The corresponding unweighted means of the 39 category-level differences are 0.0130 and 0.0119 for the CPI and hedonic methods, respectively. The component results in Tables 4 and 5 do not exactly aggregate to the Table 2 and Table 3 results because the numbers of prices collected varies from month to month. Some item comparisons therefore carry more weight than others in the overall index change for a given item.

it gives higher weight to the categories with more comparable substitutions. (The sample sizes by category range from 15 to 799.) As another piece of evidence, therefore, we normalize each of the 39 category mean differences by their standard errors. We then perform a simple one-sample t-test of whether these normalized means are drawn from a normal population with mean zero and unknown variance. Table 4 shows that we again cannot reject the null hypothesis.

The final rows of the table report on a chi-square test of whether *all* the category mean differences have expectation zero, obtained by treating the 39 normalized differences as asymptotically normal variates. This chi-square statistic, with 39 degrees of freedom, is also not nearly significant at the usual confidence levels.

Overall, Table 4 provides us with a strong and important result. If the t-statistics in the table were significant, we would conclude that there was a systematic tendency for comparable substitute items to have higher or lower quality than the items they replace. This would cast doubt on the CPI analysts' decisions to judge those items comparable. We find no evidence of such a problem.

Table 5 applies the same set of tests to the 14,351 non-comparable substitutions in our data. As expected given the earlier discussion of Table 3, the results here are more definitive than for the comparable substitutions in Table 4. The average difference in estimated monthly log-price change across all non-comparable substitutions between the CPI and hedonic methods is -0.00806, more than 28 times as large in absolute value as the corresponding Table 4 value. The hedonic index rises by more than the CPI-method index in 26 of 39 cases, which is a significant difference at about the 5 percent level using the binomial distribution. In contrast to Table 4, our chi-square test in Table 5 rejects the hypothesis that all item categories have a zero mean difference at about the eight percent level. More importantly, our two t-tests reject the hypothesis that the mean difference averaged across all item categories is zero at the 0.3 percent and 5.3 percent levels.

The Table 5 results thus cast some doubt on the CPI procedure of imputing price change for non-comparable substitutions from the observed price changes of continuing items and comparable substitutions, even in food categories. The hedonic estimates of quality change tend to be noticeably smaller than the implied quality changes yielded by the CPI technique.

One important question in this regard is whether price changes for non-comparable substitutes are more similar to those for comparable substitutes than to the price changes of continuing items. This possibility has been raised before, both within and outside the BLS, in the context of other CPI components, and provides the justification for the "class mean" technique used in many non-food areas.[21] To explore this idea, we first ask whether the price changes for comparable substitutes are systematically different from those of continuing items. For each item category we regressed the measured monthly price changes on month dummies and a dummy variable for whether the observation was a comparable substitution; non-comparable substitutions were excluded. (Item characteristic variables were excluded because under CPI assumptions any quality change in these cases is negligible. The time dummies were necessary to avoid spurious results

[21] See the discussions in Armknecht *et al.* (1997) and Lane (2001).

from the different timing of sample observations of the two types being compared.) The results are shown in Table 6. The coefficient on the comparable-substitute dummy variable was positive in 28 of 39 regressions, significantly positive in five, and significantly negative in only one. Again using a t-test and a chi-square test on the normalized coefficients (i.e., on the coefficients divided by their standard errors), we soundly reject the hypotheses that comparable-substitute and continuing items increase in price at the same rate as each other, either overall or within every item category. The procedure of combining those observations to impute price change to non-comparable substitutions is again seen to be potentially flawed.

Another suggestive comparison simply uses the mean price changes in our sample. Table 7 displays the average estimated log-price changes using different methodologies and situations. Unlike the means in the preceding tables, the means in Table 7 are unweighted averages of the item category means; this avoids having the comparisons distorted by the fact that comparable, non-comparable, and continuing items are distributed differently across item categories. The unweighted average across categories of the mean log price change for continuing items is 0.00205 per month. This corresponds to an annualized growth rate of approximately 2.5 percent per year. Using the CPI method, the average increase for non-comparable substitutes is a closely similar 0.00212, reflecting the fact that those CPI estimates are imputed primarily from the price changes of continuing items. In contrast, the hedonic estimate for non-comparable substitutes is a much higher 0.00713, or almost nine percent per year . Meanwhile, both the CPI and hedonic estimates in Table 7 for *comparable* substitutes are over 15 percent per year. Thus, the hedonic model implies that substitute items have large quality-adjusted price changes, much higher than for continuing items, regardless of whether those substitutes are comparable (i.e., small differences in characteristics) or non-comparable (large apparent differences). We also can see that the CPI estimate of price change for comparable substitutes is qualitatively similar to the hedonic estimates for substitutes but far from the average price changes for continuing items. All this argues for the potential value of a different imputation method for non-comparable substitutes in the CPI food area.

V. Concluding Remarks

The results in this paper have potential value to the BLS in its continuing efforts to improve quality adjustment procedures in the CPI. Our work in the retail market setting also contributes to the growing body of research on price setting and product change.

We find, first, that price changes for comparable substitutions are, on average, sharply greater than price changes for continuing quotes. This supports the idea described most recently in Nakamura and Steinsson (2008), but challenged in Klenow and Willis (2007), that firms take the opportunity of product replacement to make changes in the quality-adjusted price. Although this argument has frequently been made with respect to consumer durables and other technologically advancing product areas, it has seldom been examined in the market for consumer foods. The fact that changing and unchanged items display different price movements also may raise concerns about the CPI procedure of combining the two for the purpose of imputing price movements to non-comparable substitutions.

Second, comparing our quality-adjusted and unadjusted price changes for comparable substitutes, we cannot reject the hypothesis that they are equal on average. This suggests that the CPI analysts are roughly correct in determining that no significant quality changes have taken place in those cases.

Finally, for substitutes judged non-comparable, we find that on average the hedonically-adjusted price changes are significantly higher than the price changes imputed using the CPI method. This again suggests that the current CPI procedure for non-comparable substitutions may underestimate price change. Notice that this is the opposite conclusion from that reached by Bils (2008), who uses information on subsequent product penetration to argue that entering products have lower quality-adjusted prices.[22] Research on this issue should continue, perhaps by exploring a variant of the BLS "class mean" approach in which price changes are imputed from comparable substitutes rather than from all items. Alternatively, further research may indicate that the guidelines for CPI analysts should be adjusted, so that more items are judged comparable and more observed price changes used directly in the index. Any conclusions reached about the CPI treatment of product changes will, in turn, have potentially important implications for the measurement of real output and the response of prices to monetary policy.

[22] Examining historical data on apparel prices, Gordon (2009) also observes that CPI indexes demonstrate lower inflation than hedonic indexes.

Appendix

ELI FD011 - BACON, SAUSAGE AND RELATED PRODUCTS
Cluster - 02 - BREAKFAST SAUSAGE AND RELATED PRODUCTS

- TYPE*
  - A1 Country style (may contain 10% to 20% beef)
  - A2 Breakfast style
  - A99 Other

- FORM *
  - B1 Loose (unstuffed)
  - B2 Unlinked stuffed
  - B3 Linked stuffed
  - B4 Patties
  - B5 Formed, unstuffed links
  - B99 Other,

- SEASONING
  - C1 Mild
  - C2 Hot
  - C3 Highly seasoned
  - C99 Other

- PACKAGING*
  - D1 Prepackaged
  - D2 Not prepackaged

- BRAND*
  - E98 Store brand,
  - E99 Other

- PRICING UNIT
  - F1 Per pound (454 g)
  - F99 Other

- OTHER FEATURES
  - G99
  - H99
  - I99

*Substitutions between subcategories are always non-comparable

Table 1. Average Total Index Log Changes, 2003-2009

Item	CPI Method	Hedonic Method	Difference	Observations
Cereal and bakery products				
Breakfast cereal	0.114	0.126	-0.012	42,213
White bread	0.283	0.295	-0.012	24,275
Bread other than white	0.312	0.332	-0.020	16,222
Cookies	0.093	0.084	0.008	11,881
Meats				
Uncooked ground beef	0.234	0.228	0.007	43,569
Round steak	0.133	0.116	0.016	17,307
Sirloin steak	0.096	0.119	-0.023	15,020
Other steak	0.149	0.121	0.028	21,817
Bacon	0.024	0.034	-0.009	20,243
Breakfast sausage	0.111	0.105	0.006	12,974
Ham other than canned	0.055	0.109	-0.054	25,831
Canned ham	0.130	0.104	0.025	4,175
Lunchmeat (excl. bologna, liverwurst, or salami)	0.093	0.159	-0.066	19,568
Bologna, liverwurst, or salami	0.162	0.172	-0.010	11,709
Poultry				
Fresh whole chicken	0.206	0.216	-0.009	25,347
Fresh or frozen chicken parts	0.169	0.180	-0.011	46,280
Frozen whole chicken and chicken roasts	0.153	0.215	-0.062	1,990
Turkey	0.143	0.202	-0.059	20,793
Poultry other than chicken and turkey	0.139	0.210	-0.072	7,091
Fish and seafood				
Fresh fish	0.294	0.262	0.031	20,028
Fresh seafood	0.121	0.094	0.028	12,465
Packaged shelf stable fish or seafood	0.319	0.330	-0.011	16,885
Processed fish (excl. packaged shelf stable)	0.167	0.202	-0.036	9,664
Processed seafood (excl. packaged shelf stable)	0.022	-0.045	0.066	8,251
Dairy and related products				
Cheese and cheese products	0.125	0.133	-0.009	44,136
Ice cream and related products	0.066	0.087	-0.021	34,447
Nonalcoholic beverages and beverage materials				
Cola drinks	0.216	0.228	-0.012	34,644
Other carbonated drinks	0.188	0.192	-0.004	24,632
Nonfrozen noncarbonated drinks	0.149	0.151	-0.002	42,834
Roasted coffee	0.263	0.286	-0.023	17,757
Instant or freeze dried coffee	0.163	0.123	0.040	5,939
Frozen and freeze dried prepared foods				
Multiple course prepared foods	0.022	0.050	-0.028	8,594
Meat, fish, or poultry pies	0.090	0.120	-0.030	6,032
Prepared chicken	0.121	0.167	-0.046	6,192
Pizza, pasta, Italian, Mexican, or oriental foods	0.123	0.147	-0.025	12,235
Miscellaneous prepared foods	0.077	-0.026	0.103	3,740
Snacks				
Potato chips and other snacks	0.223	0.245	-0.022	32,232
Nuts	0.145	0.137	0.009	12,553
Alcoholic beverages				
Wine at home	0.073	0.176	-0.103	14,809

Table 2. Total Index Log Changes, 2003-2009
Averages Across Items

	CPI Method	Hedonic Method	Difference
Number of Categories Higher	12	27	-15
Average, Unweighted	0.1478	0.1586	-0.0108
Average, Weighted	0.1584	0.1705	-0.0122
Significance Level of Category Distribution (2 tail)			2.370%
Standard Error of Difference, Unweighted			0.0061
t statistic			-1.77
Significance Level (2 tail)			8.459%
Standard Error of Difference, Weighted			0.0042
t statistic			-2.87
Significance Level (2 tail)			0.659%

Table 3. Contributions to Index Log Changes, 2003-2009
Unweighted Averages Across Items

	CPI Method	Hedonic Method	Difference
Average	0.1478	0.1586	-0.0108
Contributions			
Continuing Items	0.1386	0.1386	0.0000
Comparable Substitutes	0.0061	0.0057	0.0004
Non-comparable Substitutes	0.0031	0.0143	-0.0112

Table 4. Average Monthly Log Price Change, Comparable Substitutes
Averages Across Items

	CPI Method	Hedonic Method	Difference
Number of Items Higher	22	16	6
Number Significantly Higher (5 percent 2 tail)	1	1	0
Unweighted Average Log-Change	0.00977	0.00948	0.00029
Significance Level of Category Distribution (2 tail)			41.77%
Standard Error of Difference			0.00071
t statistic			0.40
Significance Level (2 tail)			68.720%
Average Normalized Category Mean Difference			0.10702
t statistic			0.69
Significance Level (2 tail)			49.460%
Chi-square statistic (All differences zero)			36.13
Significance Level			60.137%

Table 5. Average Monthly Log Price Change, Non-comparable Substitutes
Averages Across Items

	CPI Method	Hedonic Method	Difference
Number of Items Higher	13	26	-13
Number Significantly Higher (5 percent 2 tail)	0	1	-1
Unweighted Average Log-Change	0.00232	0.01037	-0.00806
Significance Level of Category Distribution (2 tail)			5.33%
Standard Error of Difference			0.00274
t statistic			-2.94
Significance Level (2 tail)			0.324%
Average Normalized Category Mean Difference			-0.35509
t statistic			-2.00
Significance Level (2 tail)			5.309%
Chi-square statistic (All differences zero)			51.80
Significance Level			8.234%

Table 6. Log Price Change Regression Coefficients
Averages Across Items

Mean Coefficient on Comparable Substitute Dummy	0.0107
Number of Coefficients Positive	28
Number Significantly Positive (5 percent 2 tail)	5
Number Significantly Negative (5 percent 2 tail)	1
Mean Normalized Coefficient	0.7298
t statistic (Coefficients have zero mean)	2.32
Significance Level (2 tail)	2.59%
Chi-square statistic (All coefficients zero)	167.54
Significance Level	0.00%

Table 7. Average Monthly Log Price Changes
By Method and Situation

	Unweighted Mean Across Categories
CPI Method	
Continuing Items	0.00205
Comparable Substitutions	0.01299
Non-Comparable Substitutions	0.00212
Hedonic Method	
Continuing Items	0.00205
Comparable Substitutions	0.01192
Non-Comparable Substitutions	0.00713

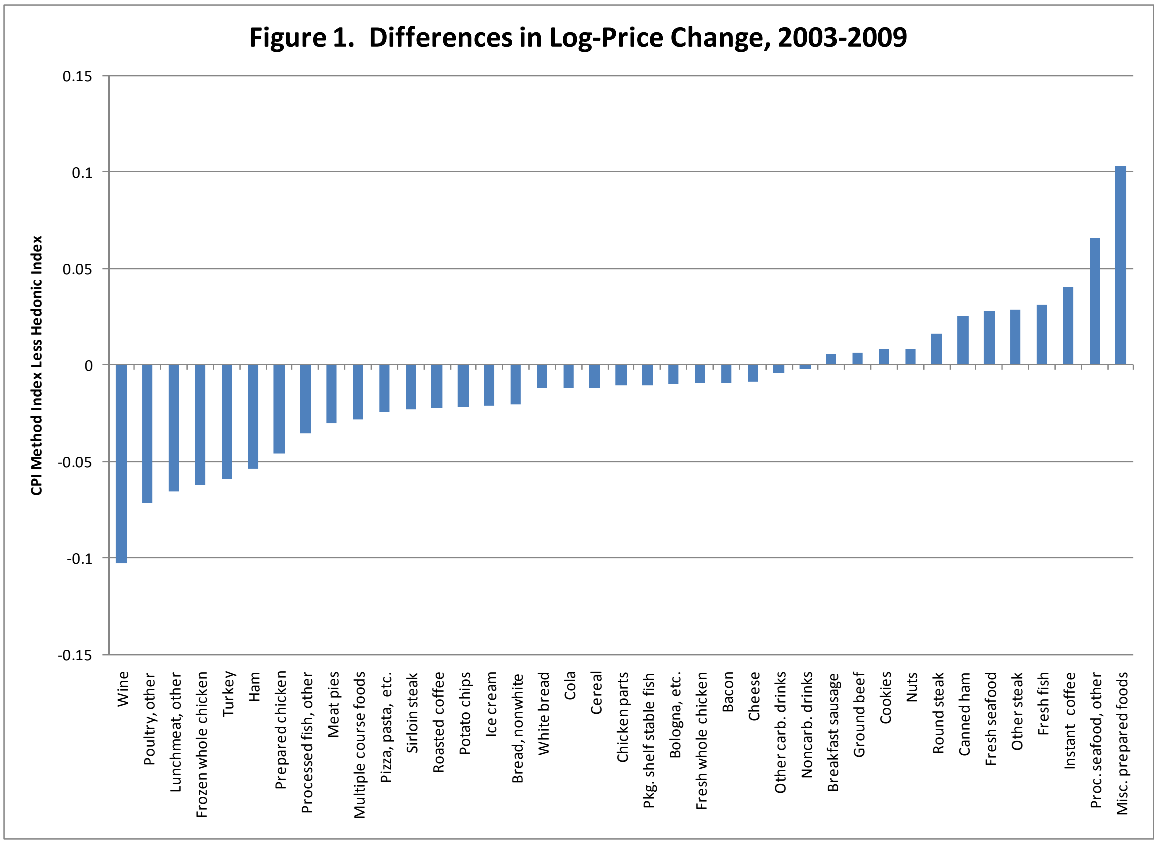

Figure 1. Differences in Log-Price Change, 2003-2009
CPI Method Index Less Hedonic Index
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
Wine
Poultry, other
Lunchmeat, other
Frozen whole chicken
Turkey
Ham
Prepared chicken
Processed fish, other
Meat pies
Multiple course foods
Pizza, pasta, etc.
Sirloin steak
Roasted coffee
Potato chips
Ice cream
Bread, nonwhite
White bread
Cola
Cereal
Chicken parts
Pkg. shelf stable fish
Bologna, etc.
Fresh whole chicken
Bacon
Cheese
Other carb. drinks
Noncarb. drinks
Breakfast sausage
Ground beef
Cookies
Nuts
Round steak
Canned ham
Fresh seafood
Other steak
Fresh fish
Instant coffee
Proc. seafood, other
Misc. prepared foods

References

Aizcorbe, Ana, Carol Corrado and Mark Doms (2000): "Constructing Price and Quantity Indexes for High Technology Goods," Industrial Output Section, Division of Research and Statistics, Board of Governors of the Federal Reserve System, July 19.

Aizcorbe, Ana, Carol Corrado and Mark Doms (2003): "When Do Matched-Model and Hedonic Techniques Yield Similar Price Measures?" FRBSF Working Paper 2003-14, June 16. Available at www.frbsf.org/publications/economics/papers/2003/wp03-14bk.pdf.

Aizcorbe, Ana and Yvon Pho (2005): "Differences in Hedonic and Matched-Model Price Indexes: Do the Weights Matter?" Bureau of Economic Analysis Working Paper WP2005-06, September. Available at www.bea.gov/papers/pdf/hedonicprice.pdf.

Armknecht, Paul A., Walter F. Lane, and Kenneth J. Stewart (1997): "New Products and the U.S. Consumer Price Index," in Bresnahan and Gordon (eds.), *The Economics of New Goods*. NBER Studies in Income and Wealth 58, University of Chicago Press, Chicago, 375-391.

Bils, Mark (2008): "Do Higher Prices for New Goods Reflect Quality Growth or Inflation?" University of Rochester, August. Available at www.econ.rochester.edu/Faculty/BilsPapers/growth_qje_revision_August_2008.pdf. Forthcoming, *Quarterly Journal of Economics*.

Bureau of Labor Statistics (2007): *BLS Handbook of Methods*, "Chapter 17: The Consumer Price Index." Available at www.bls.gov/opub/hom/pdf/homch17.pdf.

Gordon, Robert J. (2009): "Apparel Prices 1914-1993 and the Hulten-Breugel Paradox," in Diewert, Greenlees, and Hulten (eds.), *Price Index Concepts and Measurement.* NBER Studies in Income and Wealth 70, University of Chicago Press, Chicago, 85-128.

Greenlees, John S., and Robert McClelland (2007): "Price Differentials across Outlets in Consumer Price Index Data, 2002-2006," paper presented at the Tenth Meeting of the Ottawa Group, Ottawa, Canada, October. Available at www.ottawagroup.org/Ottawa/ottawagroup.nsf/51c9a3d36edfd0dfca256acb00118404/c0b785f3c0f0f765ca25727300107fd9?OpenDocument#B .

Greenlees, John S., and Robert McClelland (2008): "New Evidence on Outlet Substitution Effects in Consumer Price Index Data," Bureau of Labor Statistics Working Paper 421, November. Available at www.bls.gov/osmr/pdf/ec080070.pdf. Forthcoming, *Review of Economics and Statistics*, August 2011.

International Labour Office (2004): *Consumer Price Index Manual: Theory and Practice*. Geneva: International Labour Office. Available at www.ilo.org/public/english/bureau/stat/guides/cpi/index.htm.

Johnson, David S., Stephen B. Reed, and Kenneth J. Stewart (2006): "Price measurement in the United States: a decade after the Boskin Report" Monthly Labor Review, Vol. 129, May, 10-19. Available at www.bls.gov/opub/mlr/2006/05/art2full.pdf.

Klenow, Peter J., and Benjamin A. Malin (2009): "Microeconomic Evidence on Price-Setting," Prepared for the *Handbook of Monetary Economics* (Benjamin Friedman and Michael Woodford, editors), September.

Klenow, Peter J., and Jonathan L. Willis (2007): "Sticky information and sticky prices," *Journal of Monetary Economics*, vol. 54, September, 79-99.

Lane, Walter E. (2001): "Addressing the New Goods Problem in the Consumer Price Index" *International Working Group on Price Indices: Papers and Proceedings of the Sixth Meeting*. Canberra: Australian Bureau of Statistics, 340-368. Available at www.ottawagroup.org/Ottawa/ottawagroup.nsf/51c9a3d36edfd0dfca256acb00118404/c0b785f3c0f0f765ca25727300107fd9?OpenDocument#B .

Nakamura, Emi, and Jon Steinsson (2008): "Lost in Transit: Product Replacement Bias and Pricing to Market." Columbia University, December 1.

National Research Council (2002): *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes. Panel on Conceptual, Measurement, and Other Statistical Issues in Developing Cost-of-Living Indexes*. Charles L. Schultze and Christopher Mackie, Eds. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Reinsdorf, Marshall (1993): "The Effect of Outlet Price Differentials on the U.S. Consumer Price Index," in Foss, Manser, and Young (eds.), *Price Measurements and Their Uses*. NBER Studies in Income and Wealth 57, University of Chicago Press, Chicago, 227-254.

Silver, Mick and Saeed Heravi (2005): "A Failure in the Measurement of Inflation: Results from a Hedonic and Matched Experiment using Scanner Data." *Journal of Business and Economic Statistics*, vol. 23, no. 3, 269-281.

Triplett, Jack (2006): *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*. Paris, France: OECD Publishing.