CPI Bias from Supercenters: Does the BLS Know that Wal-Mart Exists?

Jerry Hausman and Ephraim Leibtag[1]
MIT and Economic Research Service, U.S. Department of Agriculture
Revised draft, June 10, 2005[2]

I. Introduction

Hausman (2003) discusses four sources of bias in the present calculation of the CPI. The most often discussed substitution bias is a second order bias while the other three sources of bias are all first order in their effects: "new good bias", "quality bias," and "outlet substitution bias." A "pure price" index based approach of surveying prices to estimate a COLI cannot succeed in solving the three problems of first order bias. Neither the BLS nor the recent report C. Schultze and C. Mackie, eds., At What Price (AWP, 2002), recognizes that to solve these problems, which have been long known, both quantity and price data are necessary. We discuss economic and econometric approaches to measuring the first order bias effects from outlet substitution bias. We demonstrate the use of scanner data that permits implementation of techniques that allow the problem to be solved.

Over the past decade, "non-traditional" shopping formats have captured significant share from "traditional grocery." P. Little (2004) describes the two categories of alternative retail outlets as "high-spend" outlets, which are low price, one-stop shopping destinations, and "low and medium-spend" stores which are mostly convenience stores that serve a "fill-in" role in between trips to the "high-spend" outlets He includes supercenters (Wal-Mart, Kmart, Meijer, etc.), warehouse clubs (Sam's Club, Costco and BJ's), and mass merchants (Wal-Mart, Kmart, Target, etc.) as the primary outlets for these "high-spend" expenditures. [3] Using 2003 data, he estimates that these outlets have 24.8% of food expenditures, with supercenters having 45.6% of the category. Over the

[1] Earlier draft presented at Index Number conference in Vancouver, June 2004. We thank Jie Yang and Ketan Patel for outstanding research assistance. John Greenlees and Marshall Reinsdorf provided helpful comments. Author contact: jhausman@mit.edu
[2] First draft: do not quote or cite without permission. No views in this paper reflect the U.S. Department of Agriculture's position on these issues.

past few years Wal-Mart has become the largest supermarket chain in the U.S. Wal-Mart, excluding its Sam's Club, now has supermarket- related revenues approximately 51% larger than the runner-up Kroger, and larger than Albertsons and Safeway, the third and fourth largest supermarket chains, combined. Nationwide Wal-Mart has a 14% market share (in 2003), despite not being in a number of regional markets, and an 18% share when Sam's Clubs are included. Within the "medium-low spend" category, Little estimates convenience stores that also sell gasoline as the fastest growing store type with 85.5% of the 12.4% total share for the category. Little calculates that total traditional grocery outlets, including conventional supermarkets and superstores (a larger version of the conventional supermarket), have decreased to a 56.3% dollar share in 2003. He also forecasts that in 5 years, the "high-spend category" will grow from 24.8% to 31%, with supercenters comprising 54.8% of the total while traditional grocery outlets decrease from 56.3% to 48.3%. Thus, he expects Wal-Mart to become increasingly important over the next few years, continuing the trend of change over the past decade.

Wal-Mart began selling food in 1988 and in 2002 because the U.S. largest grocery chain. Wal-Mart now is larger than Kroger, Albertsons, and Safeway, which are the next largest supermarket chains. Significant consolidation has occurred in the supermarket industry, but Wal-Mart continues to grow at a significantly faster rate than these supermarket chains. The majority of Wal-Mart's grocery sales arise from its over 1400 (as of April 2004) supercenters which are 180,000 square foot stores that are both discount stores and grocery stores, although it also has "Neighborhood Market" stores that are about the size (40,000 sq. feet) of an average supermarket.[4] While most of the stores are in the South and Southwest, Wal-Mart is increasing moving into urban centers with openings expected in Los Angeles and Chicago, along with other urban centers.[4]

Over the 10-year period from 1991-2001 margins increased in supermarkets as the price of food sold at supermarkets grew at approximately twice the rate of the PPI for

[3] Sam's Club is owned by Wal-Mart.
[4] Wal-Mart management has given guidance that it expects to open between 230-240 new supercenters in 2005 for an increase of about 16%. See Dow Jones, "Factiva," April 19, 2004. Morgan Stanley reports that Wal-Mart is seeking 16%-17% growth in supermarket sales compared with 3% industry growth. See M. Wiltamuth and R. Fariborz, "Food Retail," June 2004. Wal-Mart has grown at a 16% rate over the past three years.
[4] Wal-Mart has sometimes had difficulty in receiving planning approval for its stores. Currently, Wal-Mart has either no presence or an extremely limited presence in New England, the New York metro area, California, and the Pacific Northwest. However, its expansion into new areas has proceeded over the past few years.

food. Over this period the PPI for finished food increased by 13.9% while the CPI for food at home increased by 27.7%. Profit margins for supermarket also increased over the same time period with Kroger's operating profit margin growing from 3.3% to 4.7% and Safeway's operating profit margin growing from 3.5% to 7.9%.[5] Various studies have demonstrated that food items at Wal-Mart are 8%-27% lower priced that at the large supermarket chains, even after discounts for loyalty card and other special are taken into account.[6] After entry by Wal-Mart conventional supermarkets typically decrease their prices (or do not increase them as much as in non-Wal-Mart markets) because of the increased competition.

Remarkably, the large expansion and continuing expansion of Wal-Mart and other supercenter food outlets has almost no effect on the BLS calculation of the CPI for food.[7] The BLS employs a "linking procedure" that assumes "quality-adjusted" prices at Wal-Mart are exactly equal to prices at conventional supermarkets. Thus, when a Wal-Mart store replaces, say a Kroger, in the BLS sample of stores from which it collects prices, it links the lower Wal-Mart price to the higher Kroger price to remove any difference. Even though packaged food items are physically identical at the two stores, the BLS procedure does not recognize any price difference between the stores. This procedure is not based on any empirical study. Rather, it is based on mere assumption. The assumption is completely inconsistent with actual real world market outcomes where Wal-Mart has expanded very quickly in markets that it entered. Thus, the market impacts of Wal-Mart and other supercenters are nowhere in the food CPI so that we find that the BLS does not know that Wal-Mart "exists" in terms of the estimation of a CPI. We also believe that observed consumer behavior cannot be explained by the BLS assumption of a compensating "quality differential." We specify a theoretical model of consumer behavior that demonstrates this point below.

---

[5] Calculations based on companies SEC 10-K filings. Callahan and Zimmerman (2003) also report increased profit margins for supermarkets over this period.

[6] A recent December 2003 study by UBS Investment Research found a price gap of 17.3% to 26.2%, "Price Gap Tightens, Competition Looks Hot Hot Hot." The previous year UBS found a price gap of 20.8% to 39.1%. For example for a specified identical market basket UBS finds Wal-Mart supercenters to have an average price 19.1% less expensive in Tampa and 22.8% less expensive in Las Vegas. In 2002, Salomon Smith Barney estimated the price gap to be between 5% and 25%. See L. Cartwright, "Empty Baskets, September 12, 2002.

[7] When customers shift from conventional supermarkets to Wal-Mart no change occurs in the food CPI. To the extent that prices at Wal-Mart decrease (or increase) at a different rate than conventional stores, the food CPI will

II. Current BLS Procedure

The BLS methodology updates its samples of stores from which it collects prices periodically. It makes two adjustments. First, the BLS updates the stores at which these purchases are made. Next, the BLS updates the products in the market basket that consumers purchase.[8] Cage (1996) describes the current BLS sampling procedure, in which the "Telephone Point-of-Purchase Survey" (TPOPS) is used to provide a sampling frame of outlets visited by urban consumers. Approximately 25% of all sampling units participate in a given year. While the products can change, note that the expenditure shares across categories did not change with this procedure. The expenditure shares are only updated on a considerably less frequent basis since the TPOPs data does not collect expenditure data or quantity data. Thus, the BLS probability sampling procedure works against solving the outlet bias problem discussed in this paper.[9]

When the BLS collects data, it collects the name and address of the retail establishment reported by respondents and estimates of the total daily expenditure by TPOPS category. The expenditure weights are not used to update the expenditure weights used in the weighted average of prices, rather they are used in the selection of outlets so that those outlets with larger expenditure weights receive a greater probability of selection.

TPOPS outlet rotation allows a closer approach to actual consumer shopping patterns as they change. As more households shop at Wal-Mart, the probability of a Wal-Mart being included in a given market increases. Item rotation also occurs as discussed above. However, when an identical item is sampled at the new outlet, even if the product is physically identical to the item sampled in the old outlet, the BLS does not take account of the lower price. Thus, if a 12 ounce box of Kellogg's Rice Krispies is purchased at a Wal-Mart that is newly included to replace a Kroger that has been dropped; the BLS links the new lower price to the old higher price so no price change occurs. This linking

---

take account of this change with a lagged effect over time.

[8] The BLS sometimes takes a very long time to incorporate new products in the market basket as in the case of cellular telephones, which were not included for 15 years after their introduction. Hausman (1999) demonstrates the significant bias from the delay in the introduction of cellular telephone.

[9] We thank John Greenlees and Marshall Reinsdorf for pointing out the problem arising from using the probability sampling procedure in terms of outlet bias.

procedure creates outlet substitution bias in the estimation of the CPI. In the AWP (2002, p. 169) discussion of BLS procedures, it is claimed that consumer shopping comprises a package and that non-monetary benefits exactly balance out the effects of the lower price. This finding was based on absolutely no empirical evidence whatsoever. The finding is also completely inconsistent with the real world market facts that expenditures at supercenters grow quickly when they become available. Indeed, Wal-Mart is now the largest supermarket chain in the U.S.

This "compensating service effect" explanation is also inconsistent with the "indirect price effect" that we estimate subsequently, where we find that as expenditure at superstores increases in a given market, the prices at traditional supermarkets decrease.[10] For example, after two Wal-Mart supercenters opened in Houston, a nearby Kroger's sales dropped 10%, the Kroger store reduced worker hours by 30%-40%, and it decreased its prices.[11] Presumably this price decrease is caused by greater competition. Thus, consumers demonstrate with their expenditure choice that they prefer lower priced outlets, and the higher priced supermarket must respond in a competitive manner. The AWP description of the BLS assumption that markets are in equilibrium is inconsistent with the real world market data, which find that prices from traditional stores decrease from the increased competition.

Thus, when a new set of stores are included in the BLS sample, the linking procedure eliminates all of the price differences. Even though the box of Kellogg's Rice Krispies is identical in all respects, the BLS assumes that differences in outlet and product characteristics completely explains the difference in price. Thus, lower prices from increased expenditure at superstores have no effect on the CPI. Thus, the BLS assumes that Wal-Mart does not exist in constructing the CPI.

Reinsdorf (1993) found that food and motor-fuel prices during a two-year overlap period led to new samples prices being lower by about 1.25% compared to the outgoing samples. Since sample rotation occurs every 5 years, this finding would create a 0.25% bias per year. However, Reinsdorf's quantitative findings have not been totally accepted because of concerns about product quality as well as differences in coverage. The AWP

[10] Shapiro and Wilcox (1996) also noted this indirect effect.

[11] P. Callahan and A. Zimmerman (2003) report on these effects. The regional head of Kroger's stated, "Wal-Mart

(2002, p. 176) recommended that the BLS continue its current practice and disregard the effect of Wal-Mart and other supercenters on prices and price indices.

III. A Utility-Consistent Economic Model of Shopping Destination

The BLS assumes that an exact compensating "quality differential" exists between shopping at a supercenter store with its lower prices and a conventional supermarket. Service quality and other factors supposedly allow the BLS to assume that quality adjusted prices are exactly the same when the BLS links the prices. However, this assumption is inconsistent with real world market behavior that finds when Wal-Mart opens a store in a new geographic market, it rapidly gains share while conventional supermarkets lose share.[12] We believe that a better model than the implicit BLS model is to consider Wal-Mart supercenters as a new choice to consumers. Some consumers find the choice to be superior while others continue to shop at conventional supermarkets.[13] Thus, the arrival of Wal-Mart in a given geographic market is similar to the introduction of a new good into the geographic market. Hausman (1997, 2003) discusses how new products should be included in a correct cost of living index (COLI). Here, rather than a completely new product, e.g. cellular telephones, an existing product is expanded into a new geographic market. However, the effect on consumers is similar since they now have increased choice in their shopping trips.

For our economic model we consider the conditional choice of consumers to shop at either a conventional supermarket or at a lower price, and perhaps lower service quality, supercenter. For ease of exposition, we use a two-stage choice model in which at the lower stage the consumer considers his or her shopping behavior conditional on type of store. The consumer calculates a price index for shopping at either type of store, takes

made us look at ourselves and reinvent ourselves."

[12] Supermarket chains sometimes exit a geographic market after Wal-Mart enters. Albertsons exited the Houston market after Wal-Mart entry. However, in our model we assume that consumers continue to have access to traditional supermarkets, even if a given chain exits the market.

[13] As we discussed above, these conventional supermarkets typically decrease price because of the increased competition from Wal-Mart. If the BLS consistently applied its "quality adjustment" procedure it would ignore these price decreases at conventional supermarkets because presumably they arise from reduced service quality. However, the BLS fully incorporates these price decreases, demonstrating that its approach is based on no correct economic assumptions.

account of service and other quality differences, and then at the upper stage decides which type of store to shop at.[14] We use the two-stage approach of Hausman (1985) and Hausman, Leonard and McFadden (1995), although neither of the models was designed precisely for the situation of shopping destination choice.

We allow for consumers choice of shopping at either a conventional supermarket, j=1, or at a supercenter, j=2. Conditional on choosing to shop at one of these two types of stores the consumer has a *conditional expenditure function*

$$y = e(p_{0,} p_1{}^j, p_2{}^j, ..., p_n{}^j; \bar{u}) = e(p, \bar{u}) \text{ solves } \min \sum_i p_i x_i \textit{ such that } u(x) = \bar{u} \quad (3.1)$$

where p_0 is a vector of prices of all non-food items assumed the same for destination choice, $p^j = \{ p_1{}^j, p_2{}^j, ... p_n{}^j \}$ are the prices of the n goods in the two types of outlets denoted by the superscript j, and $\bar{u}$ is the utility level of the consumer.[15] The conditional demand for each type of product, depending on the type of outlet j chosen is:

$$x_i{}^j = \frac{\partial e(p_0, p^j, \bar{u})}{\partial p_i{}^j} = \frac{-\dfrac{\partial v(p_0, p^j, y)}{\partial p_i{}^j}}{\dfrac{\partial v(p_0, p^j, y)}{\partial y}} \quad i = 1..., n \qquad (3.2)$$

where the indirect utility function v(p,y) is derived from the duality relationship with the expenditure function. Using duality corresponding to any level of utility in equation (3.1) and any vector of prices, a price index exists that corresponds to the minimum expenditure required to achieve a given level of utility $\bar{u}$. Indeed, the utility consistent price index is the level of expenditure needed to achieve the utility level:

$$\Pi(p^j, \bar{u}) = e(p^j, \bar{u}) = y^j(p^j, \bar{u}) = y^j = \sum_i p_i{}^j x_i{}^j \qquad (3.3)$$

[14] We assume that consumers do not divide their shopping trips between different types of stores, although this behavior could be incorporated into the model.

[15] As written, equation (3.1) assumes that both types of stores carry all goods. To the extent that supermarkets carry a wider variety of products that supercenters, the prices for supercenters can be entered as virtual prices that set demand to zero. See Hausman (1997) for an explanation of virtual prices.

An “average price” $\overline{p}^{\,j}$ can then be calculated by dividing y^j by a quantity index $\overline{x}^{\,j}$ so that $y^j = \overline{p}^{\,j}\overline{x}^{\,j}$.[16]

We now move to the top level where the consumer decides whether to shop at the conventional supermarket or at the supercenter outlet. We expect $y^1 > y^2$ because most prices in supermarkets exceed the prices in supercenters. Consider the use of the binomial logit model for choice between traditional supermarkets and supercenters.[17] The probability of choosing the traditional supermarket is:

$$pr(j=1) = \frac{1}{1+\exp(\beta_0 + \beta_1(\overline{p}^1 - \overline{p}^2))} \qquad (3.4)$$

where a log price index or other type of price index (e.g. a Stone price index) can also be used depending on the precise form of the underlying expenditure (utility) and demand functions in equation (3.1) and (3.2).[18]

If we can assume that the overall units of a good are the same, e.g. Kellogg’s Rice Krispies, we can simplify so that the overall demand for good i becomes:

$$\hat{x}_i(p_0, p^1, p^2, y) = pr(j=1)x_i(p_0, p^1, y) + pr(j=2)x_i(p_0, p^2, y) \qquad (3.5)$$

where the right hand side demands are the conditional demands from equation (3.2) and have common units. Similarly, to calculate the unconditional price for the representative consumer we take overall expenditure on good i and divide by the quantity of equation (3.5):

$$\hat{p}_i(p_0, p^1, p^2, y) = \Xi_i(p_0, p^1, p^2, y) / \hat{x}_i(p_0, p^1, p^2, y) \qquad (3.6)$$

---

[16] Instead of the average price we can also divide expenditure by utility to get a “cost of utils” index.

[17] Because of only two choices, the independence of irrelevant alternative assumption does not create a problem here. With more than two choices a nested logit or multinomial probit model could be used. See Hausman et. al. (1995) for a derivation with the nested logit model.

[18] An exact aggregation approach when using a Gorman generalized polar form appears in Hausman et. al. (1995).

where Ξ_i is expenditure on good i. If choice j = 2 does not exist in a given geographic market, then the price index of equation (3.6) is just the traditional supermarket price so that $\hat{p}_i(p^1, p^{2*}, y) = p^1$ where p^{2*} are the virtual prices, which cause demand at supercenters to be zero.[19] But when supercenters become available, consumers who choose to shop at supercenters do so to maximize their utility and the correct price index is an expenditure weighted average of the two prices. This expenditure weighted approach to price averages is the procedure we use in the empirical work that follows.

Thus, the exact cost of living index becomes

$$P(p_0, p^2, p^1, \bar{u}) = \frac{y^2(p_0, p^1, p^2, \bar{u})}{y^1(p_0, p^1, p^{2*}, \bar{u})} = \frac{e(p_0, p^1, p^2, \bar{u})}{e(p_0, p^1, p^{2*}, \bar{u})} \tag{3.7}$$

which gives the ratio of the required amount of income when supercenters are present in the market compared to the situation where supercenters are not present and prices are at the virtual level p^{2*}, which causes demand to be zero. Equation (3.7) demonstrates how the new good approach applies to supercenters when the correct unit of observation is a geographic market, rather than a new product. Taking the appropriate weighted averages of equation (3.7) leads to an expenditure share weighted approach.

Thus, we do not find support for the BLS assumption of an exact compensating quality differential when consumers can choose which type of outlet at which to shop. Some consumers continue to shop at traditional supermarkets when supercenters become available, while other consumers shift to shopping at supercenters. In terms of the representative consumer we calculate the probability weights for each type of choice multiplied by the demand at each type of outlet and divide this weighted demand into expenditure to derive the price index. As more supercenters become available in a given

[19] To the extent that traditional supermarkets close because of increased supercenter competition, consumers have decreased choice, which could effect price index calculations. However, in the model we assume that consumers still have the choice to shop at one or more traditional supermarkets, i.e. that not all supermarkets in a given geographic market close. In this situation which is consistent with actual market outcomes, the effect on a theoretical price index would be extremely small. Indeed, supermarkets that close typically have the smallest customer base, which further decreases the effect of store closings on a price index.

geographic market, more consumers choose to shop at supercenters and its expenditure weight increases. We continuously update the expenditure weights to allow for this observed market determined change in shopping destination choices. Consumers in their revealed preference choices determine the appropriate weights to be used in the price index.

In this section we have specified a model of consumer outlet choice and demonstrate how to calculate an exact cost of living index (COLI) in equation (3.7). In Hausman-Leibtag (2005) we estimate this model and estimate the compensating variation that arises from the spread of supercenters. The compensating variation is the difference of the numeration and denominator of equation (3.7) while the exact cost of living index is the ratio. As discussed in e.g. Hausman (2003), the BLS has recognized (Abraham et al. 1998) that a COLI provides the correct approach. The BLS approach attempts to approximate a true COLI with the CPI so that our theoretical derivation demonstrates that the current BLS CPI approach potentially can cause significant bias. We now apply a BLS-type procedure to the data and find that the CPI bias can be significant.

We find that Wal-Mart should exist in the estimation of a price index, contrary to the current BLS procedure. However, note that as Hausman (2003) emphasized, to implement this approach both *prices and quantities* need to be available, which necessitates the use of scanner data. The BLS approach, which only collects price data, cannot implement the correct price index approach. Without quantity data, the BLS will always be required to make one or another arbitrary assumption regarding "service adjusted" quality levels. Observation of actual consumer choice in terms of quantities purchased allows us to resolve the problem.

IV. Data Description

This study uses a customized subset of the ACNielsen Homescan scanner panel data for the four years 1998-2001. The ACNielsen Homescan data is a consumer panel consisting of approximately 61,500 randomly selected households across the U.S. and includes purchase as well as demographic information for all households in the sample. Homescan households are randomly recruited to join the panel using sampling techniques to ensure household representation for demographic variables such as household income, family composition, education, and household location. Each household is equipped with an electronic home-scanning unit, and household members record every UPC-coded food purchase they make by scanning in the UPC of the food products that they buy from all retail outlets that sell food for home consumption.

The panel is recruited on a permanent basis, subject to turnover from normal attrition or adjustments to demographic targets necessitated by Census revisions.[20] The Homescan panel is considered by many in the food industry as the most reliable household based panel data due to its long-standing reputation in the marketplace and its utilization of hand-held technology that minimizes the recording burden for participants. The ACNielsen Homescan consumer panel collects consumer shopping and purchase data from all outlet channels, including grocery, drug, mass and convenience stores. The panel is geographically dispersed and is demographically balanced so the sample profile matches the US population as closely as possible. The panel data are also projected to census estimates that are updated regularly to reflect population changes.

Household panel data allow for observation of the ongoing purchase habits and practices of household and demographic groups. Tracking and analyzing this information over time can reveal the dynamics of consumer behavior such as who is buying what products, what different products are purchased during a given shopping trip, and how often a product is purchased. Panel data quantify the composition of category or brand volume which can be used to measure the impact of store choice on the purchase level of product quantities and prices. Data are collected after each panelist shopping trip. Members of the panel record their purchases, capturing not only what is purchased, but also where the purchase was made, and whether the purchase was a promotional, sale, or coupon item.

These data are useful in price analysis since we are able to observe actual purchase choices by consumers. However, in terms of food purchase behavior, the key missing information is consumer purchases of food away from home (primarily restaurant meals) so one needs to assume that the unknown levels of food away from home purchases do not somehow bias the average prices paid by an individual household for their food at home purchases. Once this assumption is made these data are useful for

[20] Households lost through attrition are replaced with others having similar key characteristics.

analysis of the impact of store choice on average prices paid for food at home items. Consumer panel information can be used to measure the average prices paid by a representative group of households over time. This measurement of average price paid can be aggregated across households and/or across time to measure price change for different categories of products.

Along with the description of each product, the price and quantity that was purchased is recorded on a daily basis. National and regional level aggregates can be calculated using transaction data from households located in 50 local U.S. markets as well as households in non-metro/rural areas that are included in this data set. For 21 of these 50 markets, a large enough number of panelists are included to enable comparisons across markets for all UPC-coded products.[21]

The Economic Research Service (ERS) of the USDA purchased a sub-sample of transaction level data from the Fresh Foods Homescan Panel[22] comprised of households that not only recorded their UPC-coded transactions, but also recorded their random-weight (non-UPC coded) food purchases over the year(s) that they participated in the panel. This sub-sample was used for this study in order to be able measure the entire market basket of household purchases of food for at-home consumption[23]. Of this group of 15,000 households per year, the sample was restricted to households that participated in the panel for at least 10 out of 12 months per year[24].

---

[21] Albany, Atlanta, Baltimore, Birmingham, Boston, Buffalo-Rochester, Charlotte, Chicago, Cincinnati, Cleveland, Columbus, Dallas, Denver, Des Moines, Detroit, Grand Rapids, Hartford-New Haven, Houston, Indianapolis, Jacksonville, Kansas City, Little Rock, Los Angeles, Louisville, Memphis, Miami, Milwaukee, Minneapolis, Nashville, New Orleans-Mobile, New York, Oklahoma City-Tulsa, Omaha, Orlando, Philadelphia, Phoenix, Pittsburgh, Portland, Raleigh-Durham, Richmond, Sacramento, Salt Lake City, San Antonio, San Diego, San Francisco, Seattle, St. Louis, Syracuse, Tampa, Washington, D.C.

[22] The Fresh Foods Homescan Panel contained 12,000 households in 1998 and 1999 and was expanded to 15,000 households in 2000 and 2001.

[23] If only UPC-coded products were used to measure food-at-home expenditures, many fruit, vegetable, meat, and poultry purchases would not be recorded in the data and food-at-home expenditure shares by store type would not accurately measure true household and market expenditure shares. This is especially true in this situation when alternative channel stores sell less random weight items than conventional retailers. Leaving out random weight items would then tend to overstate the shares of food expenditures of alternative retail outlets.

[24] In total, there were 9,501 unique households in the data with some subset participating each year creating a total of 28,996 household by year observations. In 1998 there were 7,624 households, 7,124 households in 1999, 7,523 households in 2000, and 8,216 households in 2001. Some households participated in the panel for more than one year. Of the 9,501 households in the data, 5,247 households participated for all four years, 1,877 households participated for three years, and 2,377 households were one year participants.

Standard demographic information is collected on an annual basis from each household and each household's home market/city and census region is identified for stratification purposes (see below). Each household is then assigned a projection factor (weight) based on its demographics in order to aggregate the data to be representative at the market, regional, and national level.[25]

These data were constructed based on a stratified random sample with households as the primary sampling unit. A stratified random sample is used to ensure that the sample of households matches Census-based demographic and geographic targets. One function of the design is to allow description of 8 major markets for cross-market comparisons.[26]

The strata for 1998 and 1999 are based on six cities (ACNielsen major markets) Atlanta, Baltimore/Washington, Chicago, Los Angeles, New York, and San Antonio. All other households fall into one of four census regions: East, Central, South, and West.

1998-1999

Stratum	Description
1	Atlanta
2	Baltimore-Washington
3	Chicago
4	Los Angeles
5	New York
6	San Antonio
For all other households- Census Regions are used as strata:	
7	East
8	Central
9	South
10	West

[25] Age, gender, education, occupation, of head(s) of household, number of household members, household income, household composition, race, and ethnicity.

Nielsen augmented their stratification scheme in 2000, selecting 2 additional major markets.

2000-2001

Stratum	Description
1	Atlanta
2	Baltimore-Washington
3	Chicago
4	Los Angeles
5	New York City
6	Philadelphia
7	San Antonio
8	San Francisco
For all other households- Census Regions are used as strata:	
9	East
10	Central
11	South
12	West

There was no known or intentional clustering in the sample construction. The projection factor (weight) reflects the sample design and demographic distribution within the strata.

The information that is captured on a transaction level basis includes: date of purchase, store name and channel type identifier[27], store department identifier[28], item description, brand name, number of units purchased, price paid, promotions/sales/coupons used (if any). For retail stores that ACNielsen tracks with their store-level scanner data[29], prices are verified through store-level price and promotion checks.

---

[26] Atlanta, Baltimore/Washington, Chicago, Los Angeles, New York, Philadelphia, San Antonio, San Francisco.

[27] Grocery, Drug, Mass Merchandiser, Supercenter, Club, Convenience, Other (including dollar stores, bakeries, military stores, online purchases, health food stores, and vending machines)

[28] Dry Grocery, Dairy, Frozen-Produce-Meat, Random Weight.

[29] The ACNielsen store-level sample is updated through both replacement of canceled or closed stores and *Continuous Sample Improvement Program* -- when the sample is changed intentionally to ensure that changes in the universe are reflected in the sample.

Warehouse shipment data are used to supplement scanner-generated data collected from households or provided to ACNielsen through their store-level scanner data. Warehouse shipment data are used to estimate the balance of sales moving through other food retailers. This information is Census data (i.e., non-projected, actual shipment data) supplied to ACNielsen by wholesale co-operators.

Some question the quality of household panel data when they try to reconcile it with store-level scanner data. There is the perception that the volumetric data from each source should be the same. However, panel data and store data are not always equal because measurement methodologies differ. Store-level data record millions of shopping transactions while panel data record a specific group of shoppers. In addition, panel data only represent household-based purchases, so there are no small businesses or other institutional purchases included in the panel.

Both types of information have their uses, and by combining the two, one can quantify the composition of volume, understand the reasons behind consumer behavior changes, and measure the impact of store choice on average prices. Store-level scanning data may show that sales were down in a particular store for some group of products in a given time period. Panel data provide insight into whether the lost volume is due to fewer buyers or if the existing buyers purchased less at the given store or chain of stores. Panel data also provide information on which competitors gained the lost expenditures of the store in question.

V. Effects on Prices

In producing the CPI, BLS makes the implausible assumption that all differences in price between supercenters and other stores are due to quality differences. The empirical analysis in this section proceeds from the opposite assumption that the price differences in food items represents a gain to shoppers at supercenters. While this assumption may not be entirely accurate, we believe it is much closer to the actual economic outcome than the BLS assumption. In Hausman-Leibtag (2005) we estimate an econometric model that allows us to relax this assumption. However, the results presented below illustrate the size of the CPI bias that could arise from recognition of the consumer benefits arising from supercenters. Further, since the BLS has refused to use econometric estimate of demand parameters in calculating the CPI, the approach we use here is considerably more accurate than the current BLS approach as the Hausman-Leibtag (2005) results demonstrate.

Our empirical approach first investigates the effect of supercenters, mass merchandisers, and club stores, (hereafter SMC) on prices paid by households. Two effects are present. The "total" effect is that as more of these superstores operate in a given geographic market, the average prices paid by households will decrease. Prices for food categories in superstores are typically 5%-48% less than prices for the same product in supermarkets and other conventional retail outlets. Thus, as a high proportion of households buy their food at non-traditional retail outlets, the average price paid in a market will decrease.

A. Price Difference between Supermarkets and Superstores

In Table 5.1 we calculate the ratios of average prices across different types of outlets for 20 food categories. Column 2 compares the prices for the food categories in traditional supermarkets compared to prices for these same categories in SMCs (non-traditional stores).

Table 5.1: Ratio of Supermarket and Other Outlet Prices to Superstore Prices

Product	Supermarkets/SMC	All Other/SMC
Apples	1.546	1.531
Apple Juice	1.585	1.596
Bananas	1.384	1.368
Bread	1.108	1.098
Butter/Margarine	1.096	1.096
Cereal	1.172	1.166
Chicken Breast	1.408	1.411
Coffee	1.373	1.383
Cookies	1.223	1.214
Eggs	1.312	1.305
Ground Beef	1.372	1.367
Ham	1.967	1.984
Ice Cream	1.320	1.331
Lettuce	2.117	2.107
Milk	1.207	1.199
Potatoes	1.412	1.402
Soda	0.891	0.974
Tomatoes	1.358	1.321

Bottled Water	1.058	1.165
Yogurt	1.413	1.411
Average	1.300	1.306

The largest difference in average price was for lettuce where SMC prices were about 50% lower than traditional supermarkets over the 48 month period. Bottled water was the lowest price difference with SMC prices about 5% less expensive. Soda was the only item with a lower price in traditional supermarkets than in SMCs. When we take an average across all of the food categories we find that SMCs have prices that are 27% lower than traditional supermarkets. We find this difference to be quite large.[30]

In considering the results of Table 5.1 a concern can arise that superstores and supermarkets are selling a different mix of produce, e.g. types of apples could differ across the outlets. However, we estimate approximately the same price ratio for apples and apple juice across the two types of outlets. Thus, while different product mixes remain a topic for future research, the price differences we estimate across different outlet types are unlikely to arise primarily from different product mixes. Further, price comparisons of identical products that we discuss in the introduction have found price ratios of approximately the same size that we estimate in Table 5.1.

In column 3 of Table 5.1 we compare the price in all non-SMC outlets, including traditional supermarkets, to the price of these food categories in superstores. We find the results to be quite similar with the main differences occurring in soda and bottled water. We find the same overall results that SMC stores offer significantly lower prices than other retail outlets.

We do not find any indication that SMC stores change (increase) their prices at a greater or lower rate than traditional supermarkets and other retail outlets in some additional analysis not reported here. However, we cannot do the comparison of price changes in equilibrium because as the presence of SMC stores increases, traditional retail outlets, and most importantly traditional supermarkets, decrease their prices as a competitive response.

[30] The estimated difference is in line with stock analyst reports who have previously sampled the difference in prices over a very few markets.

B. Total and Indirect Effects on Prices from Superstores

Another important effect exists from the expansion of SMC stores. Their increasing presence also increases competition among traditional food retailers. These supermarkets must decrease prices to remain competitive. The well-publicized strike in the Los Angeles area in late 2003 through early 2004 when traditional supermarkets wanted to decrease health benefits for their employees demonstrates the effect that potential entry of supercenters can have on competition. We call this SMC effect on traditional supermarkets the indirect price effect. Both the total and indirect price effects we estimate lead to lower average prices for households.

To investigate both the total and indirect effects on average prices, we do an econometric analysis using the ACNielsen Homescan data. These data are particularly useful since they provide household data and allow for a stratified random sample of all households. Importantly they provide both price and quantity data across all stores. Since Wal-Mart and some other large superstores no longer participate in the IRI or ACNielsen store level data collection, household data collection provide a source of price and quantity data that are not available elsewhere.

We analyze data at the market level using a fixed effects specification with 48 monthly observations for each market during the period 1998-2001:

$$p_{it} = \alpha_i + \delta_t + \beta e_{it} + \varepsilon_{it} \qquad i = 1,34 \quad t = 1, 48 \tag{5.1}$$

where p_{it} is the average log price paid for a given product, α_i is a fixed effect for a market, δ_t is a monthly fixed effect, e_{it} is percentage expenditure for a given product in superstores, and β is thee elasticity coefficient that wee estimate. We use market fixed effects rather than random effects because expenditure in SMC stores is unlikely to be uncorrelated with the stochastic disturbance, e.g. Hausman (1978). In this situation a fixed effects estimator yields the efficient estimator. However, we make two further econometric adjustments. First, expenditure in superstores on a given product may well not be econometrically pre-determined. Thus, we use instrumental variable estimation (2SLS) where as the instrument we use the overall proportion of food expenditure in

SMC stores in a given market as the instrumental variable. Also, we use an autoregressive model for the stochastic disturbance (AR1) to capture the time series aspect of the data and to achieve more efficient estimates. However, least squares with robust standard errors leads to quite similar results.

For our econometric investigation of 20 food products we use 34 markets, each with over 12,000 food transactions per year. The 34 markets are listed in Table 5.2:

Table 5.2: Markets Used in Econometric Analysis

BOSTON	DENVER	HARTFORD-NEW HAVEN
CHICAGO	DETROIT	PHOENIX
HOUSTON	MIAMI	SALT LAKE CITY
INDIANAPOLIS	MILWAUKEE	COLUMBUS
KANSAS CITY	MINNEAPOLIS	CHARLOTTE
LOS ANGELES	PHILADELPHIA	DES MOINES
NEW YORK	PITTSBURGH	GRAND RAPIDS
SAN FRANCISCO	PORTLAND, OR	OMAHA
SEATTLE	ST. LOUIS	SAN ANTONIO
ATLANTA	TAMPA	SYRACUSE
CINCINNATI	BALTIMORE	
CLEVELAND	BUFFALO-ROCHESTER	

For each of these markets we standardized purchases on a physical unit measure and estimated the effect of increasing purchases in SMC stores. Since we have fixed effects for each market, persistent cost and price differences should be take account of as well as seasonal effects given the presence of monthly fixed effects. We give the econometric estimates for these 20 food categories across the 34 markets in Table 5.3:

Table 5.3: Average Price for Food Products across 34 Markets

National Results

AR(1) IV Results

(Asymptotic Standard Errors)

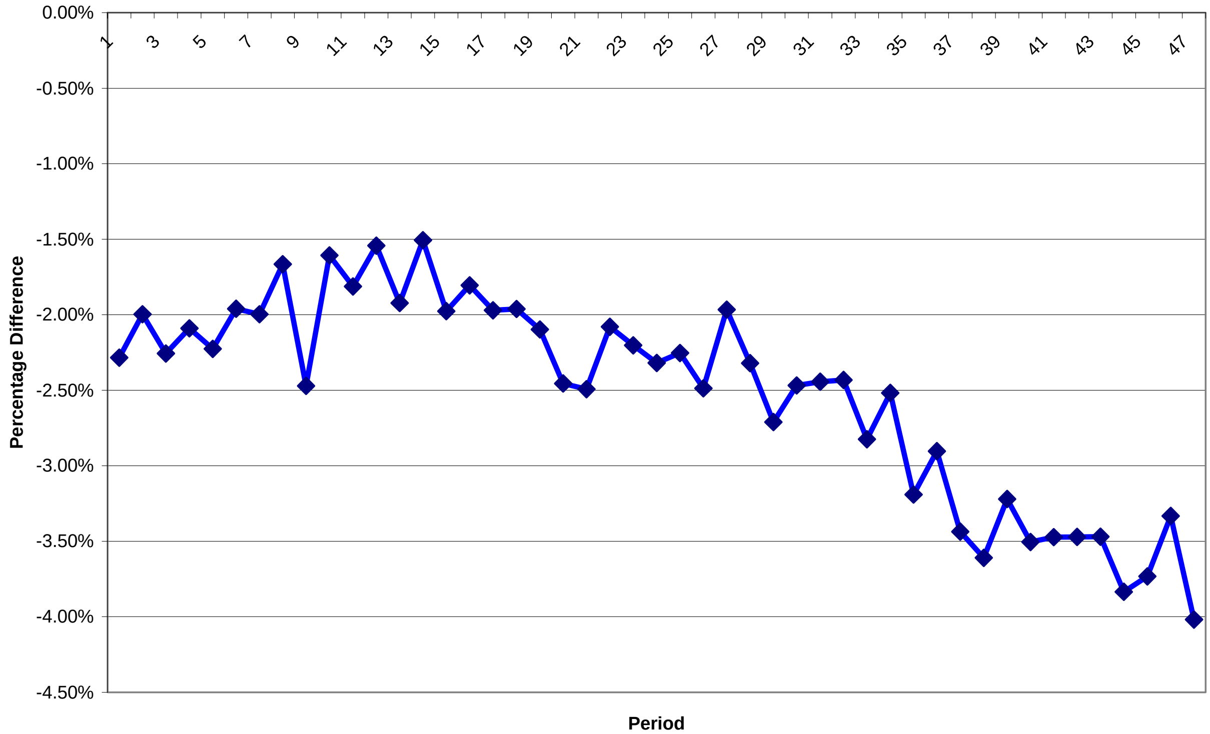
Product	**All Stores**
Apples	-0.1036
	(0.2298)

Apple Juice	-0.2769
	(0.3799)
Bananas	-0.1545
	(0.1747)
Bread	-0.0642
	(0.0898)
Butter/Margarine	-0.8192
	(0.2445)
Cereal	-0.1079
	(0.1275)
Chicken Breast	-0.5597
	(0.4402)
Coffee	-0.6548
	(0.4774)
Cookies	-0.4850
	(0.1294)
Eggs	-0.4324
	(0.0995)
Ground Beef	-0.0679
	(0.1637)
Ham	-1.3032
	(0.7580)
Ice Cream	-0.3516
	(0.3053)
Lettuce	-1.6194
	(1.0106)
Milk	-0.2411
	(0.0748)
Potatoes	-0.6406
	(0.2346)
Soda	-0.3756
	(0.1489)
Tomatoes	-0.8157
	(0.4942)
Bottled Water	-0.7231
	(0.9446)
Yogurt	-0.1832
	(0.1635)

All of the estimated elasticity coefficients are negative as expected. Thus as households spend increasing amounts of expenditure at SMCs, the average prices paid for food items decrease. While the effects are estimated with varying amount of precision, overall the results are highly significantly different from zero. No obvious pattern of coefficient size seems to exist: we find the largest effects for ham, lettuce, butter/margarine, tomatoes, potatoes, and coffee, which are a mix of branded and unbranded products. Yet, we find relatively small effects for ground beef, apples and bananas, which are typically unbranded products, but we also find relatively small effects for cereal and yogurt, which typically are branded products. Overall, we find a statistically negative effect on average prices as shopping in superstores increases. Thus, we find the "total effect" operates as household shift their expenditure from traditional supermarkets to lower priced superstore outlets.

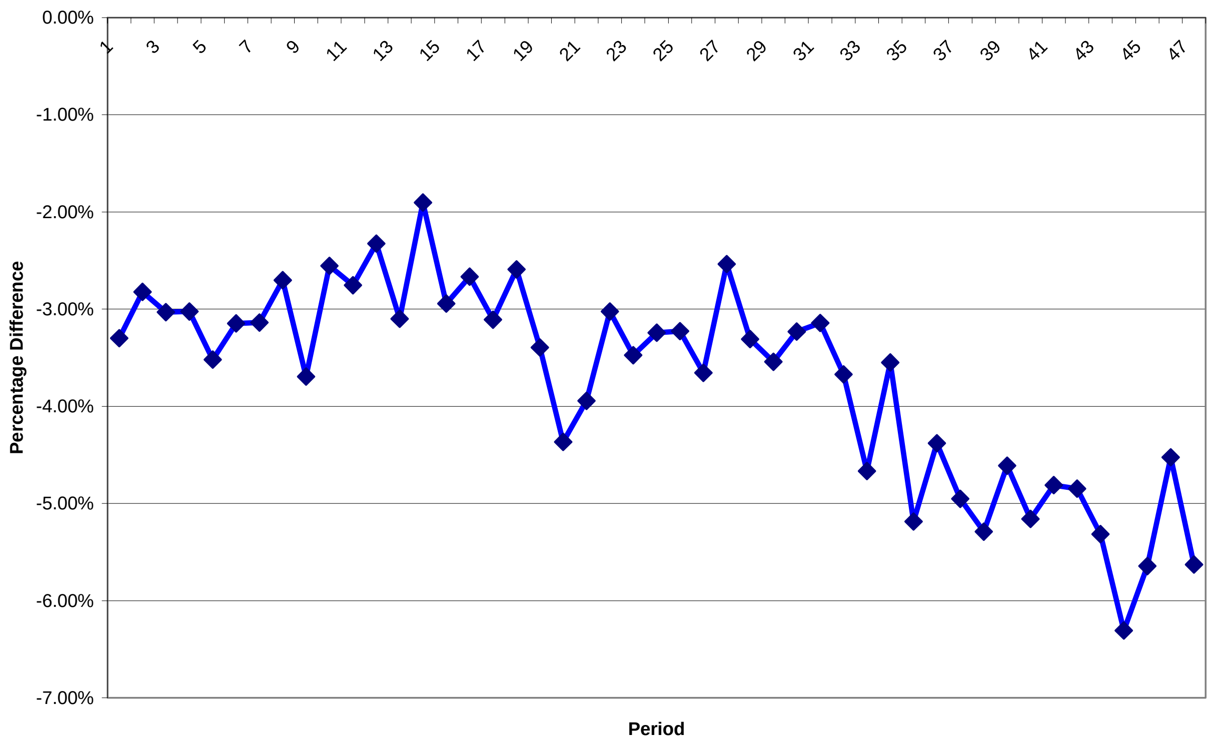
In Graph 5.1 we depict the difference in average prices paid by households due to the spread of SMC stores over the period. During the sample period from January 1998 to December 2001 the expenditure share of SMC stores increased from 10.9% to 16.9%, a 55.3% increase over the 48 months or 11.6% per year. We take the econometric estimates from Table 5.2 and use them to estimate the decrease in average price for each food category. We then average across food categories and plot the results in Graph 5.1, which demonstrates the increasing effect on average food prices as SMCs become more available and households increase their expenditures at these retail outlets. We find that food prices are 3.0% lower than otherwise, or an effect of about 0.75% year.

Graph 5.1: National Difference in Prices Due to SMC Stores

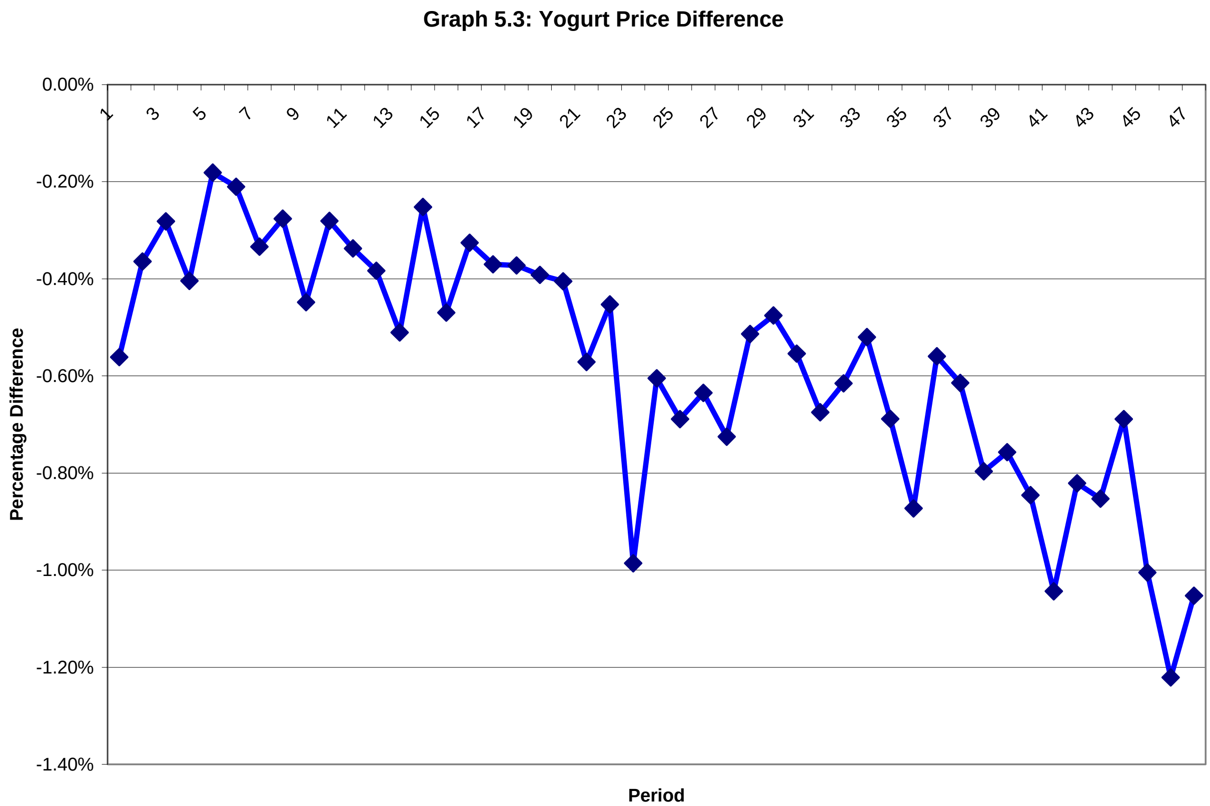


We now consider two of the individual food products. In Graph 5.2 we plot the effect of increased expenditure in superstores on the average price of butter/margarine.

Graph 5.2: Butter/Margarine Difference



The estimated coefficient for butter/margarine in Table 5.2 is quite large at –0.8192. The estimated effect of the spread of superstores on the price of butter/margarine is –5.63% over the 48 month period. The effect on the price of yogurt is presented in Graph 5.3. The estimated coefficient for the price of yogurt is considerably smaller at –0.1832.



Thus, in Graph 5.3 the effect on the average price of yogurt over the 48 month period is –1.1%. From Graphs 5.2 and 5.3 we see that significantly different price effects exist for different food products due to the spread of SMCs and increased expenditure at those SMCs by households.

We now repeat the econometrics to test for the "indirect effect" of lower conventional supermarket prices because of increased competition from superstores. In equation (5.1) we replace the left-hand variable p_{it}, which is the average log price paid for a given product, with $\tilde{p}_{it}$, which is the average price paid in supermarkets. We give the results in Table 5.3:

Table 5.3: Average Price for Food Products in Supermarkets across 34 Markets

National Results for Supermarkets

AR(1) IV Results

(Asymptotic Standard Errors)

Product	**Supermarkets**

Apples	-0.2307
	(0.2233)
Apple Juice	-0.5385
	(0.5104)
Bananas	-0.0437
	(0.1447)
Bread	0.0066
	(0.0890)
Butter/Margarine	-0.6853
	(0.2089)
Cereal	0.0832
	(0.1538)
Chicken Breast	-0.5812
	(0.5352)
Coffee	-0.4763
	(0.6005)
Cookies	-0.4366
	(0.1966)
Eggs	-0.1915
	(0.0922)
Ground Beef	-0.0303
	(0.1538)
Ham	-2.1172
	(1.2448)
Ice Cream	-0.3985
	(0.2895)
Lettuce	-2.4217
	(1.5517)
Milk	-0.1247
	(0.0887)
Potatoes	-0.5092
	(0.2244)
Soda	-0.2728
	(0.1513)
Tomatoes	-0.6956
	(0.4791)
Bottled Water	-0.5950
	(0.8155)
Yogurt	-0.0759
	(0.1833)

We estimate 18 of the 20 coefficients to be negative, with the only exceptions being bread and cereal, neither of which is statistically significant.[31] As would be expected from economic theory, the effects of increased SMC expenditures are smaller for most of the products. Thus, the estimated "total effects" on average prices paid by household arising from substitution to lower priced SMCs typically exceeds the "indirect effects" of decreased prices in supermarkets. Nevertheless, we do find some quite large indirect effects as in lettuce, butter/margarine, coffee, ice cream, potatoes, tomatoes, and bottled water. The spread of supercenters leads to lower prices both for households that shift their food shopping from supermarket to SMC stores but also for households who continue to shop at supermarkets because of lower prices caused by the increased competition from expanding food offerings at SMCs.

In principle we could decompose the total effect into an "indirect effect" and a "direct effect" by using a share weighed average of prices into supercenters and traditional supermarkets. However, the econometrics of this approach are difficult because prices and shares for a given product are not econometrically pre-determined. Further, expenditures in superstores for a given product are also unlikely to be pre-determined as we explained above. Thus, to estimate the decomposition we would need additional instrument variables, which we have been unable to determine.[32]

In terms of one of the questions we posed at the beginning of the paper, the spread of supercenters does significantly affect prices paid by households. However, to correctly estimate the effect both <u>quantities and prices</u> must be utilized. Holding prices fixed as households shift their expenditures to non-traditional retail outlets, we find the average prices they pay decrease. However, prices also change because as households shift their purchasing behavior, the increased competition forces supermarkets to lower their prices. Both of these effects, the direct effect and indirect effect, lead to lower average prices paid by households for food items.

[31] We find very similar results if we group the remaining Nielsen categories with supermarket: drug stores, convenience, and "other". These other outlet categories have relatively low expenditure levels compared to traditional supermarkets.
[32] The approach I used earlier, e.g. Hausman and Leonard (1994, 2002) of using supermarket prices in city as instruments for prices in another city does not work here since Wal-Mart has a common presence across all of the markets that we use in our econometric data set.

VI. Effect on Price Indices

Since our scanner based data set includes observation on both quantity and price, we are able to construct a price index that takes account of both increased expenditure at SMC stores as well as the effects of substitution when consumers face lower prices. Thus, we are to consider a source of first order bias in the CPI, outlet substitution bias, as well as the source of second order bias, substitution bias that occurs with the lower prices at the SMC outlets.

Food expenditures at SMC outlets have increased over the years in question. In January 1998, in our sample of 34 markets, we find an expenditure share of 0.1090. At the end of the sample, 48 months later, in December 2001 we find an expenditure share of 0.1693. Thus, the expenditure increased by .0603 or by 55.3% over the 48 months or 11.6% per year. The share has continued to increase as new SMC food outlets have continued to open and as consumers have increasingly shopped at these outlets.

We estimate the effect of this increased expenditure in lower priced SMC outlets on the 20 food categories we considered above and an overall food price index. We consider three indices in Table 6.1: (1) Continuous update: a continuously updated, value index where aggregates food expenditure shares across outlets from the current month are used to construct a share weighted average price for each food category. Note that since we have scanner data we can update both the food expenditure shares (quantity data) and the price data each month. This continuous updating allows us to control for both outlet substitution bias, a first order bias in the CPI, and substitution bias, a second order bias in the CPI. (2) BLS Constant Weights: we keep the expenditure shares constant over the 48 months. We use current prices each month, but we take a weighted average using the expenditure weights as of January 1998. Thus, both outlet substitution bias and price substitution bias are present in the calculated index. This index is probably closest to the current BLS approach, although the BLS uses geometric means while we use arithmetic means. (3) BLS with updated yearly expenditure weights: In January of each year we rotate stores and link the prices to the preceding December. We are assuming here that the BLS TPOPS procedure leads to a correctly reweighted sample each year, but that price linking removes the lower price effect of the shift by consumers to increasing expenditures at SMCs (4) Biennial Update: We now update the expenditure weights across stores based on the previous December. We continue to use the BLS linking procedure. Thus, we continue to have outlet substitution bias but we have reduced price substitution bias because of the yearly updates.

Table 6.1: Price Index Calculations for food Expenditure: 1998-2001

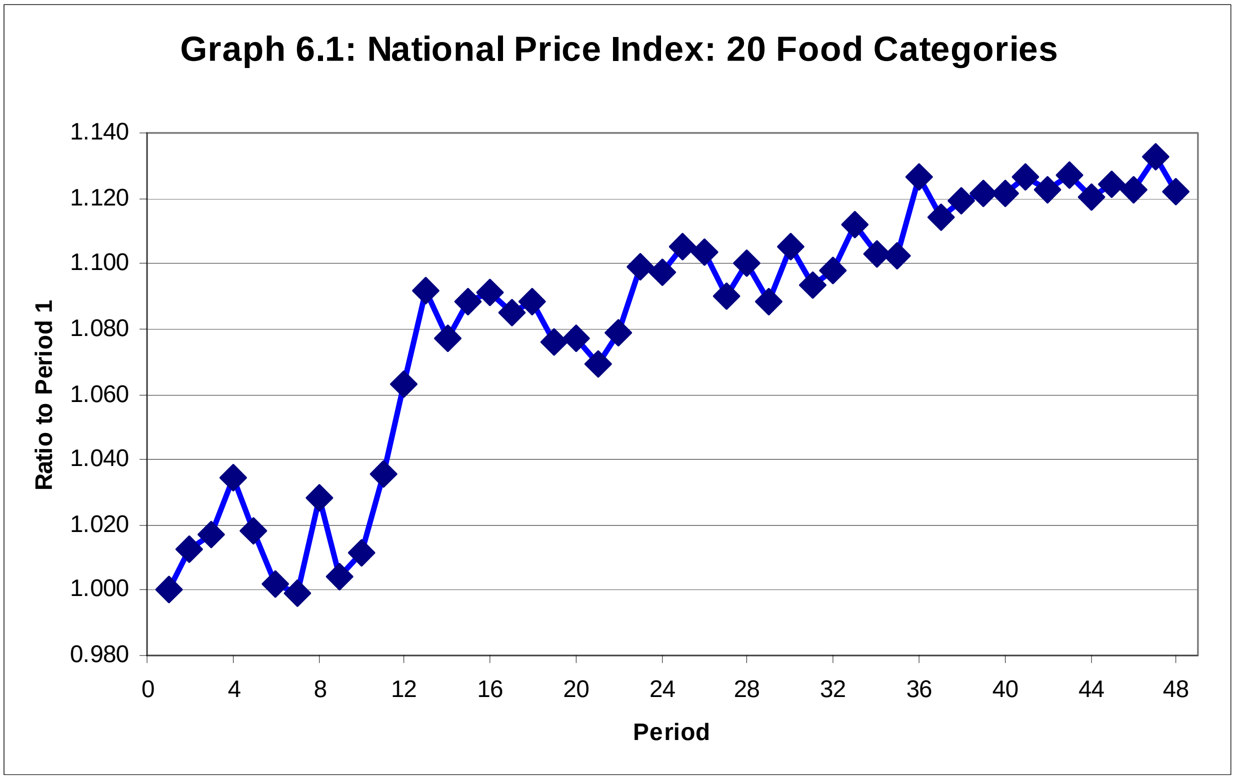
Product	**Continuous Update**	**Constant Weights**	**Yearly Update**	**Biennial Update**
Apples	1.016	1.028	1.032	1.032
Apple Juice	0.939	0.961	0.955	0.960
Bananas	0.710	0.720	0.717	0.725
Bread	1.104	1.106	1.104	1.111
Butter/Margarine	1.162	1.168	1.172	1.169
Cereal	1.043	1.054	1.051	1.056
Chicken Breast	1.731	1.765	1.768	1.762
Coffee	0.897	0.909	0.915	0.926
Cookies	1.148	1.156	1.157	1.157
Eggs	0.893	0.905	0.903	0.909
Ground Beef	1.368	1.392	1.392	1.388
Ham	0.755	0.774	0.791	0.799
Ice Cream	1.092	1.112	1.110	1.108
Lettuce	1.016	1.059	1.056	1.045
Milk	1.083	1.091	1.091	1.095
Potatoes	1.355	1.373	1.381	1.378
Soda	1.084	1.074	1.081	1.077
Tomatoes	1.569	1.581	1.582	1.599
Bottled Water	1.160	1.162	1.174	1.182
Yogurt	1.102	1.120	1.115	1.119
Average difference/year		0.0032	0.0036	0.0042

In Table 6.1 we see that Method (1), the Continuous Update procedure, almost always leads to lower price increases or greater price decreases for all food products over the 48 month period. For example, apples have a price increase estimated at 1.6%. Method (2) calculates an increase of 2.8%. The difference is 0.12 percentage points or a difference of 15.0% per year. Method (3), which allows for yearly updated expenditure weights, calculates an increase of 3.2% per year, 0.16 percentage points more than Method (1), or a difference of 18.9% per year. Lastly, Method (4), which uses biennial updates to the weights, again calculates an increase of 3.2% per year or a difference of 18.9% per year. To our initial surprise, while Method (1) finds the lowest price increase as expected, Method (2) often estimates a lower price increase than Method (3) or Method (4). However, we now recognize this outcome as the result of the BLS linking procedure that eliminates the effect of the lower prices when customers switch outlets. Method (2) captures the “indirect effect” of lower prices when the presence of supercenters increases, but Methods (3) and (4) eliminate part of this indirect effect because they update the expenditure weights. Thus, the outcome of a more continual updating of expenditure weights leads to a perverse result because of the “linking out” of lower prices in SMCs.

When we take average yearly changes across all food categories we find the estimated difference between Method (1) and Method (2) to be 0.32% a year. This estimate is the same order of magnitude, but somewhat higher than Reinsdorf’s (1993) estimate. In terms of the BLS CPI-U for food at home which averaged 2.29% over this period, the 0.32% per year difference is 14.0%. Thus, we estimate that the Method (2) has an upward bias of approximately 14.0% because of its linking procedure, which eliminates the effect of households shifting their expenditure to lower price supercenter outlets such as Wal-Mart.

We next compare Method (1) to Method (3), which allows for updated expenditure weights each year. Here we find an estimated difference between Method (1) and Method (3) of increase of about 0.36% per year. We find an upward bias in the Method (3) measure of food at home to be upward biased by 15.7%. If we compare Method (4), the biennial update method, we find the estimated average difference to be 0.42% per year. In terms of Method (4) the 0.42% per year difference is 18.3%. The years 1998-2001 were generally a period of low inflation, but we still find significant difference in estimates of the food price indices dues to shift towards lower price outlets. We find an upward bias in the range of 14.0% to 18.3% in the estimate of the CPI for food at home because of the use of the BLS linking procedure. Thus, updating the expenditure weights significantly reduces the bias in the estimated price index

In Graph 6.1 we plot the Method (1) price index where January 1998 is set equal to 1.0. Over the entire period we estimate a price increase of 12.1% or 2.5% per year.



.

While this estimate is for just the 20 food products we have investigated to date, we note that the BLS CPI-U food at home index increased by 9.48% over the same period or 2.29% per year. The estimates are quite comparable, but the CPI-U index is over a much wider range of food products than the index we have computed.

VII. Conclusion

Over the past 15 years the largest development in food retailing has been the start of Wal-Mart supercenters that compete most closely with traditional supermarkets. Wal-Mart has expanded greatly, mostly in the South and Southwest, and become the largest supermarket chain in the U.S. Wal-Mart is now expanding into additional geographic markets in California and the upper Midwest, so its effects will become even more important.[33] Wal-Mart offers identical food items at an average price about 15%-25% lower than traditional supermarkets. Wal-Mart's entry into a new geographic market creates a direct price effect by offering a lower price option to consumers and an indirect price effect by causing traditional supermarkets to lower their prices because of the increased competition.

The BLS procedure currently does not take account of the lower price option that Wal-Mart offers when it enters and expands in a given geographic market. The BLS only captures the indirect price effect. Instead, the BLS “links out” Wal-Mart’s lower prices by assuming that an exact “compensating service quality differential” exists that exactly counteracts Wal-Mart’s lower prices. If this assumption were correct, we would not see the rapid gain in market share by Wal-Mart after its entry into a market.

We find that a more appropriate approach to the analysis is to let the choice to shop at Wal-Mart be considered as a “new good” to consumers when Wal-Mart enters a geographic market. Some consumers continue to shop at traditional supermarkets while other consumers choose to shop at Wal-Mart. For the representative consumer we take a utility-consistent probability weighted average of the choice of shopping destination.[34] This approach leads to a continuously updated expenditure weighted average price calculation, which we apply to food data in 34 markets over a 48 month period. Of course, the BLS would need to implement our proposal using economic judgment because if the new market entrant were Starbucks, instead of Wal-Mart, the assumption that the higher price of coffee in Starbucks did not represent a quality differential would be difficult to justify. The difference arises because much of the food products sold by Wal-Mart are the same as the products sold by supermarkets, while in the case of Starbucks a quality adjustment is necessary as Hausman (2003) discusses.[35] The approach we recommend in this paper requires quantity data as well as price data, so the BLS would need to begin to use scanner data to implement our approach. Currently the BLS collects only price data, but does not collect quantity (or expenditure) data that it incorporates into the CPI except at lengthy intervals.

We find a significant difference between our approach and the BLS approach, even for the relatively low food inflation period of 1998-2001 that we study in this paper.

---

[33] Wal-Mart has announced plans to open 40 supercenters in California in the next 3-5 years, Wiltamuth op. cit.
[34] The BLS approach assumes that consumers are not made better off by an expanded choice set, contrary to almost all economic theory.
[35] However, if the higher quality aspect of Starbucks were correctly taken into account as Hausman (2003) discusses, the price index would typically decrease even though Starbuck coffee prices are higher than the previous coffee shop charged.

Our estimates are that the BLS CPI-U food at home inflation is too high by about 0.32 to 0.42 percentage points, which leads to an upward bias in the estimated inflation rate of about 15% per year. We intend to expand our approach to more food categories in further research, but we find that the BLS should take account of Wal-Mart and other non-traditional retail outlets, rather than making believe that Wal-Mart does not exist.

References

Abraham, K., J. Greenlees, and B. Moulton (1998), “Working to Improve the Consumer Price Index”, Journal of Economic Perspectives, 12

Cage, R., “New Methodology for Selecting CPI Outlet Samples,” Monthly Labor Review, December 1996, p. 49

Callahan, P. and A. Zimmerman, “Grocery Chains Fighting Wal-Mart for Market Share,” Wall Street Journal, May 31, 2003.

Hausman, J., "Specification Tests in Econometrics," Econometrica, 46, 1978

Hausman, J., "The Econometrics of Nonlinear Budget Sets," Econometrica, 53, 1985

Hausman, J., "Valuation of New Goods Under Perfect and Imperfect Competition," ed. T. Bresnahan and R. Gordon, The Economics of New Goods, University of Chicago Press, 1997.

Hausman, J., "Cellular Telephone, New Products and the CPI," Journal of Business and Economics Statistics, 1999.

Hausman, J., “Sources of Bias and Solutions to Bias in the CPI”, Journal of Economic Perspectives, 2003

Hausman, J., G. Leonard, and D. Zona, "Competitive Analysis with Differentiated Products," Annales, D'Economie et de Statistique, 34, 1994

Hausman, J., G. Leonard, and D. McFadden, "A Utility-Consistent Combined Discrete Choice and Count Data Model: Assessing Recreational Use Losses Due to Natural Resource Damage," Journal of Public Economics, 56, 1995.

Hausman, J. and G. Leonard, “The Competitive Effects of a New Product Introduction: A Case Study,” with G. Leonard, Journal of Industrial Economics, 50, 2002

Hausman, J. and E. Leibtag, “Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart,” MIT mimeo, revised May 2005.

Little, P. “Channel Blurring Redefines the Grocery Market,” Competitive Edge, June 2004.

Reinsdorf, M., "The Effect of Outlet Price differentials in the U.S. Consumer Price Index," in M.F. Foss et. al. eds, Price Measurements and their Use, Chicago; University of Chicago Press, 1993.

Schultze C. and C. Mackie, eds.; At What Price? , Washington: Nation Academy of Sciences Press, 2002.

Shapiro, M. and D. Wilcox, "Mismeasurement of the Consumer Price Index: An Evaluation," NBER Macroeconomics Annual, 1996.

Appendix A: Reply to Professor Silver

Prof. Silver correctly draws attention to the questions of sample selection and composition for the ACNeilsen Homescan panel data that we use. The use of panel data seems especially attractive since households remain the same when the presence of supercenter outlets increases in a given market. Indeed, in a related paper, Hausman and Leibtag (HL, 2005), we use econometric methods to estimate fixed effects for households, which is only possible with panel data. Hausman (1978) demonstrated that unobserved household effects that are correlated with observed variables can create significant bias in econometric and statistical analysis. Prof. Silver raises the problem of sample selection of more price sensitive households can lead to an estimate upward bias in effects. As we describe in the current paper Nielsen chooses household using random sampling techniques. Prof. Silver speculates that households that agree to join the Homescan panel "are likely [to] have lower search costs, be better informed about prices and be more price sensitive." He gives no reasons for this conclusion. However, we understand that Nielsen, and its competitor IRI, both use household panels to supplement their point of sale (POS) scanner data, which Prof. Silver views as nearly problem-free. Thus, neither the companies nor their customers find the panel data to suffer from unacceptable bias.[36]

Prof. Silver also discusses attrition and sample replacement and the possible problems that may arise with non-random attrition. Again Prof. Silver speculates that only the more "price sensitive" households remain in the sample. He correctly states that non-random attrition "can potentially create estimation problems for researchers." Hausman and Wise (1979) were among the first authors to demonstrate the problems that can arise from non-random attrition in panel data. Prof. Silver is also correct that "biased sample selection comes around in biased results." He states that whether the problem exists is "difficult to test."

To investigate the possible effect of non-random attrition we calculated the proportion of shopping trips to supercenters by 3 groups. Group 1 is households who exit in the sample in year1; Group 2 exit in year 2; Group 3 exit in year 3; and Group 4 remain in the sample for the entire sample period. If non-random attrition created a problem, we would expect to see the proportion of trips to supercenters be greater the higher the group number since household exit

[36] Hausman has used POS scanner data in many previous papers. However, since Wal-Mart no longer sells its scanner data to either Neilsen or IRI, we used panel data in this paper to do the estimation. If the BLS were to use scanner data to estimate its CPI correctly as we suggest, hopefully it could buy the necessary scanner data from Wal-Mart.

the sample earlier may not be as "price sensitive" as other groups. The proportions and standard errors (S.E.) are given in Table A1.

Table A1: Supercenter Shopping Proportions

	Group 1		Group 2		Group 3		Group 4	
Period	**Proportion**	**S.E.**	**Proportion**	**S.E.**	**Proportion**	**S.E.**	**Proportion**	**S.E.**
3	0.151	0.009	0.143	0.012	0.139	0.013	0.136	0.008
12	0.184	0.010	0.149	0.013	0.154	0.014	0.147	0.008
24			0.160	0.011	0.173	0.013	0.166	0.008
36					0.173	0.010	0.173	0.009
48							0.191	0.009

Table A1 calculates the proportion of supercenter shopping trips across all market at 3 month intervals. We do not find evidence that households that exit the sample earlier are more likely to shop at supercenters. For example in month 3 of the sample Group1 shops more at supercenters than Groups 2-4. In period 12 Group 3 shops more at supercenters that Group 2 but also more that Group 4, households who never exit the sample. If non-random attrition were a problem we would expect to see the proportions increase from left to right in each period. We do not observe this pattern. Further, the estimated standard errors are sufficiently small that we would very likely have found the pattern if it existed in the population.

While the question of a non-random sample cannot be answered definitively since it depends on unobserved household characteristics, these estimate and tests demonstrates that it is unlikely to be particularly large. Further, the main recommendation of our paper that the BLS not ignore the price differences of Wal-Mart and other supercenters compared to traditional outlets when they enter the BLS sample remains unchanged.

References

Hausman, J., "Specification Tests in Econometrics," Econometrica, 46, 1978

Hausman J. and E. Leibtag, “Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart,” MIT mimeo, revised May 2005.

Hausman J. and D. Wise, "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," Econometrica, 47, 1979