

BLS WORKING PAPERS

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



New Evidence on Outlet Substitution Effects in Consumer Price Index Data

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

Working Paper 421
November 2008

New Evidence on Outlet Substitution Effects in Consumer Price Index Data

by

John S. Greenlees
Robert McClelland

U. S. Bureau of Labor Statistics

November 24, 2008

The authors thank Aylin Kumcu for valuable research assistance. We also thank Ronald Johnson, Peter Klenow, Marshall Reinsdorf, and the participants at the 2007 meeting of the International Working Group on Price Indices, in the Division of Price and Index Number Research seminar series, at a 2008 Summer Institute session of the Conference on Research on Income and Wealth, and at the 2008 meeting of the Southern Economic Association for helpful comments and suggestions on earlier drafts. 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

In this paper we provide new and detailed evidence on the impact on the U.S. CPI of the appearance and growth of new types of product outlets. Using actual CPI microdata for 2002-2007, we find that the changing mix of outlets had a statistically significantly negative impact on average prices in most of the 14 item food categories we study. In contrast to previous studies of this issue, our approach allows us to examine the effects of changes in outlet mix both across outlet types (such as among large groceries, discount department stores, and warehouse club stores) and within those outlet categories. We also adjust for numerous differences in item characteristics such as brand name, organic certification, and, importantly, package size. In our sample we find that the upward impact on price from increased item quality has offset most of the downward impact of lower-priced outlets. We also provide evidence showing that a simulated “matched-model” approach similar to that used in the CPI yields indexes that differ to a surprising extent from our baseline hedonic indexes, which also hold outlet and item mix constant.

I. Introduction

In this paper we provide new evidence on the impact on the U.S. Consumer Price Index (CPI) of the appearance and growth of new types of product outlets. For decades, analysts both within and outside the Bureau of Labor Statistics (BLS), the agency that produces the CPI, have known that consumers can benefit when new stores and delivery channels offer lower prices. Examples of these new outlets have included chain store supermarkets, supercenters, “big box” discounters, warehouse clubs, and the internet, and many of the associated trends in consumer shopping patterns are still continuing.

Unfortunately, obstacles both conceptual and operational have precluded statistical agencies like the BLS from fully incorporating those benefits into price indexes. Some of these same factors have made it difficult for researchers to estimate the resulting potential index bias. The most recent analysis using BLS data, which has informed almost all expert estimates of overall CPI bias, is based on the period 1987-1989.

The research we present here uses regression analysis to compare food prices across CPI outlets during the years 2002 through 2007. In addition to providing estimates for a more recent time period, we are able to go beyond previous work in several ways by using the CPI Research Database developed by BLS. Notably, we have detailed information on outlet type, as well as on the detailed characteristics of individual items priced in the CPI. Ours is the first research to adjust for differences across outlets in the specific characteristics of items sold.

Over the time period we study, we find that CPI food samples exhibited steadily increasing shares of prices from discount department stores and from warehouse club stores. We observe this trend for each of the 14 item categories we study. This is consistent with the national trends reported for the grocery industry as a whole. Despite these trends, however, large grocery stores remain the predominant outlet type in our samples.

We analyze the new outlets issue by estimating, for each of our 14 item categories, a regression model in which the price of an item is a function of variables representing time and item characteristics, plus fixed effects for each of our sample outlets. This enables us to perform statistical tests of whether these outlet fixed effects vary over time and thereby whether outlet mix affects the estimate of price change. We find that the changing mix of outlets between 2002 and 2007 had a statistically significantly negative impact on average prices in most of the 14 item categories.

We then analyze the ways in which this price-reducing impact of the changing outlet mix is related to the growth or decline of particular outlet types. In doing so we are able to distinguish not just between new and traditional store types but also between the categories of low-cost outlets. Notably, we find that the patterns of prices and sample shares for discount department stores differed from those for warehouse club stores. We also estimate the effects of changes in outlet mix both across outlet types (such as between large groceries and discount department stores) and within those outlet categories, and find that within-category changes account for more than a third of the total outlet effect.

We also are able to adjust for numerous differences in item characteristics, which exist even within the relatively homogeneous item categories on which we, following previous authors, focus. In our sample we find that the upward impact on price from increased item quality has offset most of the downward impact of lower-priced outlets.

II. New Outlet Bias

Analysts have long recognized the potential problems caused for a Consumer Price Index by the appearance of new outlets. Feasible solutions for those problems have been difficult to identify, however.

It is important at the outset to distinguish the problem of new outlets from the substitution bias that can arise when there is a change in the relative prices charged at different outlets. For example, in response to an increase in sales or excise taxes in one local jurisdiction, consumers may shift their purchases of gasoline or apparel to outlets in an adjoining area. In this situation, changes in a CPI exceed changes in a cost-of-living index (COLI) unless (1) the CPI is based on a representative sample of outlets in different jurisdictions, and (2) the CPI employs an index formula that allows for consumer response to relative price change. This substitution bias is addressed in the U.S. CPI through its probability sampling and continuous rotation of outlets—albeit with a lag—and by its use of a geometric mean formula, which will approximate a COLI if consumers exhibit a roughly unitary elasticity of substitution across outlets.

As noted in the recent *Consumer Price Index Manual* published by the International Labour Office,¹ the bias from new outlets is conceptually identical to the well-known problem of new product bias. The introduction of a replacement model of computer with improved speed and storage capability is equivalent to the introduction of a remodeled grocery store with better lighting and faster checkout handling. The appearance of a wholly new product type, such as a mobile telephone that can take photographs, is conceptually equivalent to the appearance of a new outlet type, such as an Internet site that offers DVD rentals. In some cases the new good and new outlet are combined, as in the example given in the Boskin Commission report on the CPI of Tuscan and Thai restaurants that brought to American consumers a wider variety of ethnic food specialties.²

The concern of this paper, however, is with the appearance of new outlets that offer lower prices for products essentially identical to those available at existing stores. That issue has been the focus of most prior discussions, and empirical analyses, of outlet bias.

In general, statistical agencies do not construct basic CPI indexes by averaging together prices drawn from different outlets. First, samples of items and outlets are selected, and then the item prices are collected on a monthly or other recurring basis within the sample of outlets. The index is computed as an average (the exact form of which depends on formula and weighting) of the changes over time for the sampled item-outlet pairs. Those changes are measured as ratios of prices, and longer run changes are estimated by multiplying those ratios together.

¹ International Labour Office (2004), p. 213.

² U.S. Senate, Committee on Finance (1996), p. 24.

For example, elementary item/area indexes for food in the CPI employ a geometric mean formula. The log change in the index between times 1 and 2 for a sample of outlets $i=1,\dots,N$ is given by

$$\ln\left(\frac{I_2}{I_1}\right) = \sum_i w_i \ln\left(\frac{P_{i,2}}{P_{i,1}}\right) \quad (1)$$

where we assume for simplicity that only one item is priced in each outlet and we abstract from some computational details in the calculation of the sampling weights w_i attached to the different outlets. For convenience we also assume that the w_i are share weights summing to unity.

The log change in the index between times 1 and 3 is given by

$$\ln\left(\frac{I_3}{I_1}\right) = \sum_i w_i \ln\left(\frac{P_{i,2}}{P_{i,1}}\right) + \sum_i w_i \ln\left(\frac{P_{i,3}}{P_{i,2}}\right) \quad (1a)$$

Rearranging,

$$\ln\left(\frac{I_3}{I_1}\right) = \sum_i w_i \ln(P_{i,3}) - \sum_i w_i \ln(P_{i,1}) \quad (1b)$$

Thus, the log change in the index is the difference between the weighted averages of log prices in periods 3 and 1.

Now let time 2 be an “overlap” period in which a new outlet sample $j=1,\dots,M$ is introduced. This new outlet sample will be accompanied by new sampling share weights w_j^n that reflect purchasing patterns in a more recent period than do the weights for the units in the outgoing sample. Prices P_j^n are the prices from the new outlet sample.

Then the change in the index between times 2 and 3 is defined by

$$\ln\left(\frac{I_3}{I_2}\right) = \sum_j w_j^n \ln\left(\frac{P_{j,3}^n}{P_{j,2}^n}\right) \quad (2)$$

In this case, the log change in the index between times 1 and 3 is found by combining (1) and (2),

$$\ln\left(\frac{I_3}{I_1}\right) = \left[\sum_i w_i \ln\left(\frac{P_{i,2}}{P_{i,1}}\right) \right] + \left[\sum_j w_j^n \ln\left(\frac{P_{j,3}^n}{P_{j,2}^n}\right) \right] \quad (3)$$

and rearranging

$$\ln\left(\frac{I_3}{I_1}\right) = \left[\sum_j w_j^n \ln(P_{j,3}) - \sum_i w_i \ln(P_{i,1}) \right] - \left[\sum_j w_j^n \ln(P_{j,2}^n) - \sum_i w_i \ln(P_{i,2}) \right] \quad (4)$$

Equation (4) shows that the change in log index level can be written as the difference between the log-mean price in period 3 in the new sample and the log-mean price in period 1 in the old sample, *less* the difference in log-mean prices charged by the two sets

of outlets in time 2. The method used in the CPI implicitly subtracts the difference in average prices from the direct comparison measure. Only if that difference is zero will the two-period change be the difference in weighted averages, as was true in (1b). Whether one views that subtraction as appropriate depends on one's views concerning the observed differences in prices across outlets. If consumers view outlets as equivalent except for the prices those outlets charge, then the first term in (4) would provide a better approximation than the CPI index to changes in the cost of living. Conversely, if prices in different outlets are considered equal on a quality-adjusted basis, then incorporating the second term in (4) is essential in order to avoid index bias.

The recent Committee on National Statistics report *At What Price*³ provides a clear and careful discussion of the specific issues raised by the handling of new outlets in the U.S. CPI. Within each item and area category in the CPI, the BLS develops an outlet sampling frame using the Telephone Point-of-Purchase Survey, or TPOPS. Outlets are sampled from the TPOPS frame in proportion to their estimated sales within the item category. Then, BLS staff select individual items for pricing within the store, again using a probability-proportional-to-size procedure.⁴ This process ensures that the CPI sample will include a wide range of specific items in each category. At the same time, it makes it unlikely that the sets of outlets entering and leaving the sample will be represented by identical items, even when their distributions of products sold are similar. This complicates the analysis of potential outlet bias and would likely also complicate the implementation of any solutions.

The implicit assumption used in the CPI is that any cross-sectional differences in the prices charged in different outlets for the same item are attributable to outlet-related variation in "quality": stores offering lower prices may be less conveniently located, have inferior customer service, offer more limited product selection or hours of operation, and so on. Intuitively, in a static equilibrium in which outlets offer different prices there must be exactly offsetting differences in outlet quality. If not, one outlet would increase its share of the market.

The CPI assumption of equal quality-adjusted prices across outlets is not just consistent with the equilibrium assumptions used in numerous economic analyses, it is convenient to implement. It is called into question, however, by observable trends in consumer shopping patterns such as the growth in chain-store supermarkets in the 1950s and 1960s. More recently, the ongoing increase in the market shares of supercenters and warehouse club stores has been a prominent feature of many product markets.⁵ One explanation for this increase would be that, even after quality adjustment, prices at those stores are lower than at more traditional stores.

In this paper we do not attempt to reach definitive conclusions about quality-adjusted price differentials. Examination of store-related quality characteristics and estimation of their value to consumers are left for future research. Our focus here is on whether, in CPI data, prices are systematically lower at some outlets than at others, and whether there are trends in the average outlet-related premium or discount. In estimating the size and

³ National Research Council (2002), pp. 167-177.

⁴ For details on this and other aspects of CPI procedures see Bureau of Labor Statistics (2007).

⁵ See, for example, Michael Stroppe (2006).

statistical significance of these differences we are able to adjust for detailed characteristics of the items sold at sample outlets rather than assuming that all products within an item category are essentially equivalent.

III. Previous Empirical Research on CPI New Outlet Bias

As far back as the 1960s, the BLS carried out an empirical examination of potential bias in the CPI from the appearance of new outlet types. Ethel Hoover and Margaret Stotz (1964) cited Census data showing that the percentage of U.S. food sales accounted for by chain stores rose from 34 percent to 44 percent between 1948 and 1958. The BLS introduced those 1948 weights into the CPI at the end of 1955 and the 1958 weights late in 1961, with several interim adjustments during the intervening years. In each case, however, the new weights were introduced in such a way as to eliminate any impact on the index level of the difference between the mean price levels in chain stores and traditional stores. Hoover and Stotz re-computed the index without that linking procedure for five selected cities. Their results indicated that food prices rose 7.3 percent percentage points over the 1955-1961 period, compared to 8.0 percent for the corresponding CPI five-city average—a difference of about 0.1 percentage point per year.

Alan White (2000) analyzed Canadian CPI indexes for ‘other household equipment,’ non-prescribed medicines, and audio equipment for Ontario for 1990-1996. He showed that those indexes had higher rates of inflation than alternative indexes based on either a unit value approach or one that explicitly calculates changes in the market shares of different outlet types. He also estimated the potential bias from using of an unrepresentative sample of outlets. Those two biases combined were estimated as between 0.2 and 0.4 percentage points per year for the Canadian CPI as a whole.

Unquestionably the most influential study of outlet bias in the CPI has been Marshall Reinsdorf’s 1993 paper. After carefully reviewing the relevant theoretical and measurement considerations, Reinsdorf presented a comparison of prices in incoming and outgoing CPI rotation samples that is closely related to the method used in this paper. During the 1987-1989 period he analyzed, the BLS introduced entirely new outlet and item samples in one-fifth of the CPI geographic areas each year. (We discuss the current four-year TPOPS rotation process in Section IV below.) Reinsdorf selected and pooled 35 reasonably homogeneous CPI food categories, such as flour, eggs, and butter, and computed the percentage changes in price between the old and new samples in 16 cities that underwent rotation during calendar year 1987 or July 1988-June 1989. For all areas pooled, the new sample average prices were 1.23 percent lower than the old sample average, that difference being statistically significant at the five percent level. Given a five-year rotation cycle, this would imply an upward bias in the CPI food at home component of 0.24 percentage point per year. The estimate is an upper bound, however; it “... may possibly overstate the true outlet substitution bias because average quality in the new samples may have declined along with average prices.”⁶ Reinsdorf obtained a similar difference for motor fuel, although that estimate was not statistically significant.

These results of Reinsdorf have provided the basis for almost all subsequent estimates of overall CPI new outlet bias. David Lebow, John Roberts, and David Stockton (1994)

⁶ Reinsdorf (1993), p. 239.

estimated that 40 percent of the CPI was subject to outlet bias; multiplying this by Reinsdorf's bias estimate for food and energy they obtained a 0.1 percentage point estimate for the CPI as a whole. Because of the possible effect of outlet quality differentials, their paper presented both a high-end bias estimate of 0.1 percentage point and a low-end estimate of zero. The Boskin Commission used Lebow *et al.*'s high-end 0.1 percentage point estimate in their report to the Senate Finance Committee.⁷ Matthew Shapiro and David Wilcox (1996) elaborated on this by assigning a log-normal distribution to their outlet bias estimate, with a mean of 0.1 percentage point per year and 90 percent of its mass to the left of 0.2 percentage point. Finally, Lebow and Jeremy Rudd (2003) employed the 0.05 percentage point center of the Lebow-Roberts-Stockton range as their point estimate of new outlet bias, with a confidence interval ranging from zero to 0.2 percentage point annually.

In contrast to all these estimates, Jerry Hausman and Ephraim Leibtag have recently evaluated CPI new outlet bias using data from the ACNielsen Homescan survey. For our present purposes, their most relevant results are comparisons of prices between different store types, in 37 U.S. cities, for 20 relatively homogeneous grocery store food categories. These 20 item categories include thirteen that were also studied by Reinsdorf (1993). Pooling across the cities, Hausman and Leibtag computed the ratios of unit value average prices in traditional supermarkets to those in supercenters, mass merchandisers, and club stores (SMCs). The ratios averaged 1.300 and ranged as high as 2.117 (for lettuce). For only one item category—soda—was the ratio less than unity. Similar ratios with supermarkets replaced by all non-SMC stores were very similar.

Hausman and Leibtag (2004) go on to model how the growing SMC market penetration affects market-average prices, both directly and indirectly through the prices charged by non-SMC stores. They conclude that annual CPI food-at-home inflation is too high by 0.32 to 0.42 percentage point. In Hausman and Leibtag (2005), they employ a discrete choice model of household shopping choice to conclude that the compensating variation value to consumers of SMC entry is 25 percent of food expenditure.

IV. Methodological Approach and Data

As we noted in Sections II and III, discussions of outlet bias in the CPI have focused on the differences in prices between incoming and outgoing outlets at the time of sample rotation. The Conference Board's Study Group on the CPI, for example, recommended that⁸:

“When outlet rotation shows price changes on the same items between the old and new sales outlets, the BLS, instead of (as now) assuming that all of it represents differences in the quality and convenience of the transactions, should estimate what portion of the price change represents a difference in quality and convenience vs. what portion represents a “true” change in price.”

⁷ U.S. Senate (1996), p. 43.

⁸ Conference Board (1999), p. 23.

Our primary goal in this paper is to determine the potential quantitative impact of changing the current BLS approach.⁹ For that purpose we examined detailed CPI microdata on the 69 months from January 2002 through September 2007. Our analysis was made possible by the BLS development of a CPI Research Data Base providing detailed information on the items priced in the index since 1987.¹⁰ Previous studies have been limited by the difficulty of assembling large files of incoming and outgoing items along with their quality characteristics.

Following Reinsdorf (1993) and Hausman and Leibtag (2004), we selected a number of relatively homogeneous food categories in order to limit, as much as possible, the influence of differences across outlets in the characteristics of items being sold. These 14 categories are shown in Table 1. With the exception of fruit and vegetable juices, our list roughly corresponds to item categories that were studied by both Reinsdorf (1993) and Hausman and Leibtag (2004). Together, the CPI item strata in which these categories fall comprised approximately one-quarter of the weight of the Food at Home in the CPI in December 2006, although in the interest of reducing heterogeneity we have further limited some of the samples by including, for example, only yellow bananas within the Bananas item stratum. Even within these limited categories, our study differs from others by explicitly adjusting for the varying quality of goods sold by different outlet types. A large grocery store might sell name-brand yellow bananas, while a discount department store might sell unbranded bananas.

Each of the categories in Table 1 is drawn from a different “Entry-Level Item” or ELI, the ultimate sampling unit for items as defined by the BLS national office. ELIs comprise the level of item definition from which data collectors begin item sampling within each sample outlet. Some of our 14 items, such as apples, constitute an entire ELI, while others comprise only a subset. National cola brands, for example, fall within the Carbonated Drinks ELI.

As is true for the great majority of CPI items, the TPOPS rotation process brings in new outlet samples for these categories on a semi-annual basis, during four months of the year. The outlets chosen for pricing in each of the 87 areas in the CPI geographic sample (primary sampling units or PSUs) are selected from frames generated using spending patterns reported in the household TPOPS survey, which is conducted for BLS by the Census Bureau. Within each CPI item category, the outlet sample is replaced in one-eighth of the areas during each semi-annual rotation; thus, the entire sample is replaced every four years.¹¹ For example, in the bimonthly even metropolitan area of San Francisco-Oakland-San Jose, the ELI sample for soda was rotated in April 2004, coffee in April 2005, eggs and apples in October 2005, and bread in April 2006. By contrast, in Philadelphia-Wilmington-Atlantic City, coffee and eggs were rotated in April 2004,

⁹ As noted in Section II, the structure of the TPOPS probability-sampling process, which selects first outlets and then items, complicates the design of operational solutions to the outlet bias problem. For a discussion of some potential approaches, see Walter Lane (2001).

¹⁰ See Dennis Fixler and Robert McClelland (2000), p. 6.

¹¹ The major exception to this process is rental housing, which is not subject to regular rotation. A few other “Non-POPS” items are rotated using other means. These items include, for example, postage and state vehicle registration. A more detailed discussion of pricing and sample rotation is given in Bureau of Labor Statistics (2007), pp. 13-17.

apples and bread in October 2004, and soda in October 2006. This balanced schedule smoothes the workload for CPI data collectors and, for our purposes, it yields a roughly constant number of incoming and outgoing item prices over our sample years.

CPI PSUs are classified as either monthly or bimonthly according to the frequency of CPI price collection. New York, Chicago, and Los Angeles are monthly areas, indicating that BLS collects prices for virtually all item categories each month. In other areas, collection of most prices takes place only in odd or even months. The BLS, however, prices food at home, energy, and selected other items on a monthly basis in all areas.

For our empirical analysis we constructed a sample of all item prices—what the BLS calls “quotes”—for each month from January 2002 through September 2007. For the 14 item categories above, this yields a total sample of about 350,000 price quotes in approximately 8,000 outlets. Note that the same individual item in a given store will be observed in multiple months until it rotates out of the sample or when it disappears from the shelves and the BLS data collector must substitute a similar item. CPI terminology refers to the substituted item as a new “version.” When such a substitution occurs in a store, the CPI analyst in Washington decides 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 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).¹² For the purposes of this paper 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. Our dataset for the 14 item categories contains about 16,000 item strings.

The CPI Research Database enables us to identify for each priced item the “business type” of store in which it is sold. Sample outlets are coded into hundreds of categories. Most of these categories—pet stores, banks, etc.—are not relevant for the items we study in this paper, but our data still provide great detail on store type. Roasted coffee, for example, is represented in our CPI sample primarily by three business types: Large Grocery Stores, Discount Department Stores (the supercenter category in which Wal-Mart would appear), and Warehouse Clubs and Other Membership Retail Outlets (which would include Sam’s Club or Costco). Among the other store types represented are small grocery stores, chain drug stores, limited-service food service establishments (into which a Starbucks offering snacks would logically be classified), and miscellaneous food at home stores (such as a store selling only coffee), along with catalog and internet outlets. This detail enables us to obtain a clearer understanding of the impact of outlet type trends on the CPI than would be possible with a simple classification of outlets into, for example, traditional and non-traditional stores.

¹² Examples of the characteristics recorded by the CPI and used in comparability decisions are given in Section V below. Exceptions to the substitution-handling process described here, such as the use of hedonic regression for quality adjustment, are very rare in the CPI food categories.

Figures 1 through 3 provide information on the distribution of outlet types in our sample and on the trends in the mix between 2002 and 2007. For several of the analyses in this paper we group outlets into six categories: Large Grocery Stores; Discount Department Stores; Warehouse Clubs and Other Membership Retail Outlets; Small Grocery Stores; Convenience Stores; and Other Outlet Types. The second and third of these categories comprise the SMC (supercenters, mass merchandisers, and club stores) group on which Hausman and Leibtag focus in their recent papers. In Figure 1 we show the percentages of our total item sample for the five outlet categories other than Large Grocery Stores in each of our 69 sample months. Note that these are unweighted counts. For CPI index calculation, individual item prices will have different weights depending on their item stratum, their geographic area, and the specific way in which the probability sampling process was designed and carried out for that outlet and ELI. For our purposes, however, the use of unweighted counts is both more convenient and more useful.

As Figure 1 demonstrates, the aggregate market share in our CPI food samples of the five outlet categories shown has been growing steadily, from about 16 percent in January 2002 to about 25 percent in September 2007, implying a corresponding reduction in the residual Large Grocery Stores share from 84 percent to 75 percent. The two SMC categories have exhibited the most striking growth. Discount Department Stores increased from 3.6 percent to 9.2 percent, and Warehouse Club stores from 3.1 percent to 6.4 percent. Among the three remaining categories, increases in the small shares of Convenience Stores and Small Grocery Stores approximately offset a decline for Other Outlet Types.

The aggregate 15.6 percent share of the two SMC categories in the last month of our data approaches, but is somewhat lower than, the share reported by the federal government's Economic Research Service (ERS) for all sales of food at home. According to ERS, warehouse clubs, supercenters, and mass merchandisers accounted for 20.0 percent of food at home sales in 2007. The unavoidable lags in the TPOPS rotation process account for some of this difference.

Figures 2 and 3 demonstrate that the growth in the share of these lower-price outlets in our sample has not been limited to any one item category. In Figure 2 we compare the percentages of quotes priced in discount department stores for the first and last calendar years of our study period, and show that those percentages increased sharply in all of our 14 categories except one, juice. Figure 3 makes the same comparison for warehouse club stores. Between 2002 and 2007 the share of quotes from that outlet group increased in 11 of the 14 item categories; the exceptions were potatoes, lettuce (where the warehouse share was very small in 2002 and zero in 2007) and juice.¹³

The previous empirical studies of outlet bias have relied primarily on examinations of average-price differences or on comparisons of published indexes to simulated series such as unit value indexes. A key goal of this paper, however, is to adjust not only for outlet-type differences but also for the effects of numerous item-specific characteristics.

¹³ For three item categories—ground beef, milk, and juice—our study period began shortly after January 2002 due to changes in their CPI checklist formats. For milk and juice, the first year represented in Figures 2 and 3 is 2003, rather than 2002.

Therefore, the analysis we report in the next section is based on a multiple regression approach.

V. Regression Results

For each of our 14 item categories, we estimate a semi-logarithmic regression model in which price is a function of both item characteristics and outlet effects, as well as time period. The individual observations in our data comprise the price, characteristics, and outlet codes of a given product version in a given month. As discussed above, a version string is observed in each month of our study period from the time it appears until it is rotated out of the CPI sample or its place is taken by a new version through forced item substitution. In any given month there may be more than one version string being observed within the same sample outlet. Because of the rolling pattern of CPI rotation, outlets and their associated version strings are entering and leaving our samples throughout the period of study.

Indicating a version string by the subscript s and its outlet by j , and using t to indicate month (the values of t run from 0 in January 2002 to 68 in September 2007), our model is

$$\ln P_{st} = \beta_0 + \pi_t^h + X_s \beta + \gamma_j + \varepsilon_{st} \quad (5)$$

In (5) β_0 is an intercept term and the disturbance terms ε_{st} are assumed to be independently and identically distributed for all s and t . Item characteristics are represented by a set of dummy variables X_s specific to the version string, and β is the associated vector of coefficients.¹⁴ The term γ_j is an outlet fixed effect for the j -th outlet, where j is, again, the outlet in which version s is sold. The values of π_t^h are of particular interest because they comprise the price index implied by (5), with item characteristics and outlets held fixed. That is, for t running from 1 to 68, the values of π_t^h are the estimated logarithms of the price level in period t relative to period 0 for a given set of item characteristics and in a specific outlet. We attach the superscript h , for hedonic, to distinguish this price index from others we will present later in this section.

Although the item categories studied here are relatively homogeneous, an important purpose of this paper is to determine whether some of the variation in prices across outlets arises from variation in the characteristics of items sold at those outlets. For example, part of the difference in observed prices between outlet types might be explained by discount department stores selling items with different characteristics from those sold at large grocery stores. The item characteristics that we have available in the CPI data are included in the variables X_s for each of our 14 item categories. Each ELI has one or more checklists that allow BLS employees to locate and price the same item in successive periods. We display the checklist for tomatoes in the appendix as an example. The checklists include categories for most relevant characteristics, and several additional

¹⁴ In estimating this model we ignore any differential correlations among the disturbance terms within and across item strings in a given outlet. Note also that the use of outlet fixed effects precludes modeling locational effects, such as with dummy variables for metropolitan area location. Given that the CPI geographic sample did not change over our sample period, however, we are not concerned that the intertemporal role of outlet changes will be distorted.

write-in categories. For tomatoes, this includes such information as the variety of tomato (cherry, plum, etc.), whether the tomatoes are organic, whether they are greenhouse-grown, and whether they are loose or packaged. Here we employ dummy variables for virtually all non-write-in checklist categories for each ELI. If appropriate, we also include dummy variables coded from some of the write-in fields, such as H99. For example, the regression model for ham includes dummy variables for characteristics such as spiral cut and honey baked.

It is also important to note that the dependent variable in our regressions, the logarithm of price, is measured including sales tax and is also measured on a per-unit basis, consistent with the general CPI practice for food items. For example, the dependent variable in the cola and butter regressions is the logarithm of price per ounce. Thus, variations in container or package size across items and outlets will not affect the price variable directly. Indirectly, however, the per-unit price may well vary with package size if markets are characterized by volume discounts. The use of package size as a characteristic is particularly important because larger sizes clearly are less attractive to consumers, *ceteris paribus*, and because average package size is related to outlet type. As we discuss later in this paper, warehouse club stores generally offer both the lowest per-unit prices and the largest package sizes.

To capture the effect of such discounts, we define a variable, Size, to measure the amount, usually measured in units of volume or weight, corresponding to the particular item offered for sale. Size is measured as the product of the “derived quantity” and “derived size” variables coded in the CPI database. A six-pack of 12-ounce cans of cola selling for \$2.99 would have a derived quantity of 6 and a derived size of 12, and the per-unit CPI price would be approximately four cents. In this case our Size variable would be 72.

Although Size is ordinarily constant within a version string, it can change in some instances. One example is two-for-one sales, which appear in the data as a temporary doubling of derived quantity and a 50-percent drop in per-unit price. To prevent temporary sales from distorting the relationship between price and package size we defined the variable LNSIZE as the version-string mean value of the logarithm of Size, thereby making LNSIZE constant within a version string, and each of our 14 item regressions included LNSIZE and its square as explanatory variables. In addition, we include dummy variables for container size in X_s whenever it is included in the ELI checklist specification, as in the cases of cola and juice.

As noted above, we estimate the model in (5) separately for each of our 14 item categories. Consider the category of butter as one example. For butter, our sample contains 12,347 observations, corresponding to 503 version strings (465 item strings) in 376 outlets observed over an average of 25 months per version string (27 months per item string). Thus, there are 376 outlet fixed effects γ_j in the butter regression, along with 68 dummy variables for months. The X_s matrix includes, for example, a set of dummy variables representing whether the product is whipped as opposed to regular creamery butter, whether it is Grade AA or some other quality, and the weight of the item (8 ounces, 16 ounces, etc.). We also included two dummy variables for well-known national brands of butter.

We estimated this fixed effects model without weighting the observations. The resulting price index terms π_t^h indicate wide variation within our sample period but little overall price change. The value for April 2003 is -.223, implying an approximately 20 percent price decrease from January 2002. The index then rises to a value of .185 in June 2004, or about 20 percent above the January 2002 level. By September 2007, however, the log-index has returned to .029, or about a three percent overall increase.

It is instructive to compare these index movements to those of an index derived using methods closer to those used in the CPI. For that we estimate

$$\ln P_{st} = \beta_0 + \pi_t^m + \delta_q + \varepsilon_{st} \quad (6)$$

where δ_q is a dummy variable for the item string of which version string s is a component. Recall that the CPI treats prices within an item string as comparable, but essentially treats any difference between the prices of different item strings as due to product quality differences. Similarly, the price index terms π_t^m effectively hold the item string fixed, so that inflation is only estimated within each item string. Consequently, we refer to π_t^m as a matched model index, with a model corresponding to an item string.¹⁵

The distinction between our matched model and hedonic indexes is in the treatment of changes in item characteristics. When one item is substituted for another, the matched model index attributes the price difference to quality if there is a change in item string, and to “pure price change” if there is no change in item string—that is, if the CPI considered the new and old versions comparable. In calculation of the hedonic index π_t^h , the price difference is decomposed into quality and “pure price” differences based on the regression coefficients associated with the item characteristics that differ between the two versions. For butter, the price index in (6) closely follows π_t^h : in April 2003 π_t^m equals -.225, and it rises to .180 in June 2004 before falling to .017 in September 2007.

We emphasize that our indexes π_t^h and π_t^m differ from the actual CPI for many reasons. 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. Nevertheless, the CPI index movements for butter are quite similar to those of our regression-based indexes. The CPI for butter fell 21 percent between January 2002 and April 2003, then rose by June 2004 to a level 19 percent higher than that in January 2002. For our study period as a whole, the CPI rose only 0.1 percent.¹⁶

We next turn to a consideration of our estimated outlet effects for butter. Because of the changing outlet mix in our CPI sample, we do not have a balanced sample: not all outlets

¹⁵ Since an item string can comprise versions with slightly different characteristics, one could alternatively define a matched model index as one in which the model corresponds to a version string. We use our formulation because it corresponds to the approach used in the CPI.

¹⁶ There are no published CPI series at the detailed level of several of our item categories, and we make no attempt in this paper to compare the CPI to any of our estimated price indexes except that for butter.

appear in any given period. The γ_j terms do not have to average to zero in each period. Rather, they are all measured relative to one arbitrarily chosen outlet whose γ_j is set to zero. In Figure 4 we display the outlet effects in each period for 500 randomly chosen outlet-period combinations in our sample. (There are 11,798 such pairs in our butter sample, which if included would make the chart unreadable.) Also included is the simple linear trend estimated using those 500 data points. The figure shows the wide variation in fixed effects across outlets. It also shows a slight downward trend, which suggests that the outlets in the sample at the end of the study period tended to offer lower prices for butter, *ceteris paribus*.

We can ask two questions about these outlet fixed effects, testing whether the trend displayed in Figure 3 is indicative of an actual market phenomenon. The first question is whether the γ_j are significantly different from zero. Using a standard F test, the null hypothesis that all the outlet effects are zero is rejected with a very high confidence level. We therefore conclude that some outlets in our sample charge higher prices than others for butter products with the same measured characteristics. We will henceforth use the term “premium,” which may be either positive or negative, to denote the fixed effect for a given outlet.

The second question we can ask about the fixed effects is whether there is a correlation between the outlet premiums and time. That is, does the average premium value change according to some linear, or non-linear, or even non-monotonic function of time? We address this question in two ways. First, we simply regress the average monthly average outlet premium against a time trend, as pictured in Figure 4. For butter, the coefficient of time is negative and highly significant, with a t-statistic of -6.72. Second, we apply a more general Hausman-type specification test. We can estimate equation (6) using a random effects rather than fixed effects specification of the γ_j terms. If the random effects model is valid for our data, it offers a more efficient estimator of the outlet effects. However, the maintained assumption of the random effects model is that the γ_j are uncorrelated with the other explanatory variables in the equation, including the dummies for time periods. Therefore, we apply a Hausman-type test of the null hypothesis that the outlet effects are independent of time. For butter, this null hypothesis can be rejected at the .0086 significance level.¹⁷ Using either test, therefore, we conclude that the average level of the outlet premium for butter is not constant over time.

The CPI methodology implicitly assumes that differences in outlet premiums reflect differences in outlet quality as viewed by consumers, and that the price index should therefore be computed conditional on an average outlet premium level, as is done in the computation of π_t^h . If consumers are indifferent between outlets except for price, however, a change in the average outlet premium should be treated as a change in the

¹⁷ To apply the test, we first estimated a random effects model of equation (5). This yields estimates of the component within- and between-outlet variances. Given these estimates, the random effects model can be equivalently estimated by subtracting from each regression variable its outlet-mean value multiplied by a term θ which is a function of the two variances and the sample size of the outlet. Our specification test was then an F test of whether, conditional on these θ -adjusted variables, we could accept the null hypothesis that the unadjusted differences of the time dummy variables from their outlet means significantly added to the explanatory value of the regression. Under the random effects model they should not.

price index. To represent the latter view, we define a third log-price index by setting π_t^o in time t to equal π_t^h plus the average outlet premium in the sample of price quotes in time t .

$$\pi_t^o = \pi_t^h + (\overline{\gamma_t} - \overline{\gamma_0}) \quad (7)$$

If the average premium is declining over time, as suggested by the trend in Figure 4, the index series π^o will show a slower rate of inflation than π_t^h . This is what we observe for butter. As noted earlier, the value of π_t^h in September 2007 is .029, or about a three percent increase after five years and eight months. By contrast, the ending value of π_t^o is .010, indicating only a one percent increase.

Next, return to equation (5) and note that with dummy variables included in the regression for each month in our sample except the first, the average residual will be zero for each month.¹⁸ Then, letting $\ln \hat{P}_t$ be the average predicted value in period t , we can write

$$\begin{aligned} \ln \hat{P}_t - \ln \hat{P}_0 &= (\overline{X_t} - \overline{X_0})\beta + \pi_t^h + (\overline{\gamma_t} - \overline{\gamma_0}) \\ &= (\overline{X_t} - \overline{X_0})\beta + \pi_t^o \end{aligned} \quad (8)$$

We now define a fourth price index π_t^u obtained by regressing $\ln P_{st}$ on a set of time dummy variables alone, that is, equation (5) without the item characteristic variables or outlet fixed effects. The estimated values of π_t^u will be the differences in mean log prices in each period relative to period 0, and we have seen that these will equal the differences in mean predicted values on the left hand side of (8). The series π^u would be an appropriate measure of price change under the assumption that consumers are indifferent between different outlets and that they also view the item category as perfectly homogeneous; they are indifferent among all the items represented.

The differences among π^u , π_t^h , and π^o , therefore, measure two components of change over time: an item characteristics component and an outlet premium component.

Rewriting equation (7) gives the outlet premium component between periods 0 and t as

$$\pi_t^o - \pi_t^h = (\overline{\gamma_t} - \overline{\gamma_0}) \quad (9)$$

and the item characteristics component is given by

$$\pi_t^u - \pi_t^o = (\overline{X_t} - \overline{X_0})\beta \quad (10)$$

Our alternative index results for butter and the other 13 item categories are shown in Table 2. The numbers in the table are estimated log-changes in price over the 69 months in our study period. The columns labeled U, H, O, and M contain the ending log-index

¹⁸ If this were not true for some time period t the sum of squared residuals could be reduced by a change in the estimated π_t^h .

levels for π'' , π_t^h , π^o , and π^m , respectively, and columns B through E display the index differences. As previously noted, for example, our hedonic index for butter ends the study period at a value of .029, and after adjustment by the average outlet effects we obtain an ending level of .010. For butter, the item characteristics component is in the same direction as the outlet component but is about three times as large (.055 compared to .018). The unadjusted index π'' , which does not hold the mix of item characteristics constant, ends at a level of -.045. In the case of butter, the most important changes in item characteristics implied a decreasing value to consumers: a decrease over time in the sample share of the top national brands, and the increased frequency of unusual weight values, such as three- or four-pound sizes. Adjusting for these changes raises the hedonic estimate of price change. Column E shows that the matched model index ends at a level 0.012 below the hedonic index. This implies that the matched model approach estimates about 1.2 percentage points more in quality change over the period than does the hedonic index.

The other rows of Table 2 highlight the widely varying rates of index change across our item categories, and the variation in the item characteristics component of index differences. The latter range from an approximate 7.8 percent decrease in item “quality”¹⁹ for cola to more than an 7.5 percent increase for ham.²⁰ The item characteristics effect is positive, indicating an increase in “quality,” in ten of 14 categories. Meanwhile, the estimates of the outlet effect cover a much narrower range and are negative in all but four item categories: apples, eggs, coffee and lettuce.

In the last row of the table we display weighted averages of the category results. The BLS does not construct consumer expenditure weights for all ELIs, and many of our categories comprise only part of an ELI; iceberg lettuce, for example, is only a small part of the Lettuce ELI. For this paper we used sample quote counts along with 2003-2004 CPI item stratum weights to yield rough estimates of the weights of our item categories in the CPI. Aggregating the category results using these weight estimates, we find that the average outlet component is -.0154. This implies that the effect of changing outlet mixes has had a negative effect on the price level for the 14 categories together of approximately 0.27 percent per year.²¹ The negative sign is consistent with the hypothesis that discount department stores, warehouse club stores, and/or other new outlet types are offering lower prices to consumers, even after the characteristics of items sold in the stores are taken into account. In the next section of the paper we analyze the contributions of different outlet categories to this overall effect.

Perhaps surprisingly, the table also shows that the aggregate estimate of the item characteristic component is the same order of magnitude as the outlet component, but opposite in sign. Consequently, our aggregate movements in the hedonic and unadjusted

¹⁹ As discussed in Section VII, we recognize that some aspects of what consumers view as item quality cannot be observed in our data. Also, some of the valuation of item characteristics, such as brand, may reflect market structure rather than quality per se.

²⁰ In the remainder of our discussion we treat “percent changes” and “log-changes” as synonyms, for reasons of expositional convenience.

²¹ As noted in footnote 13, for three item categories—ground beef, milk, and juice—our study period began after January 2002. All the estimates presented for those three categories were adjusted by imputing their price changes back to January 2002 from the indexes for the other categories..

price indexes are relatively close. Overall, our results imply an improvement in the quality of items sold of about 0.19 percent per year. As might be expected, large effects are in coffee and juice, which have numerous measured characteristics such as the flavor of juice or whether the coffee is decaffeinated. Relatively large effects are also observed, however, in such apparently homogeneous categories as butter and fresh whole milk. One explanation is that package size is very important as an explanatory variable even in cases in which the individual product may seem very standardized aside from packaging. These results highlight the importance of examining differences in item characteristics across stores.

Equally surprising is the relationship between the matched model index π^m and the other indexes. In principle, the matched model index should behave similarly to the hedonic index. Both do not include price changes occurring from outlet changes, and if analyst decisions about the comparability of new and old item versions are consistent with quality differences as reflected in the hedonic regressions then the matched model and hedonic indexes should approximate each other. In 12 of the 14 item categories, however, the matched model index indicates less inflation than the hedonic index. The implication is that either (i) the matched model approach is systematically mis-estimating the market value of quality differences, or (ii) there are unmeasured differences in product quality that are reflected in the matched model index but not taken into account by the hedonic regressions. Coincidentally, the differences shown in Column E of Table 2 lead the matched model index to be quite close to π^o , the quality adjusted price index that does not hold constant price changes due to outlet substitutions.

We close this section by noting that although our focus is on consumer price indexes, our results also have implications for product price regressions in general. Just as our time dummy coefficients are affected by the inclusion or exclusion of outlet fixed effects, a failure to include outlet-type variables can potentially distort other coefficient estimates in price regressions, such as the coefficients on item quality variables in hedonic regression models.

VI. Outlet Categories and Decomposition of Outlet Fixed Effects

We have shown in Section IV that the shares of discount store types have been growing rapidly in CPI food data. We have also shown in Section V, using our fixed effects regressions, that the changing mix of outlets has had an overall negative impact on the price level in the categories we study. In this section we analyze the extent to which this negative impact has been due to the changing sample shares of discount department stores, warehouse club stores, and other outlet types.

We begin with simple comparisons of price levels in the six outlet categories from Section IV—Large Grocery Stores, Discount Department Stores, Warehouse Clubs and Other Membership Retail Outlets, Small Grocery Stores, Convenience Stores, and Other Outlet Types—without any adjustment for item quality. We estimated 14 regressions, one for each item category, with the logarithm of price, $\ln P_{st}$, as the dependent variable, and dummy variables for outlet category and time period as independent variables. The results are shown in Table 3 and Figure 5.

The table demonstrates a clear pattern of lower prices in the two SMC categories compared to prices in large grocery stores. Discount department store prices are lower in every category except milk, which is also the only category where the difference is not statistically significant. The weighted average difference across all the item categories is 12.3 percent. The differences are frequently much wider, although less consistent in direction, between large grocery stores and warehouse club stores. The weighted average discount observed at warehouse clubs (23.4 percent) is almost as large as the largest found for any item category at discount department stores (24.3 percent for butter). In seven categories the warehouse club discount exceeds 30 percent.²² Figure 5 displays these price differentials graphically, highlighting that in 10 of the 14 item categories the CPI sample prices in warehouse clubs are lower than prices in either large grocery stores or discount department stores.²³

For the other outlet types--small grocery stores, convenience stores, and other—Table 3 shows many large and statistically significant differences from large grocery prices, both in a positive and negative direction. Convenience stores are the only outlet type in which we find higher prices than in large grocery stores for more than half of the item categories.

Although Table 3 demonstrates that discount department stores and warehouse club stores have distinctly different patterns of prices, the table does not adjust the prices for differences in the characteristics of the items sold. In Table 4, we present price level comparisons derived from regressions similar to those underlying Table 3 but with the addition of the item quality variables we used in the models discussed in Section V above.²⁴ Comparison of Table 4 and Table 3 shows that, as expected, adjusting for item characteristics tends to reduce the estimated price differences across store categories. Averaged across all item categories, the price differentials relative to large grocery stores in Table 3 ranged from a positive 13.1 percent for convenience stores to the negative 23.4 percent for warehouse club stores. By contrast, in Table 4 the differentials are all negative and range only from 1.8 percent for convenience stores to 10.0 percent for discount department stores.²⁵

The most dramatic differences between Table 3 and Table 4 are in the category of warehouse club stores. Their estimated overall price advantage relative to large grocery stores is reduced by almost two thirds, and in five of 14 item categories a negative premium relative to large grocery stores becomes a positive premium. Only the estimate

²² As elsewhere, this is the average logarithmic differential.

²³ Hausman and Leibtag make a similar comparison of price levels across outlet types in their 2004 paper, showing the ratio of average prices in supermarkets relative to SMCs in each of their 20 item categories for their 1998-2001 period of study. On average, they found supermarket prices to be 30.0 percent higher, which is a larger average differential than we observe in our CPI data. Their pattern of differentials across items also differs from ours; for example, they found a relatively small differential for butter/margarine, whereas butter shows the largest differentials in our data.

²⁴ Another way to present quality-adjusted prices by outlet category would have been to compute category averages of the fixed effects from the Section V regressions. The results would have differed only slightly from those in Table 4 due to weighting, and would have been less directly comparable to the Table 3 results.

²⁵ Some estimated differentials for individual item categories remain wide even in Table 4, often reflecting small sample sizes; the CPI lettuce sample contains very few convenience stores, for example.

for ground beef is relatively unaffected. Given the nature of warehouse club stores, it is natural to expect that product size is the characteristic responsible for these sharp changes.

We explore this size relationship in Table 5, which compares the product size distributions between warehouse club and all other stores. The first row of the table shows that the median size for a priced apple item in the CPI in warehouse club stores in our sample was 96 ounces (six pounds) compared to one pound for all other stores. There was essentially no overlap in the distributions; no amount of apples priced in a sample warehouse club was smaller than five pounds, and six pounds was at the 99th percentile of the size distribution in the rest of the CPI sample. Similarly distinctive size distributions are found for warehouse club stores in most of the other item categories. (Three of our 14 item categories have the same median sizes in both outlet samples and are excluded from Table 5 and Figure 6. These are eggs, which are almost always sold in one-dozen amounts, and ground beef and ham, which are generally offered for sale on a per-pound basis and for which the exact size is usually not recorded in the CPI data.)

In Figure 6 we compare the quality adjusted prices in warehouse club stores with prices at those stores that are adjusted for all characteristics *except* size. The figure confirms that adjusting for the negative relationship between size and price in our regressions largely offsets the apparent price advantage of warehouse club stores.²⁶

Trading low prices for very large sizes is clearly a fundamental feature of warehouse club stores. Product size is an item characteristic that can, therefore, also be thought of as a characteristic of the warehouse club outlet category. Because little or no data are available on other outlet-level tradeoffs for low prices, our results present a unique opportunity to evaluate the importance of an outlet characteristic. Viewed in this way, our results suggest that a substantial portion of the difference in prices between warehouse club stores and large grocery stores is due to a measurable quality of the outlets.

This discussion also highlights the value of examining warehouse club stores separately from the rest of the SMC category. Although there is an important and growing economic literature on the impact of Wal-Mart,²⁷ discount department stores like Wal-Mart represent only one aspect of the impact of new outlets on the overall level of consumer prices. Aggregating unadjusted prices of discount department stores with the prices of similar items at warehouse club outlets can severely exaggerate the impact of discount department stores. Aggregating quality adjusted prices does not appear to suffer from that problem to the same degree, although even in Table 4 there are still many large discrepancies between the prices of discount department stores and the prices of warehouse club stores.

²⁶ It might be thought that the strong relationship between size and outlet type would make it impossible to separately identify the impact of size and the outlet fixed effects for warehouse club stores. However, this was not the case; warehouse club stores account for only a small percentage of our outlets, and there was sufficient variation in size across the entire sample to make LNSIZE and its square statistically significant in every regression underlying our hedonic indexes in Table 2.

²⁷ Basker (2005) is a prominent example.

We now separate the effects of changing market shares of outlet categories from the changing market shares of outlets within outlet categories. To distinguish those two effects, we return to equation (4) of Section II and assume that each of the old and new sample outlets indexed by i and j falls into one of a set of outlet categories $k=1,\dots,S$. We define the share weight of category k as the sum of the weights of the outlets in that category, i.e., $W_k = \sum w_i$ and $W_k^n = \sum w_j^n$ for all outlets i and j in category k in the overlap time period 2. We also define \bar{P}_k and \bar{P}_k^n as the weighted average prices in outlet category k in period 2. The difference in outlet effects in (4) can be written as:

$$\sum_j w_j^n \ln(P_j^n) - \sum_i w_i \ln(P_i) = \sum_k W_k^n \ln(\bar{P}_k^n) - \sum_k W_k \ln(\bar{P}_k) \quad (11a)$$

Rearranging,

$$\sum_k W_k^n \ln(\bar{P}_k^n) - \sum_k W_k \ln(\bar{P}_k) = \sum_k \frac{\ln(\bar{P}_k^n) + \ln(\bar{P}_k)}{2} [W_k^n - W_k] + \sum_k \frac{W_k^n + W_k}{2} [\ln(\bar{P}_k^n) - \ln(\bar{P}_k)] \quad (11b)$$

The first sum on the right-hand side measures the difference in prices due to the difference in expenditure shares of each category. That difference would be due to, for example, a greater expenditure share of discount department stores in the new sample. The second sum measures the difference in average prices within each category, including changes due to shifting consumption patterns. For example, changing shopping patterns within the category of large grocery stores would result in a change in average prices for the category.

The two equations above can be estimated using the outlet fixed effect coefficients from our regression analysis in Section V. By dividing outlets into categories, we obtain an equation for the difference in average outlet premiums at the beginning and end of our sample period:

$$\bar{\gamma}_T - \bar{\gamma}_0 \equiv \sum_k w_{kT} \bar{\gamma}_{kT} - \sum_k w_{k0} \bar{\gamma}_{k0} \quad (12a)$$

With outlet premiums substituted for prices, equation (12a) corresponds to equation (11a), where the categories k are our six store types. Periods 0 and T correspond to January 2002 and September 2007, respectively, and the terms $\bar{\gamma}_{kt}$ in (12a) are the mean outlet premiums for quotes in outlet category k in those two periods.²⁸ The terms w are the shares of quotes represented by each outlet category.

The use of outlet fixed effects in (12a) differs from the use of average prices in (11a) in two respects. First, the outlet fixed effects are explicitly drawn from samples in two different time periods rather than being drawn from two contemporaneous incoming and outgoing samples. However, the outlet fixed effects are estimated in an equation that includes dummy variables for the time periods, so the variation due to time has been purged from the outlet fixed effects. Second, differences in the average prices in the two

²⁸ Note that because the quotes and outlets in our sample change only slowly over time, we are not introducing volatility by defining the terms in (12a) for specific months rather than for longer time periods.

samples in (11a) did not solely reflect differences in the outlets, because of potential differences in item characteristics in the two samples. The outlet fixed effects do not suffer from this problem because they are derived from a regression that included dummy variables for item characteristics.

Again following our earlier equation (11b), equation (12a) can be rewritten as:

$$\bar{\gamma}_T - \bar{\gamma}_0 \equiv \sum_k (\bar{\gamma}_{kT} - \bar{\gamma}_{k0}) \bar{w}_{k.} + \sum_k (w_{kT} - w_{k0}) \bar{\gamma}_{k.} \quad (12b)$$

Finally, without changing the value of the right-hand side we can subtract from each term $\bar{\gamma}_{k.}$ the overall mean outlet premium in the two periods $\bar{\gamma}_{..}$, yielding

$$\bar{\gamma}_T - \bar{\gamma}_0 \equiv \sum_k (\bar{\gamma}_{kT} - \bar{\gamma}_{k0}) \bar{w}_{k.} + \sum_k (w_{kT} - w_{k0}) (\bar{\gamma}_{k.} - \bar{\gamma}_{..}) \quad (12c)$$

The first summation is a set of within-category effects, weighted by the average of the period 0 and period T category weights. Each term in parentheses in this summation is the change over time in the average outlet premium for the category.²⁹ For example, the mean outlet premium for the Large Grocery category could change because the mix of upscale and low-price grocery stores changed.³⁰

The second summation in (12c) gives the between-category effects, the effects of the changing sample shares of the categories. Here these effects are weighted by the category's average outlet premium relative to the overall mean $\bar{\gamma}_{..}$. Calculated in this way, an outlet category with an increasing sample share will add to the overall outlet component on the left hand side of (12c) if that category has a relatively high average outlet premium. If the category has a relatively low average outlet premium, an increase in its share will lower $\bar{\gamma}_T$ relative to $\bar{\gamma}_0$.

The results of our decomposition are shown in Table 6. The results for individual item categories were aggregated using estimated CPI weights as in Section V, and the total of -.0154 in Table 6 is the same value given in the last row of column C of Table 2. We can see that most of the total outlet component is explained by the between-category effects, the changes in category shares over time. A total of -.0097, or about 0.17 percent per year, can be attributed to the growth in the shares of the two SMC categories at the expense of large grocery stores, along with smaller effects from changes in the shares of the other three outlet groups. The growth in the discount department store category directly accounts for more than three-fifths of this total. These estimates of the impact of the growth in low-cost stores are qualitatively consistent with previous evidence and conventional wisdom. When we examine the item category data underlying Table 6, we find that the total between-outlet category effect is negative for 11 of our 14 items.

²⁹ As noted above, our regressions calculated outlet effects relative to an arbitrarily chosen outlet set to zero. For our decomposition exercise we adjusted the outlet effects so the mean over periods 0 and T would be zero in each category.

³⁰ It is important to distinguish our within-category effects from what Hausman and Leibtag called the "indirect effect" of SMC growth on the prices charged by regular supermarkets. If individual grocery stores lower prices in response to SMC entry, this will be picked up in our regression model by the price index terms π . Our within-category effect refers to the impact on the price level of outlet sample changes within a category.

In addition to the between-category results, Table 6 also indicates a relatively large total within-category effect, arising primarily from the contribution of decreasing relative prices within the mix of Other stores. This is a very heterogeneous group that differs widely from item category to item category, comprising such outlets as delicatessens, bakeries, and drug stores.

VII. Concluding Remarks

This paper confirms the potential importance of new outlets bias in the CPI. Using BLS-collected price data for 2002-2007, we observe a continuous increase in the market share of discount department stores and warehouse club stores. We also observe significantly lower prices at discount department stores and at warehouse club stores than at large grocery stores, both before and after adjusting for a large number of item characteristics. In 11 of 14 item categories examined, the increasing shares of lower-priced store categories reduced the average prices collected by the BLS. Changes in the distribution of outlets within categories also led to a substantial decline in average prices. Combined, changes in the distribution of outlets within and between categories lowered prices by about 1.5 percent over the 2002-2007 time period.

We also find a surprising degree of variation over time in the value of item characteristics. While the item categories might, at first glance, appear to contain relatively homogeneous goods, the average value of item characteristics has offset much of the decrease in average price due to the change in the distribution of outlets. Even items such as eggs and fresh whole milk have shown item quality increases of 3.8 percent and 3.7 percent, respectively.

The price variation accompanying that variation in characteristics leads to increases in the quality adjusted prices. The matched model index in this paper captures some, but not all of that increase. In 12 of the 14 item categories, the hedonic index increases faster than the matched model index.

The evidence described here by no means offers conclusive evidence of CPI bias. Most importantly, our analysis holds observable item characteristics constant, but does not address most outlet characteristics. Our results do take into account that the lower per-unit prices offered at warehouse club stores are partly due to larger package sizes, which are less valued by consumers. However, only by assuming that consumers are indifferent among stores on all dimensions, including locational convenience, service quality, and item selection variety, can our estimated outlet effects be taken at face value. Moreover, there may be differences in item characteristics across outlet types that are unobserved in the CPI data we employ. Some outlets may allow fruits and vegetables to lose freshness by remaining longer on the shelves, for example. Finally, we do not estimate and compare a model that reflects the precise current BLS procedures for calculating the CPI. Nevertheless, the fact that the market shares of discount stores are growing suggests that many consumers are benefiting from the lower prices at those stores.

We also know from our data that there are some countervailing trends, such as the increasing market share of outlet types that sell coffee at higher than average prices. Consumers shifting to those stores must attach some value either to the characteristics of

those outlet types or to unmeasured characteristics of the items sold there. Thus, our results suggest that outlet characteristics are not negligible factors.

Appendix

CPI Tomatoes Checklist

BUREAU OF LABOR STATISTICS

U.S. DEPARTMENT OF LABOR

CONSUMER PRICE INDEX - ELI CHECKLIST

| | | | |
|-----------------|-----------------|---------------|---------------|
| collection | outlet | quote | arranging |
| period: _ _ _ _ | number: _ _ _ _ | code: _ _ _ _ | code: _ _ _ _ |

| | |
|-----------------------------|-----------------|
| ELI No./ | cluster |
| title FL031 TOMATOES | code 01A |

| | | | |
|-----------------------|-------------|----------------|-------------------|
| item availability: | 1-AVAILABLE | 2-ELI NOT SOLD | 3-INIT INCOMPLETE |
| purpose of checklist: | 1-INIT | 2-INIT COMPL | 3-SPEC CORR |
| | 4-SUB | 5-REINIT | 6-CHECK REV |

CURRENT PERIOD

price: _ . _ _ _

type of price: REG SALE

quantity: _ _ _

size: _ . _ _ _ pair: YES NO

unit of size: _____

SALES TAX

included: YES NO

| | | | | | | | | | | | | | |
|------------|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| YEAR-ROUND | in-season: | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|------------|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

respondent: _____ location: _____

field message: _____

VARIETY

A1 Cherry Tomatoes

B1 Grape tomatoes

B98 Other (if specified),

A2 Round Red (Regular or Slicing)

Tomato Varieties

** B2 Variety of Round Red

Not Specified

** B99 Specified variety,

A3 Plum/Roma/Italian

A97 Other,

TYPE

C1 Field Grown/Vine Ripe

C2 Green House/Hot House

** D2 Hydroponic

** D98 Other (if specified),

C3 Not specified/Unable to determine

C99 Other,

ORGANIC CERTIFICATION

E1 Not USDA Certified organic

E2 USDA Certified organic

E3 Other Organic Claim

** PACKAGING

F1 Loose

F2 Packaged (Box, Tray, etc.)

** SIZE REPRESENTS

G1 Weight labeled

G2 One Package Weighed

(Qty. = the # of packages priced)

G3 Weighed 2 Tomatoes,

circled YES for PAIR

(Qty. = the # of tomatoes priced)

OTHER FEATURES

H99 _____

I99 _____

** OTHER ITEM IDENTIFIERS

J99 _____

K99 _____

L99 _____

Figure 1
Sample Shares by Outlet Type and Period

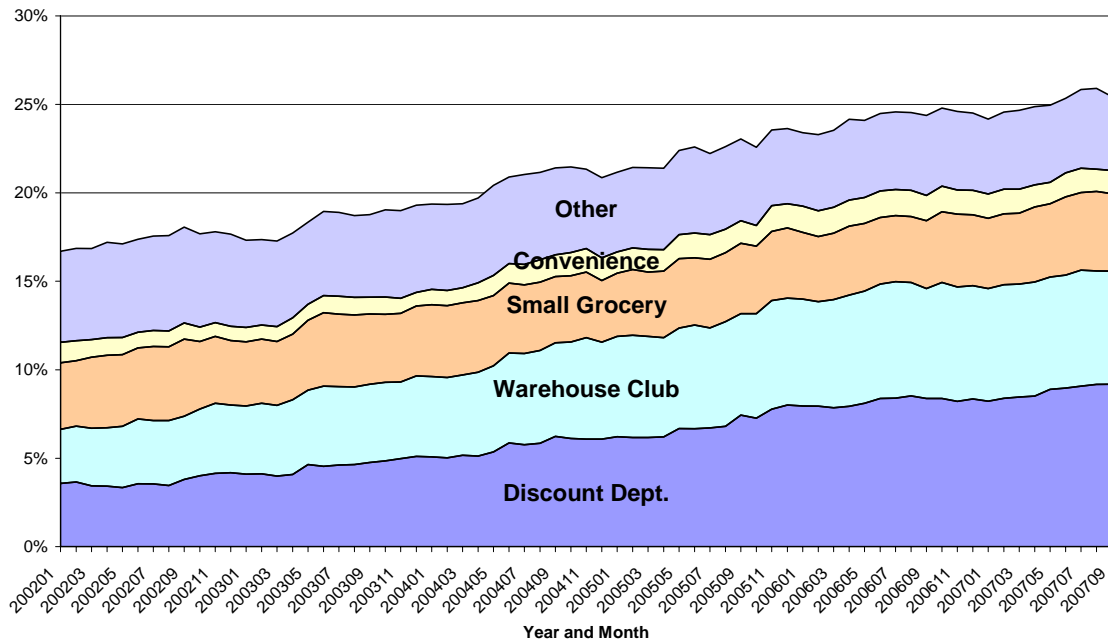


Figure 2
Discount Department Store Shares, by Item Category

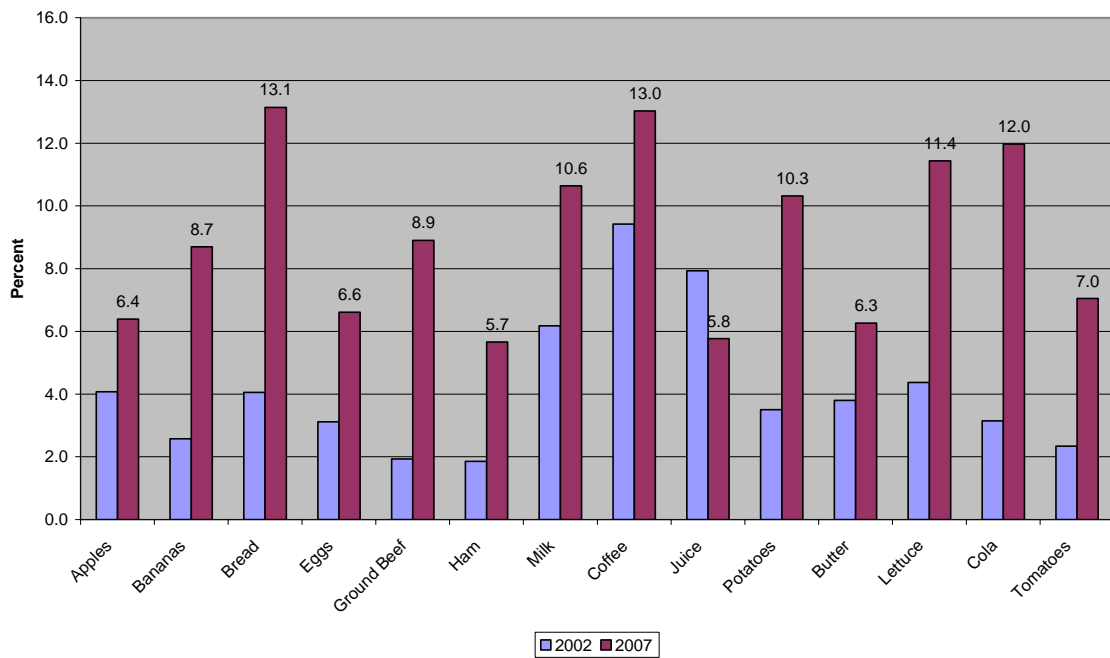


Figure 3
Warehouse Club Store Shares, by Item Category

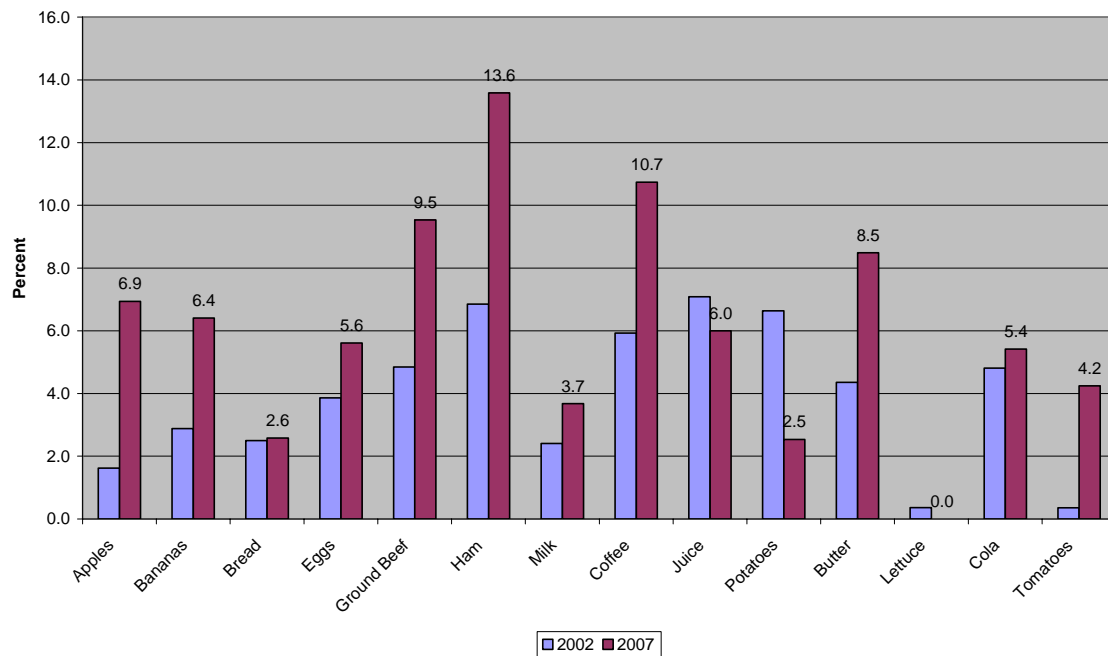


Figure 4
Sample of Outlet Effects by Month for Butter

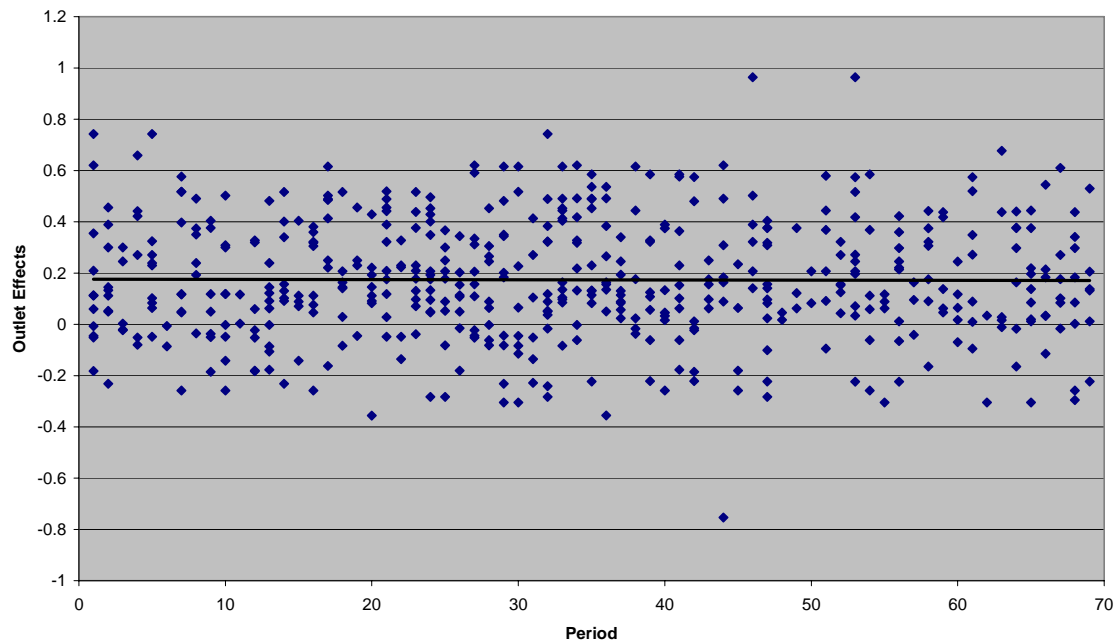


Figure 5
Unadjusted Price Differentials Relative to Large Grocery Stores

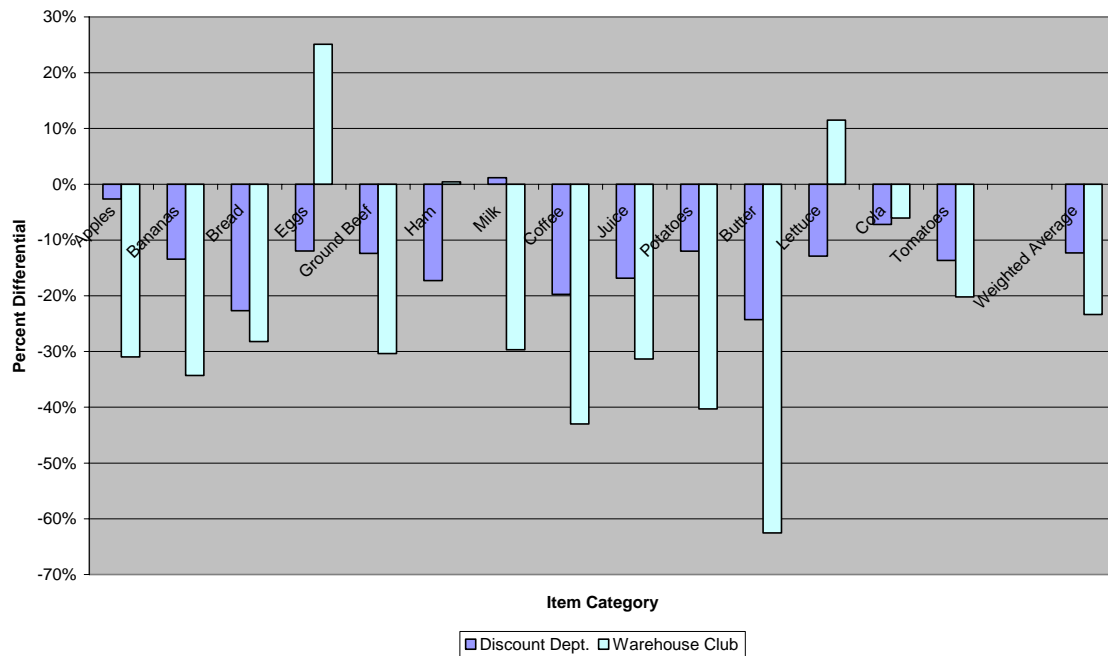


Figure 6
Warehouse Club Prices Relative to Large Grocery Stores
With and Without Adjusting for Product Size

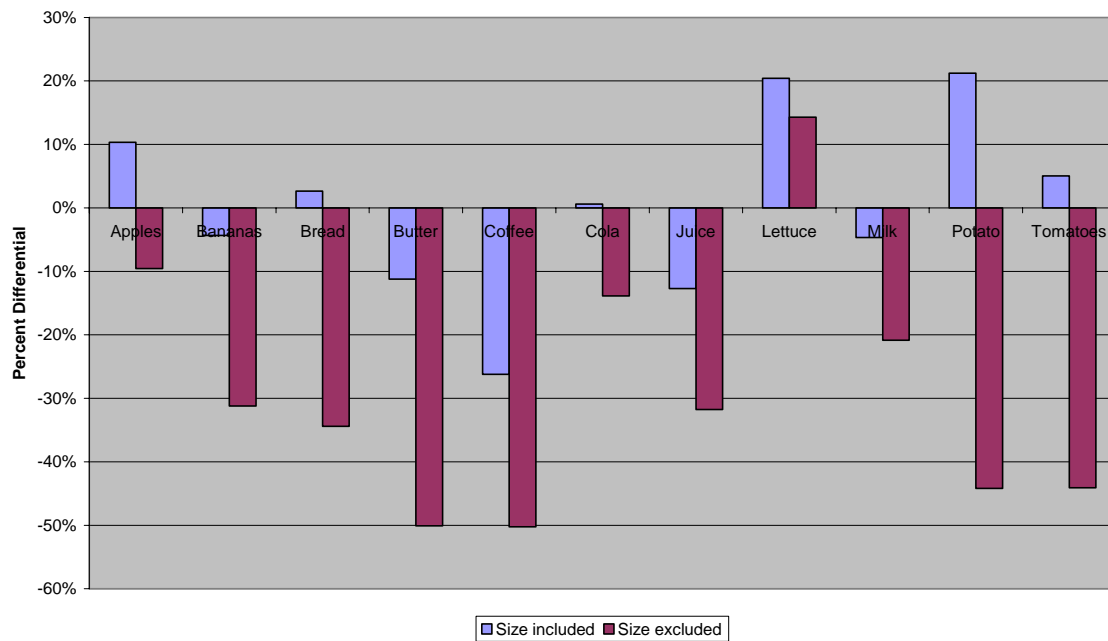


Table 1
Item Categories

White Bread
Yellow Bananas
Chicken Eggs
Ground Beef
Ham, Excluding Canned
Apples
Fresh Whole Milk
Potatoes
Tomatoes
Cola, National Brands
100% Fruit or Vegetable Juices
Roasted Coffee
Butter
Iceberg Lettuce

Table 2
Alternative Indexes
Log-change Jan 2002-Sept 2007, By Item Category

| Item Category A | Unadjusted Index U | Hedonic Index H | Hedonic Index Plus Average Outlet Premium O | Matched Model M |
|--|--------------------------|--------------------|--|--------------------|
| Apples | 0.297 | 0.324 | 0.329 | 0.321 |
| Bananas | -0.031 | 0.020 | -0.032 | 0.018 |
| Bread | 0.179 | 0.226 | 0.190 | 0.208 |
| Eggs | 0.535 | 0.465 | 0.497 | 0.465 |
| Ground Beef | 0.280 | 0.289 | 0.263 | 0.271 |
| Ham | 0.170 | 0.145 | 0.094 | 0.131 |
| Milk | 0.382 | 0.359 | 0.345 | 0.347 |
| Coffee | 0.304 | 0.219 | 0.237 | 0.234 |
| Juice | 0.236 | 0.179 | 0.173 | 0.137 |
| Potatoes | 0.212 | 0.189 | 0.168 | 0.193 |
| Butter | -0.045 | 0.029 | 0.010 | 0.017 |
| Lettuce | -0.057 | -0.077 | -0.076 | -0.083 |
| Cola | 0.108 | 0.187 | 0.185 | 0.166 |
| Tomatoes | 0.004 | 0.018 | 0.003 | 0.017 |
| Weighted Average Across Items | 0.2138 | 0.2186 | 0.2033 | 0.2053 |

| Item Category A | Difference between Hedonic and Unadjusted Indexes (U-H) B | Outlet Component (O-H) C | Item Characteristics Component (U-O) D | Difference between Hedonic and Matched Model Indexes (M-H) E | P values, Hausman test Random effects vs. Hedonic | P values, significance of time trend in explaining outlet effects |
|--|--|-----------------------------------|--|---|--|---|
| Apples | -0.027 | 0.005 | -0.032 | -0.003 | * | 0.0005 |
| Bananas | -0.052 | -0.053 | 0.001 | -0.003 | 0.0333 | * |
| Bread | -0.047 | -0.036 | -0.011 | -0.018 | 0.0885 | * |
| Eggs | 0.070 | 0.032 | 0.038 | -0.001 | 0.0002 | + |
| Ground Beef | -0.009 | -0.026 | 0.017 | -0.018 | 0.0114 | * |
| Ham | 0.025 | -0.051 | 0.075 | -0.015 | 0.1972 | * |
| Milk | 0.023 | -0.014 | 0.037 | -0.012 | 0.1186 | * |
| Coffee | 0.085 | 0.018 | 0.067 | 0.015 | 0.5146 | 0.7284 |
| Juice | 0.058 | -0.006 | 0.064 | -0.042 | 0.1302 | 0.2844 |
| Potatoes | 0.023 | -0.021 | 0.044 | 0.004 | 0.0846 | * |
| Butter | -0.073 | -0.018 | -0.055 | -0.012 | 0.0086 | * |
| Lettuce | 0.019 | 0.001 | 0.019 | -0.006 | 0.0004 | + 0.0002 |
| Cola | -0.080 | -0.002 | -0.078 | -0.021 | 0.0222 | * |
| Tomatoes | -0.014 | -0.014 | 0.001 | -0.001 | 0.0073 | 0.1513 |
| Weighted Average Across Items | -0.0048 | -0.0154 | 0.0105 | -0.0134 | | |

* indicates p-value less than 0.0001. + indicates coefficient is positive.

Table 3
Price Differentials Relative to Large Grocery Stores
Not Adjusted for Item Characteristics

| Item Category | Discount Dept. | Warehouse Club | Small Grocery | Convenience | Other |
|------------------|----------------|----------------|---------------|-------------|--------------|
| Apples | -2.7% | -31.0% | -17.8% | 0.0% | -40.5% |
| Bananas | -13.5% | -34.3% | 13.2% | 25.0% | -1.1% |
| Bread | -22.7% | -28.2% | -15.3% | -8.8% | -11.2% |
| Eggs | -12.0% | 25.1% | -18.3% | -13.1% | 21.0% |
| Ground Beef | -12.4% | -30.4% | -14.6% | -18.5% | 1.1% |
| Ham | -17.3% | 0.4% | 0.8% | 36.2% | 31.6% |
| Milk | 1.2% | -29.6% | -8.7% | 5.5% | -1.2% |
| Coffee | -19.8% | -43.0% | -4.1% | 71.6% | 87.8% |
| Juice | -16.8% | -31.3% | 9.9% | 58.9% | 5.2% |
| Potatoes | -12.0% | -40.3% | -33.3% | -20.2% | -11.7% |
| Butter | -24.3% | -62.5% | -13.0% | 10.2% | -20.5% |
| Lettuce | -12.9% | 11.5% | -14.8% | 27.7% | -10.2% |
| Cola | -7.2% | -6.1% | 25.1% | 44.7% | 26.4% |
| Tomatoes | -13.7% | -20.2% | -35.2% | -21.2% | -44.4% |
| Weighted Average | -12.3% | -23.4% | -6.4% | 13.1% | 3.7% |

Bold values indicate that the difference is not statistically significant at the .05 level

Table 4
Price Differentials Relative to Large Grocery Stores
Adjusted for Item Characteristics

| Item Category | Discount Dept. | Warehouse Club | Small Grocery | Convenience | Other |
|------------------|----------------|----------------|---------------|--------------|--------------|
| Apples | -5.7% | 10.3% | -20.7% | 0.0% | -39.2% |
| Bananas | -11.6% | -4.3% | 0.2% | 10.7% | -1.5% |
| Bread | -4.0% | 2.6% | -14.1% | -5.5% | -17.8% |
| Eggs | -17.3% | -6.4% | -17.3% | -2.7% | 11.0% |
| Ground Beef | -7.4% | -35.4% | -8.1% | -11.0% | -3.0% |
| Ham | -20.2% | -20.3% | -12.7% | -16.8% | 33.8% |
| Milk | -1.2% | -4.7% | -8.8% | 1.7% | -5.3% |
| Coffee | -20.9% | -26.2% | -12.3% | 50.4% | 25.0% |
| Juice | -20.9% | -12.7% | -0.1% | -14.4% | -14.1% |
| Potatoes | 1.8% | 21.2% | -17.8% | -24.6% | -20.9% |
| Butter | -21.0% | -11.2% | -11.4% | 22.8% | -4.3% |
| Lettuce | -15.8% | 20.4% | -10.9% | 34.8% | -13.9% |
| Cola | -6.2% | 0.6% | 5.7% | 6.1% | 3.5% |
| Tomatoes | -11.2% | 5.0% | -28.4% | -16.4% | -34.2% |
| Weighted Average | -10.0% | -8.5% | -9.3% | -1.8% | -5.4% |

Bold values indicate that the difference is not statistically significant at the .05 level

Table 5
Product Size Comparisons

| | Warehouse Club Stores | | All Other Stores | |
|----------|-----------------------|--------------------------------|-----------------------|-------------------------------------|
| | Median Size in Ounces | Percentile at All-Other Median | Median Size in Ounces | Percentile at Warehouse/Club Median |
| Apples | 96 | 0 | 16 | 99 |
| Bananas | 48 | 9 | 16 | 95 |
| Bread | 48 | 3 | 20 | 97 |
| Milk | 192 | 46 | 128 | 96 |
| Coffee | 39 | 2 | 16 | 84 |
| Juice | 192 | 0 | 64 | 99 |
| Potatoes | 160 | 0 | 80 | 74 |
| Butter | 64 | 0 | 16 | 99 |
| Lettuce | 46 | 58 | 48 | 41 |
| Cola | 288 | 0 | 144 | 77 |
| Tomatoes | 42 | 0 | 16 | 98 |

Table 6
Outlet Effect Decomposition by Outlet Category, January 2002 to September 2007
Weighted Sample Average of Item Categories

| | Large Grocery | Discount Dept. | Warehouse Club | Small Grocery | Convenience | Other | Total |
|--|---------------|----------------|----------------|---------------|-------------|--------|--------|
| Change in Average Outlet Premium Within Category | -0.07% | 0.07% | 0.03% | -0.21% | 0.02% | -0.41% | -0.57% |
| Change in Weight of Category | -0.15% | -0.60% | -0.26% | 0.01% | -0.05% | 0.08% | -0.97% |
| Total | -0.21% | -0.53% | -0.23% | -0.20% | -0.03% | -0.33% | -1.54% |

References

- Basker, Emek (2005): "Selling a Cheaper Mousetrap: Wal-Mart's Effect on Retail Prices," *Journal of Urban Economics* 58 (September), 203-229.
- Bureau of Labor Statistics (2007): *BLS Handbook of Methods*, "Chapter 17: The Consumer Price Index." Available online at <https://www.bls.gov/opub/hom/pdf/homch17.pdf>
- Conference Board, Inc. (1999): *Measuring Prices in a Dynamic Economy: Re-Examining the CPI*. Special Report 1260-99-SR.
- Fixler, Dennis, and Robert McClelland (2000): "Do Prices Change Uniformly Across Regions?" Bureau of Labor Statistics manuscript.
- Hausman, Jerry, and Ephraim Leibtag (2004): "CPI Bias from Supercenters: Does the BLS Know that Wal-Mart Exists?" National Bureau of Economic Research Working Paper 10712.
- Hausman, Jerry and Ephraim Leibtag (2005): "Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart," NBER Working Paper No. 11809.
- Hoover, Ethel D., and Margaret S. Stotz (1964): "Food Distribution Changes and the CPI," *Monthly Labor Review* 87 (January), 58-64.
- International Labour Office (2004): *Consumer Price Index Manual: Theory and Practice*. Geneva: International Labour Office. Available online at <http://www.ilo.org/public/english/bureau/stat/guides/cpi/index.htm>.
- Lane, Walter E. (20013): "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 online at [http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+6/\\$file/2001%206th%20Meeting%20-%20Lane%20Walter%20-%20Addressing%20the%20New%20Goods%20Problem%20in%20the%20Consumer%20Price%20Index.pdf](http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+6/$file/2001%206th%20Meeting%20-%20Lane%20Walter%20-%20Addressing%20the%20New%20Goods%20Problem%20in%20the%20Consumer%20Price%20Index.pdf).
- Lebow, David E., and Jeremy B. Rudd (2003): "Measurement Error in the Consumer Price Index: Where Do We Stand?" *Journal of Economic Literature* XLI (March), 159-201.
- Lebow, David E., John M. Roberts, and David J. Stockton (1994): "Monetary Policy and 'The Price Level,'" Board of Governors of the Federal Reserve System, July.
- 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.
- Shapiro, Matthew D. and David W. Wilcox (1996): "Mismeasurement in the Consumer Price Index: An Evaluation," *NBER Macroeconomics Annual 1996*, 93-142.
- Strople, Michael H. (2006): "From supermarkets to supercenters: employment shifts to the one-stop shop," *Monthly Labor Review* 129 (February), 39-46.
- U.S. Senate, Committee on Finance (1996): *Final Report of the Advisory Commission to Study the Consumer Price Index*. Print 104-72, 104 Cong., 2 sess. Washington, D.C.: Government Printing Office.
- White, Alan G. (2000): "Outlet Types and the Canadian Consumer Price Index," *Canadian Journal of Economics* 33 (May), 488-505.