Item replacement and quality change in apparel price indexes

Simulations removing commodity analyst intervention in the production of price indexes, including the application of direct hedonic quality adjustments, indicate that experimental apparel price indexes would tend to rise more or fall less than their published counterparts

Craig Brown
and
Anya Stockburger

A principal role of a Bureau of Labor Statistics (BLS) commodity analyst in the Consumer Price Index (CPI) program is that of data reviewer. In that role, commodity analysts are responsible primarily for monitoring the prices and characteristics of the goods and services tracked for the CPI market basket, verifying that the reported price movements of those goods and services meet the criteria for the collection of accurate consumer prices, and developing tools and procedures that aid in the careful collection and evaluation of the price and characteristic data used in producing the CPI.[1] More broadly, the commodity analyst's role is to ensure the accurate calculation of the CPI by eliminating the effects of inaccurate microlevel data.[2]

A commodity analyst's microlevel data review consists largely of examining problematic observations (or price quotes) that meet one or more of the conditions for review defined in the CPI'S data-processing system. Commodity analysts must intervene and resolve the issues related to their assigned quotes before the price data from those quotes are used in computing the CPI. Two of the more common conditions prompting a review are an unusually large price change and the selection of a replacement item as a substitute for an original item that is no longer available. The latter of these conditions—the selection of replacement items—is the primary focus of this article.

The analysis to be presented assesses the impact of commodity analyst intervention on price indexes for apparel items. The definition of "commodity analyst intervention" is limited to "effort spent in the review and analysis of replacement items and changes in quality." Issues inherent in the calculation of apparel price indexes are discussed in order to identify the circumstances in which commodity analyst intervention is most critical. These issues are then further developed through the decomposition of the concept of commodity analyst intervention into two stages of microlevel data review. Following that development, experimental indexes that simulate the effects of removing commodity analyst intervention are calculated and compared with official apparel indexes.[3] Finally, a brief summary of the findings and a few caveats regarding the analysis are offered.

The apparel component of the CPI

The apparel component of the CPI was chosen as the subject of the research described in this article due in part to the difficulty of correctly measuring price change for apparel items and the labor-intensive nature of the microlevel review of price and characteristic data associated with apparel. The seasonal nature of apparel marketing, the large number of item replacement scenarios that occur in the collection of apparel price data, and the emphasis on maintaining a constant-quality price index all necessitate an intensive review.[4] The remainder of this section discusses these three issues in greater detail and explains how commodity analysts address them.

Craig Brown and Anya Stockburger are economists in the Office of Prices and Living Conditions, Bureau of Labor Statistics. E-mail: brown.craig@bls.gov; stockburger.anya@bls.gov

Seasonal nature. Typically, apparel retailers use a seasonal pricing scheme to accommodate the characteristically seasonal nature of apparel commodities. In September 2005, 36 percent of the apparel sample consisted of seasonal merchandise.[5] Prices of seasonal apparel items usually are at their highest levels at the beginning of a season and gradually are reduced until they reach a very low clearance price at the end of the season. At that point, the next season's merchandise is introduced.

Under a seasonal pricing scheme, an apparel item usually is first priced by BLS data collectors at its high, beginning-of-season price. Then the item typically will be priced month after month at continually declining prices, until it is priced at its clearance level and eventually becomes unavailable for sale in the store either because it sold out or was removed from the shelves to make room for new merchandise.

When a seasonal item returns to the store after being unavailable during the months of the opposing season, the data collector verifies that the characteristics of the available item match those of the item that was last priced. If the characteristics match, a price is reported for that item—usually a high, markup price at the beginning of the season. Because the last prices collected for these seasonal items were generally low, end-of-season sale prices, a large price increase typically will occur.

Large price increases of a certain magnitude qualify as triggers for review for all apparel item categories. Commodity analysts are required to examine the quotes in question and determine whether the large price increases are due strictly to the price changes of identical items or whether other factors are contributing. Identical items with large price changes are noted and accepted for price index calculation. Though rare, when it is determined that some other factor is affecting the price change, the usual explanation is that a different item with varying characteristics was inadvertently priced. In this case, commodity analysts classify the situation as an item replacement and incorporate the price of the item into the index.

Item replacement. Along with price fluctuations, the seasonal marketing of apparel goods results in an ever-changing variety of goods that reflects the current fashions and trends. Constantly evolving product mixes mean that data collectors must often find replacements for those items which have exited the market. Apparel commodity analysts allocate a significant amount of their microlevel data review to determining whether the data on prices and characteristics are recorded correctly when item replacements are reported. In 2005, 15 percent of apparel sector quotes were reported as item replacements.

Normally, item replacements are reported correctly. However, on a rare occasion, an error will occur. The types of reporting errors that commodity analysts usually can identify and correct are those in which an item has been replaced, but was not reported as such, and those in which an item was reported as having been replaced, but in fact was not. These infrequent errors occur primarily because of the complexity of official CPI data collection procedures and documents. When such errors are identified, it is the responsibility of the commodity analyst to edit a quote and correctly identify its replacement status before the quote is cleared for use in index calculations.

Quality change. Operationally, the CPI measures the average price change of a fixed market basket of goods and services. The aforementioned two issues surrounding the apparel component of the CPI illustrate how pricing the same apparel goods month after month is particularly difficult. The seasonal nature of apparel marketing leads to frequent item replacement scenarios. Data collectors are instructed to select a replacement item that is similar in quality to the item that is no longer available, but often there are no similar substitute items from which to choose. As a result, a less comparable item is selected, potentially introducing quality change and an associated price differential into the index. Apparel commodity analysts are responsible for reviewing item replacement scenarios and minimizing the unwanted effects of quality changes on the apparel index.

Once the commodity analyst establishes that an item has been replaced, the next step is to determine whether the new item is of similar quality to the old one. This process consists of reviewing the differences between the characteristics of both items and determining whether those differences are insignificant (exhibiting little to no quality change) and the new item is comparable to the old one, or whether the differences are significant and the new item is not comparable to the old one. When a comparable replacement is identified, the price of the new item is compared directly with the price of the old one and the measure of price change between the two items is used in calculating the CPI. When a noncomparable replacement is identified, the price of the new item is held out of the calculation of the index until the next pricing cycle, and the price change for the quote is imputed.

In the apparel component of the CPI, commodity analysts have a third price comparison option available to them when reviewing item replacements. Over the last 15 years, apparel analysts have developed and applied hedonic regression models for making direct quality adjustments to price data.[6] The hedonic method allows analysts to remove any price differential attributed to a change in quality by adding or subtracting the estimated value of that change from the price of the old item. The three price comparison options available to apparel analysts and the hedonic method are discussed in more detail next.

Commodity analyst intervention

In this article, "commodity analyst intervention" is defined in two stages. Stage 1 occurs at the initial phase of the commodity

analyst's review of microlevel price and characteristic data and consists of determining whether a new item has been selected to replace an old one. Stage 2 intervention occurs if a new item was identified in Stage 1. In Stage 2, commodity analysts evaluate the degree of comparability between the new item and the old one and determine how to incorporate its price into the calculation of the index. The remainder of this section decomposes the two stages of commodity analyst intervention further and discusses the ramifications of that intervention on the calculation of the price index.

Stage 1 intervention. The primary function of intervention in Stage 1 of the commodity analyst's review is to determine whether an item replacement scenario has occurred. As mentioned in the previous section, item replacement scenarios are reported correctly in nearly every instance. However, as a result of their specialized knowledge about their assigned commodities, commodity analysts occasionally discover (1) instances in which an item is reported as having been replaced when, in fact, it was not (overreporting) and (2) instances in which an item was replaced, but the replacement was not reported (underreporting).

During the 36 months (September 2002 to August 2005) that data were reviewed for this article, 233,709 prices were collected for use in the calculation of the CPI for Apparel. Data collectors reported 34,788 item replacement scenarios in the apparel sector of the CPI,[7] and apparel commodity analysts recognized and corrected 2,622 instances of overreporting and 1,253 instances of underreporting. These figures represent the lower bound of the actual number of errors incurred, because some errors will go unidentified.

Errors in the identification of item replacement scenarios result in miscalculation of the imputed price relative for noncomparable item replacements. For certain categories of items in the CPI, including all apparel items, this imputed value—the substitution relative—is calculated via a class-mean imputation method. Briefly, the imputed value is the geometric mean of the price changes of comparable and quality-adjusted item replacements in the same index cell.[8]

When an error in reporting an item replacement is identified, the commodity analyst changes the status of the price quote in order for it to be processed correctly in index calculations. If a quote is reported to have undergone item replacement when in fact it did not, changing the status of the quote removes it from the pool of other item replacement quotes used to impute the price change of noncomparable item replacements. Conversely, if it is determined that a quote has undergone item replacement, but the change was not reported, the status of the quote is modified to make it eligible for the processing of the substitution relative.

Another type of imputation methodology, the cell-relative method, is used to calculate price relatives for noncomparable replacements in the remaining item categories of the CPI.[9] The method differs from the class-mean method in its choice of imputation source quotes. Whereas the class-mean method uses only other item replacement quotes, the cell-relative method utilizes both replacement and nonreplacement quotes to impute the price changes of noncomparable replacements. The imputation method selected is critical for items such as apparel, for which price changes are closely associated with the periodic or seasonal introduction of new lines or models. Recall that new items are usually introduced at high prices and are discounted throughout the season. By excluding nonreplacement items from the source pool, the class-mean method estimates a price relative from quotes for replacement items that typically are registering large, positive price changes at the beginning of a season.

Stage 2 intervention. The primary function of commodity analyst intervention in Stage 2 review is to ensure that a constant-quality index is maintained by limiting the index's exposure to price differentials attributed to quality changes introduced through item replacement scenarios. To achieve this objective, apparel commodity analysts rely on their knowledge of the retail industry, apparel commodities, fashion and trends, and marketing to evaluate the comparability of item replacements. Toward that end, item replacements are reviewed and classified as comparable, comparable by means of a quality adjustment, or noncomparable.

In order to make direct quality adjustments to price data when differences in quality are observed in item replacement scenarios, apparel commodity analysts develop and apply hedonic regression models. In the hedonic sense, a product is viewed as a bundle of value-adding characteristics, and the product's price is the sum of the values of its characteristics. The coefficients produced by the hedonic regression model are estimates of these values, or *shadow prices*. When characteristics change, shadow prices are used to estimate the value of the change in quality. This value is then used to adjust the observed price change between the old and new items.

Of the 34,788 item replacements selected during the period examined in this article, 45 percent (15,533) were deemed comparable by commodity analysts. Twenty-seven percent (9,528) were deemed noncomparable. The remaining 28 percent (9,727) received direct quality adjustments.

Beginning in September 2002, apparel commodity analysts began recording alternative comparability decisions—comparable or noncomparable—for the item replacement scenarios that require direct quality adjustments. The decisions describe how commodity analysts would have treated the quality-adjusted item replacements under the hypothetical condition that they could not make hedonic quality adjustments.

Removing hedonic quality adjustments from the calculation of the index illustrates the devolution of a hedonically adjusted

price index into a class-mean price index.[10] The impact of commodity analyst intervention in Stage 2 review is measured by the differences between the hedonic index series and the class-mean index series. The effects of removing the intervention from that stage are further simulated by calculating experimental indexes under the assumption that commodity analysts do not have the choice of classifying highly dissimilar item replacements as noncomparable.

If commodity analysts cannot make comparability decisions (that is, commodity analysts do not intervene), item replacements can be handled in either of two ways. Under the direct comparison method, the prices of new items are directly compared with the prices of the old ones.[11] This method assumes that there is no quality difference between the two items and thus does not require any intervention by commodity analysts in Stage 2.[12] Alternatively, under the matched-model method, only the price comparisons of nonreplacement items are used in calculating the index. The prices of all replacement items are held out of the index, and a price relative is imputed via the cell-relative method. The full impact of commodity analyst intervention during the Stage 2 review is measured by the differences between the hedonic price index series and the two indexes calculated without commodity analyst intervention.

Impact on calculation of index

Explanation of experimental-index simulations. After the scenarios under which commodity analysts intervene in the review of data used in calculating apparel price indexes were identified, the "fingerprints" of such intervention also were identified and were removed from the price data in varying combinations. The price data were then used to calculate five kinds of experimental price indexes, which are presented in exhibit 1 in order of decreasing complexity.[13] The least complex indexes—the matched-model index (5) and the direct comparison index (4)—simulate the removal of all identifiable analyst intervention from the calculation. The class-mean imputation indexes (1 and 3) include the commodity analyst's original comparability decisions, but not the use of hedonic quality adjustments. The hedonic quality adjustment indexes (0 and 2) allow comparability decisions and the use of hedonic quality adjustments. Indexes 2 and 3 do not include Stage 1 intervention. Index 0 is the official CPI for Apparel. The remainder of this section discusses the experimental indexes in the order presented in exhibit 1.

Index 1: class-mean imputation with item replacement verification. Table 1 shows the percent change in Index 1 and in the official CPI between August 2002 and August 2005, and the difference between them, for the aggregate and subaggregate apparel item categories. The table indicates that if apparel commodity analysts did not make hedonic quality adjustments during the period studied, the aggregate apparel index would have exhibited an upward discrepancy of 0.6 percent by falling an average of 0.2 percent less per year than the official hedonic apparel index.[14]

Particularly following the 1996 report by the Senate Finance Committee's Advisory Commission to Study the Consumer Price Index,[15] much has been written about the possible presence of quality change bias in the calculation of the CPI. The Commission reported a 0.6-percent annual bias in the All-Items CPI due to quality change, as well as a large upward bias of around 1.0 percent per year for the apparel sector of the CPI. However, David Lebow and Jeremy Rudd concluded that there is no quality change bias in the apparel component of the CPI.[16] The issue is further complicated by arguments put forth by Charles Hulten and Bart Hobijn concluding that perhaps quality change biases indexes *downward*.[17] Another study, by Robert Gordon, found that quality change bias in apparel items may not be consistently upward or downward and even may be negligible over the long term.[18] The results of this study indicate that, without the application of hedonic adjustments to control for quality differences, the apparel index would have been greater by 0.2 percent per year during the period the study examined.

The results for the aggregate apparel index are in contrast to prior, similar experiments that attempted to measure the impact of applying hedonic quality adjustments. Paul Liegey calculated experimental nonhedonic indexes for seven subaggregate, specific apparel item categories and three aggregate apparel item categories for the latter half of 1991.[19] During that period, the nonhedonic apparel index exhibited a downward discrepancy, rising 0.2 percent less than the official apparel index. As disparate as these results are, a categorical comparison between the results from Liegey's study of the last half of 1991 and the results of the research presented in this article is not advised, because many differing factors contribute to those results.

Among the major factors that could contribute to dissimilar results between the nonhedonic indexes calculated by Liegey and those arrived at herein are (1) changes to the methodology for imputing the price change of noncomparable replacements, (2) an increase in the number of apparel item categories for which hedonic quality adjustments are used, and (3) the different periods for which the nonhedonic indexes are calculated. In 1992, the CPI began using the class-mean imputation method for noncomparable replacements in the apparel sector. Prior to that, a different method was employed.[20] As mentioned in the discussion of Stage 1 intervention, these methodologies can lead to different results. In addition, differences in nonhedonic indexes from 1991 and 2005 are to be expected, because the use of hedonic quality adjustments has expanded and hedonic models are updated periodically. Hedonic quality adjustments were used in 7 item categories in 1991, compared with 10 in 2005.

Exhibit 1. Descriptions of experimental indexes

Index number	Index name	Stage 1 intervention: item replacement verification	Stage 2 intervention	
			Comparability decisions	Quality adjustments
0	Hedonic quality adjustment with item replacement verification (official index)	Yes	Yes	Yes
1	Class-mean imputation with item replacement verification	Yes	Yes	No
2	Hedonic quality adjustment without item replacement verification	No	Yes	Yes
3	Class-mean imputation without item replacement verification	No	Yes	No
4	Direct comparison	No[1]	No	No
5	Matched model without item replacement verification	No	No	No

[1] There is no need for verification with an index calculated under the condition that all item replacements are directly compared.

Finally, in his study of the last 6 months of 1991, Liegey recalculated indexes for those months, while the current article does the same for a full 3 years of data, from September 2002 to August 2005. The choices of the duration and the beginning and ending periods of the simulation may affect the interpretation of the results, due to seasonal pricing patterns. For example, a downward discrepancy of 0.1 percent develops during the first 6 months (August 2002 to January 2003) of the study presented herein, but an upward discrepancy of 0.5 percent develops from January 2003 to July 2003.

The analysis of Index 1 is best continued by decomposing the apparel index into its component indexes. At this level, item category indexes behaved quite differently from each other during the experiment. As shown in table 1, men's apparel, women's apparel, and footwear produced the largest discrepancies between their official and experimental indexes; the reason was that a large majority of the hedonic quality adjustments occurred within those three categories. During the 36-month span of the study, the experimental men's apparel index exhibited a downward discrepancy, falling more than the official hedonic men's apparel index. Contrary to the results for men's apparel, the experimental women's apparel index showed an *upward* discrepancy, falling less than the official hedonic women's apparel index. Similarly, the experimental footwear index produced an upward discrepancy, rising more than the official hedonic footwear index.

Determining what causes these disparate effects across the apparel item categories requires a focused review of available data—a review that will offer the most valuable clues to the behavior of the experimental indexes. The data are the frequency of hedonic quality adjustments in each apparel item category, the distribution of alternative comparability decisions with regard to those quality adjustments, and the difference in the price changes of item replacements affected by the experiment.

Table 2 presents the number of item replacement scenarios in the specific men's apparel, women's apparel, and footwear item categories to which hedonic quality adjustments are applied. A total of 21,905 scenarios occurred in these categories during the 3 years covered by the study. A large majority of the scenarios (63 percent) occurred in women's apparel. The table also presents the distribution of official comparability decisions for item replacements, by specific item category, within the three apparel categories selected. During the 3 years examined, 9,284 hedonic quality adjustments were made to the prices of item replacements in these three apparel categories. Seventy-nine percent of the adjustments were made to the prices of women's apparel items.

The rightmost two columns of table 2 list the distribution of commodity analysts' alternative comparability decisions for

Table 1. Percent change in Index1, August 2002–August 2005

Item category	Index 1	Official CPI	Difference	Average annual difference
Apparel	–3.3	–3.9	0.6	0.2
Men's apparel	–5.1	–3.8	–1.3	–.4
Boys' apparel	–10.2	–9.8	–.4	–.1
Women's apparel	–3.1	–5.1	1.9	.7
Girls' apparel	–6.7	–6.9	.2	.1
Footwear	2.7	1.7	1.0	.3
Infants	–8.3	–8.7	.4	.1
Jewelry and watches	–1.0	–1.0	.0	.0

quality-adjusted item replacements. The distribution used in the experiment varied greatly across specific apparel item categories. Quality-adjusted item replacement scenarios were deemed comparable in commodity analysts' alternative decisions in 50 percent of men's apparel scenarios, 54 percent of women's apparel scenarios, and 37 percent of footwear scenarios. The greater percentage of comparable alternative comparability decisions in men's and women's apparel compared with that in footwear indicates that, on average, item replacements are more similar in quality to the items they are replacing in the former two categories than in the latter.

The implications of the comparability ratios among quality-adjusted item replacement scenarios require further explanation. Commodity analysts found that quality change was minimal in a majority of item replacement scenarios involving item categories with high comparability ratios. Therefore, quality-adjusted item replacements that were deemed comparable in the experiment should have similar price changes in both the official and experimental indexes. By contrast, commodity analysts found that quality change was significant in a majority of item replacement scenarios involving item categories with low comparability ratios. Quality-adjusted item replacements that were deemed noncomparable in the experiment received imputed price relatives by class-mean imputation. The direction and magnitude of these price relatives are unpredictable.

Chart 1 illustrates the relationships between the distribution of percent changes in price for item replacements that were either quality adjusted or deemed noncomparable in the official index (the *y*-axis) and the corresponding percent changes in the experimental index (the *x*-axis). The top panel plots the percent changes for quality-adjusted item replacements that were deemed comparable in the experiment, and the middle panel plots the percent changes for quality-adjusted item replacements that were deemed noncomparable in the experiment. The bottom panel plots the percent changes for noncomparable item replacements in the official index that remained noncomparable in the experimental index.

The top panel of the chart exhibits a linear distribution, implying that the percent changes in price for quality-adjusted item replacements deemed comparable in the experiment differ very little from the corresponding percent changes for official quality-adjusted item replacements. The middle panel exhibits no clear distribution: item replacements that had a large percent change in price in the official index may have a small percent change in the experimental index, and conversely. Like the top panel, the bottom panel exhibits a (somewhat less) linear distribution, illustrating that the imputed substitution relatives varied only slightly after the removal of hedonic quality adjustments from the pool of item replacement quotes used to estimate the substitution relative. When item replacements are analyzed by apparel category and by specific item category, the same relationships hold. Thus, the tendency for discrepancies to develop between the experimental and official indexes can be attributed to the differences in the percent changes in price for quality-adjusted item replacements that were deemed noncomparable in the experiment. However, these data alone cannot predict, to any reasonable degree of accuracy, whether upward or downward discrepancies will develop between the experimental indexes and their official counterparts when hedonic quality adjustments are removed.

Another possible explanation for the differing effects of hedonic quality adjustments on apparel item category indexes is the greater impact of seasonal pricing behavior in women's apparel than in men's apparel and footwear. In an article examining pricing regularities, Peter Pashigian and Brian Bowen explained that seasonal fluctuations in the prices of women's apparel are greater than those in the prices of men's apparel and that prices of women's apparel start relatively higher because retailers are uncertain about what styles will be popular with consumers and eventually resort to reducing prices through sales to clear out leftover inventory.[21] Many apparel items exhibit seasonal price fluctuations, of course, but the severity of the fluctuations is greater in women's apparel than in any other apparel category and may help to

Table 2. Number of item replacement scenarios, by official and alternative comparability decisions

Item category	Number of prices collected	Number of item replacement scenarios	Official comparability decisions			Alternative comparability decisions	
			Comparable	Non-comparable	Quality adjusted	Comparable	Non-comparable
Total	135,005	21,905	7,515	5,106	9,284	4,833	4,451
Footwear	29,072	3,447	1,966	796	685	252	433
Men's footwear	9,949	871	524	253	94	22	72
Women's footwear	9,123	2,576	1,442	543	591	230	361
Men's apparel	58,468	4,744	2,374	1,120	1,250	631	619
Pants	15,066	971	582	211	178	88	90
Suits, sportcoats, and outerwear	22,384	1,763	867	389	507	204	303
Shirts and sweaters	21,018	2,010	925	520	565	339	226
Women's apparel	47,465	13,714	3,175	3,190	7,349	3,950	3,399
Outerwear	11,380	3,393	1,020	776	1,597	643	954
Dresses	12,168	4,664	817	1,188	2,659	1,274	1,385
Separates	23,917	5,657	1,338	1,226	3,093	2,033	1,060

explain the different experimental-index values between item categories. Further investigation of this topic is left for future research.

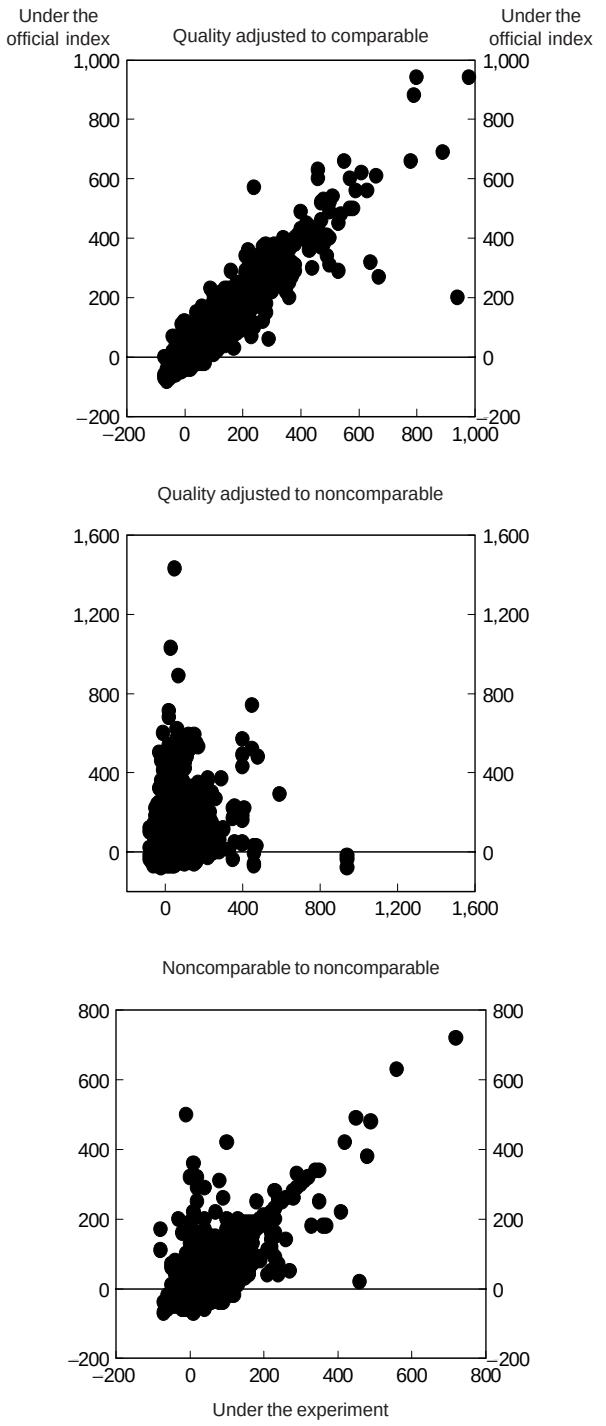
The experiments described here offer further evidence that, although discrepancies between hedonic indexes (those with quality adjustments) and nonhedonic indexes (those without quality adjustments) do exist, no consistent differences appear across item categories or their lower level components. This finding mirrors that of Liegey:[22] the experiments carried out here resulted in seven downward discrepancies and seven upward discrepancies between the experimental, nonhedonic indexes and the official indexes, while Liegey's analysis resulted in five downward discrepancies, three upward discrepancies, and two instances of no discrepancy between the two types of index.[23]

Liegey also suggested that, along with the percentage of item replacements deemed noncomparable after the removal of quality adjustments and differences in relative importance, the complexity of the apparel items in each category, the degree of knowledge of the product exhibited by the commodity analyst, and the capacity of the data collection document to capture the qualities of an item being priced account for some of the differences across item categories.[24] Jack Triplett concluded that discrepancies are determined, not solely by the degree of quality change in the marketplace, but also by marketing practices for different products and by the interaction between those practices and the BLS procedures that attempt to limit the amount of quality change allowed into price comparisons.[25] Other factors that were not researched, however, may contribute to the differences as well. Among these factors are exogenous forces such as seasonal patterns that may influence the availability and pricing of certain categories of items more so than others and endogenous factors such as peculiarities in the calculation of the substitution relative for noncomparable item replacements, variations in the final weights used in calculating aggregate indexes, and variations in the quote weights used in calculating basic indexes. These topics fall outside the scope of this study and are left for future research.

Indexes 2 and 3: hedonic quality adjustment and class-mean imputation without item replacement verification. Experimental Indexes 2 and 3, described in exhibit 1, simulate the behavior of the apparel index had apparel commodity analysts not intervened in the Stage 1 review. This means that all item replacements reported by data collectors entered into the calculation of the index as item replacements, regardless of whether they actually were. In total, only 3,875 price quotes were directly affected by the removal of Stage 1 intervention. For Index 2, commodity analyst intervention was removed from Stage 1, but commodity analysts were allowed an unrestricted Stage 2 review of item replacements, with the full range of comparability options (comparable, noncomparable, and quality adjusted). For Index 3, Stage 1 intervention was removed and Stage 2 review was restricted by removing the option to apply hedonic quality adjustments to item replacements. Table 3 shows the percent changes in Indexes 2 and 3 and in the official CPI between August 2000 and August 2005, and the differences of those indexes from the official CPI, for the aggregate and subaggregate apparel item categories.

The results obtained for Index 2 indicate that Stage 1 intervention had small and ambiguous effects on apparel indexes during the period studied. The removal of Stage 1 intervention resulted in monthly discrepancies between –0.3 percent and 0.2 percent of the official index during the 3 years covered in the study. The average annual percent discrepancy was less than one-tenth of 1 percent of the official apparel index. Commodity analyst intervention during a Stage 1 review veri-

Chart 1. Distribution of percent changes in prices for item replacements affected by the removal of hedonic quality adjustments



fies that item replacement has occurred, thereby ensuring that the pool of quotes used to calculate the substitution relative consists only of quotes undergoing item replacement. It is difficult to formulate credible expectations of the behavior of price indexes for which the substitution relatives have been altered. Therefore, the inconsistency in the results obtained for Index 2 across item categories is not surprising.

Combining the removal of commodity analyst intervention in the Stage 1 review with the removal of hedonic quality adjustments from the Stage 2 review produces an upward discrepancy between Index 3 and the official apparel index. Index 3 fell an average of 0.2 percent per year less than the official index. Although the effect of removing commodity analyst intervention from Stage 1 is somewhat small and ambiguous, as evidenced by the results for Index 2, Index 3 reacts to the upward discrepancy that develops from the removal of hedonic quality adjustments from the Stage 2 review.

Indexes 4 and 5: direct comparison and matched model without item replacement verification. Indexes 4 and 5 simulate the CPI for Apparel under the condition that commodity analysts did not intervene at all in the review of item replacements. If there is no analyst review of quality change, then an assumption must be made about the existence of quality change in price comparisons. Index 4 is calculated under the assumption that there is no quality change, and the prices of new items are directly compared with the prices of the old items they replaced. Computationally, all item replacements are deemed comparable in the calculation of Index 4, resulting in an index calculated on the basis of direct comparisons of items. Because the only implication of Stage 1 intervention is the effect on the imputed price relative for noncomparable item replacements, removing Stage 1 intervention has no effect in this experiment, given that all item replacements are directly compared. By contrast, Index 5 is calculated under the assumption that quality is changing, and the prices of new items are not compared against the prices of the items they replaced. This matched-model index is generated by comparing prices of nonreplacement items, and the price change of replacement items is imputed via the cell-relative imputation scheme. Unlike the situation with Index 4, removing Stage 1 intervention from the calculation of Index 5 affects the source pool for the imputation of the noncomparable price relatives in a matched-model index. Table 4 shows the percent changes in Indexes 4 and 5 and in the official CPI between August 2000 and August 2005, and the differences of those indexes from the official CPI, for the aggregate and subaggregate apparel item categories.

The calculation of Index 4 results in an upward discrepancy from the official index. Given the direct-comparison nature of Index 4, proportionally more low, end-of-season prices are compared with high, beginning-of-season prices, causing the experimental index to fall less than the official index. The earlier discussions of the class-mean indexes (Indexes 1 and 3) illus-

Table 3. Percent changes in Indexes 2 and 3, August 2002–August 2005

Item category	Index 2	Index 3	Official CPI	Difference		Average annual difference	
				Index 2 minus official CPI	Index 3 minus official CPI	Index 2 minus official CPI	Index 3 minus official CPI
Apparel	–3.8	–3.2	–3.9	0.1	0.7	0.0	0.2
Men's apparel	–4.1	–5.1	–3.8	–.3	–1.3	–.1	–.4
Boys' apparel	–8.7	–9.6	–9.8	1.1	.2	.4	.1
Women's apparel	–5.2	–3.3	–5.1	–.1	1.8	.0	.6
Girls' apparel	–7.1	–7.0	–6.9	–.2	–.1	–.1	.0
Footwear	1.5	2.3	1.7	–.2	.6	–.1	.2
Infants' apparel	–8.5	–8.3	–8.7	.2	.4	.1	.1
Jewelry and watches	1.2	1.2	–1.0	2.2	2.2	.7	.7

Table 4. Percent changes in Indexes 4 and 5, August 2002–August 2005

Item category	Index 4	Index 5	Official CPI	Difference		Average annual difference	
				Index 4 minus official CPI	Index 5 minus offical CPI	Index 4 minus official CPI	Index 5 minus official CPI
Apparel	–1.7	–46.9	–3.9	2.2	–43.1	0.7	–14.4
Men's apparel	–3.0	–33.1	–3.8	.8	–29.3	.3	–9.8
Boys' apparel	–4.0	–49.0	–9.8	5.8	–39.2	1.9	–13.1
Women's apparel	–3.8	–65.9	–5.1	1.3	–60.8	.4	–20.3
Girls' apparel	1.6	–57.4	–6.9	8.5	–50.5	2.8	–16.8
Footwear	.4	–42.5	1.7	–1.3	–44.2	–.4	–14.7
Infants' apparel	–7.4	–36.7	–8.7	1.3	–28.0	.4	–9.4
Jewelry and watches	9.9	–5.3	–1.0	10.9	–4.3	3.6	–1.4

trated how the partial removal of commodity analyst intervention in the Stage 2 review tends to produce an upward discrepancy between the experimental and official apparel indexes. Under the assumption that quality change is minimal or nonexistent, and with the removal of commodity analyst intervention in Stage 2 review, Index 4 falls at an average annual rate of 0.7 percent less than the rate at which the official index falls.

The calculation of Index 5 results in a downward discrepancy from the official index. Given the matched-model nature of Index 5, no low end-of-season prices are compared with high beginning-of-season prices. Under the assumption that quality change errors are introduced with all item replacements, this matched-model index falls at the high average annual rate of 14.4 percent more than the rate at which the official index falls.

THIS ARTICLE HAS SHOWN THAT SEVERAL EXPERIMENTAL PRICE INDEXES produced without commodity analyst intervention tend to either rise more or fall less than official BLS indexes produced with commodity analyst intervention. The lone exception is a matched-model index, which falls much more rapidly than the official index. Over the 3 years of data used in the study, the experimental indexes in which all levels of commodity analyst intervention were removed produced the largest discrepancies from the official indexes. Experimental indexes simulating the removal of hedonic quality adjustments also produced discrepancies from the official indexes, although these appeared with varying signs and magnitudes at lower levels of aggregation.

Clearly, commodity analysts have an impact on the CPI for Apparel, not only by affecting the data, but also by maintaining the sample of price quotes, used in calculating the index. In truth, it is impossible to measure the total impact of commodity analyst intervention on the calculation of price indexes simply by defining and conducting the experiments presented in this article. For instance, the impact of hedonic regression models goes beyond their application by commodity analysts in quality adjusting item replacements during the Stage 2 review. The alternative comparability decisions used to calculate the class-mean experimental indexes were made by apparel commodity analysts, whose knowledge of the existing hedonic models no doubt affected their judgment. In addition, commodity analysts are responsible for the design and functionality of the electronic data collection tools used by CPI data collectors. The unique market knowledge commodity analysts possess, as well as their

knowledge of hedonic models, is instrumental in the creation of data collection tools. As a result, commodity analysts indirectly affect the CPI by regulating the composition of the goods and services that are sampled. The tasks involved in the commodity analyst's intervention in producing the CPI are far too complex and indivisible to extract fully from the calculation of the index. Therefore, the findings presented in this article should be regarded as an estimate of the impact of only a portion of the effect of commodity analyst intervention on apparel price indexes. □

Notes

[1] See Chapter 17 of the *BLS Handbook of Methods* (Bureau of Labor Statistics, 2006) for a complete discussion of the construction of the CPI and the role of commodity analysts.

[2] The Bureau took a significant step forward in improving the quality of microlevel data used to compute the CPI when it began using computer-assisted data collection in 2002. Prior to that changeover, the collection, transmission, and transcription of data were subject to a large degree of human error. Now, automatic checks and reviews help improve the completeness and accuracy of CPI data.

[3] The Bureau uses the term "experimental" in contrast to "official" to denote statistics that it produces outside of its regular production systems and, consequently, with less than full production quality. For security reasons, BLS researchers cannot produce experimental statistics until after the publication of the corresponding official statistics. To obtain experimental series referred to in this article, contact either of the authors.

[4] See Dennis Fixler, Charles Fortuna, John Greenlees, and Walter Lane, "The Use of Hedonic Regressions to Handle Quality Change: The Experience in the U.S. CPI," paper presented at the Fifth Meeting of the International Working Group of Price Indices, Reykjavik, Iceland, August 1999, pp. 6–9, for a review of the difficulties inherent in constructing accurate price indexes for apparel items.

[5] In September 2005, the apparel sample of the CPI consisted of 6,316 priced quotes. At that time, 4,018 quotes represented unique items that were identified as being available for sale year round (nonseasonally). The remaining 2,298 seasonal quotes represented unique items that were identified as being available for sale only during certain months of the year.

[6] Modern research on hedonic theories and methods generally starts with Zvi Griliches, "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," in *The Price Statistics of the Federal Government*, General Series No. 73 (New York, Columbia University Press, 1961). For more information on early work with hedonic regression models for apparel items in the CPI, see Efthemia V. Georges and Paul R. Liegey, Jr., *An Examination Using Hedonic Regression Techniques to Measure the Effects of Quality Adjustment on Apparel Indexes*, internal report (Bureau of Labor Statistics, 1988); Paul A. Armknecht and Donald Weyback, "Adjusting for Quality Change in the U.S. Consumer Price Index," *Journal of Official Statistics*, June 1989, pp. 107–23; Paul R. Liegey, Jr., "Adjusting Apparel Indexes in the Consumer Price Index for Quality Differences," in Murray F. Foss, Marilyn E. Manser, and Allan H. Young, *Price Measurements and Their Uses*, National Bureau of Economic Research Studies in Income and Wealth, 57 (Chicago, University of Chicago Press, 1993), pp. 209–26; Paul R. Liegey, Jr., "Apparel price indexes: effects of hedonic adjustment," *Monthly Labor Review*, May 1994, pp. 38–45; and Nicole Shepler, *Analysis of Hedonic Regression: Applied to Women's Apparel in the Consumer Price Index*, internal manuscript (Bureau of Labor Statistics, 1994).

[7] The number of apparel item replacement scenarios reported during the 3 years of research for this article (34,788) includes 948 scenarios (2.7 percent) that are not technically item replacements. These scenarios are referred to as *reinitiations* and are the result of resampling the price quotes and selecting replacement items through the initiation procedures determined by the Bureau. Reinitiations differ from item replacements in that data collectors are not directed to select the next-most-similar item to sample when reinitiating a price quote. Data collectors are directed to reinitiate quotes for a number of reasons, most notably when an item is ineligible in the specific category being sampled or when a category is selected for item rotation (a method of keeping the sample current by periodically reinitiating the quotes). Commodity analysts are required to review reinitiations in the same manner that they review item replacements. Reinitiations are included in this analysis because they share many characteristics with item replacements in terms of how they are reviewed.

[8] Class-mean imputation is explained in greater detail in the *BLS Handbook of Methods*, Chapter 17, p. 23.

[9] Cell-relative imputation for noncomparable item replacements is used primarily in the foods and services sectors of the CPI. The method also is used to impute a price relative for nonreplacement items for which prices are unavailable during a collection period. Cell-relative imputation is explained further in the *BLS Handbook of Methods*, Chapter 17, p. 23.

[10] A "hedonic price index" is often defined as an index that uses prices imputed directly from a hedonic regression, either as a time dummy coefficient or by using the coefficients of the characteristics variables to impute a value for the dependent variable. The hedonic adjustment method used by the Bureau is what Mick Silver and Saeed Heravi call "patching" and simply means that adjustments for quality differences are made to noncomparable models and the adjusted, "patched" price is used for price comparisons. (See Mick Silver and Saeed Heravi, "Scanner Data and the Measurement of Inflation," *Economic Journal*, June 2001, pp. 384–405.)

[11] A complete academic description of quality adjustment and conventional price index methodologies can be found in Jack Triplett, *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, OECD Science, Technology and Industry Working Papers, No. 9 (Geneva, OECD Publishing, 2004).

[12] Triplett, *Handbook on Hedonic Indexes*, notes that the direct comparison method may be the appropriate way of handling changes in quality under the "most similar item" replacement rule employed by the Bureau. The rule instructs its data collectors to choose item replacements with characteristics similar to those of the item that "disappeared."

[13] The experimental indexes were calculated with an in-house program that mimics the functions of the official CPI algorithm. This approach represents a noteworthy improvement over previous BLS experimental indexes used for research purposes, which were limited in their ability to replicate the official algorithm.

[14] The distinction between upward and downward discrepancies is made in Jack Triplett, "Determining the effects of quality change on the CPI," *Monthly Labor Review*, May 1971, pp. 27–32 (see especially note to table 4, p. 30), and in Liegey, "Apparel price indexes," p. 41. Borrowing from Liegey, an "upward discrepancy" means that the experimental index rises more than the official price index if prices are rising or falls less than the official index if prices are falling. Conversely, a "downward discrepancy" means that the experimental index rises less than the official price index if prices are rising or falls more than the official index if prices are falling.

[15] Michael J. Boskin, Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, and Dale W. Jorgenson, *Final Report of the Advisory Commission to Study the Consumer Price Index* (Washington, DC, U.S. Government Printing Office, 1996).

[16] David E. Lebow and Jeremy B. Rudd, "Measurement Error in the Consumer Price Index: Where Do We Stand?" *Journal of Economic Literature*, March 2003, pp. 159–201.

[17] Charles R. Hulten, "Quality Change in the CPI," *Federal Reserve Bank of St. Louis Review*, May/June 1997, pp. 87–111; and Bart Hobijn, "On Both Sides of the Quality Bias in Price Indexes," *Federal Reserve Bank of New York Staff Reports*, no. 157, December 2002, pp. 1–40. Neither of these articles offers an estimate of quality change bias specifically for the apparel item sector of the CPI.

[18] Robert J. Gordon, "Apparel Prices 1914–93 and the Hulten/Brueghel Paradox," paper presented at the CRIW Conference on Price Index Concepts and Measurement, Vancouver, British Columbia, Canada, June 28–29, 2004.

[19] Liegey, "Apparel price indexes."

[20] Before the substitution relative procedure began in 1992, the apparel section used a linking method to ensure that apparel items did not exit the sample while on sale. Instead of using cell-relative imputation, analysts compared an overlap price estimated from the old item's last collected regular price with the new item's price.

[21] B. Peter Pashigian and Brian Bowen, "Why Are Products Sold on Sale?: Explanations of Pricing Regularities," *Quarterly Journal of Economics*, November 1991, pp. 1015–38.

[22] Liegey, "Apparel price indexes."

[23] Liegey recalculated indexes for 7 item categories (3 men's apparel items and 4 women's apparel items), 2 aggregate indexes (men's and women's apparel), and an aggregate apparel index, for a total of 10 experimental, nonhedonic indexes. The analysis presented herein calculates experimental indexes for 10 item categories (3 men's apparel items, 3 women's apparel items, 1 boys' apparel item, 1 girls' apparel item, and 2 footwear items), 3 aggregate indexes (men's apparel, women's apparel, and footwear), and an aggregate apparel index, for a total of 14 experimental, nonhedonic indexes.

[24] Liegey, "Apparel price indexes."

[25] Triplett, "Determining the effects of quality change."