ABSTRACT

BALL, MELANIE JEAN. Measuring the Effect of Supercenters on Price Indexes: Evidence from Canned Beans and Milk Using Household Scanner Data. (Under the direction of Dr. Walter Thurman).

U.S. government agencies that administer federal assistance programs use the Consumer Price Index (CPI) to determine benefit amounts and income eligibility requirements for those programs. To best estimate the amount by which the cost of living has changed, federal agencies must rely on the accuracy of the CPI, which is published by the Bureau of Labor Statistics (BLS). There is evidence that BLS substantially underestimates the presence of supercenters (Hausman and Leibtag 2009; Griffith, et. al. 2009). This is relevant because research has shown that when supercenters enter new marketplaces, consumers tend to switch outlets from traditional grocery stores to the lower-priced supercenters (Arnold, et. al. 1998; Singh, Hansen, and Blattberg 2006). Previous studies have used scanner data to examine the outlet-substitution bias that may exist in BLS's calculation of the CPI but few studies have examined the difference in price indexes caused by increased supercenter presence. In some geographic markets, supercenters have little to no presence. In other geographic markets, supercenters have a substantial market share for food products. This study employs household scanner data from the 2004-2006 Nielsen Homescan panel to econometrically examine the geographic market-level effect of supercenters on the price index for food using the proportion of household transactions at supercenters and the price index of two goods - canned beans and milk. The house price index is used as an instrumental variable to control for endogeneity in supercenter location. This study found a significant causal effect of supercenters on the price index for canned beans; however, there was no evidence of an effect of supercenters on the price index for milk. Results from this study may be used to measure the extent to which supercenters affect the price

index for food and whether the effect differs by the type of food (e.g. perishable versus non-perishable items).

Measuring the Effect of Supercenters on Price Indexes: Evidence from Canned Beans and Milk Using Household Scanner Data

by
Melanie Jean Ball

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Economics

Raleigh, North Carolina

2012

APPROVED BY:

Ivan Kandilov	Xiaoyong Zheng

Walter Thurman
Committee Chair

DEDICATION

This is dedicated to Bryce. He is my best friend, my soul-mate, my rock, and my inspiration.

BIOGRAPHY

Melanie J. Ball (neé Smith) is a native of Elon, North Carolina, currently residing in Mebane, North Carolina with her husband, Bryce. Much more than thesis writing, she enjoys gardening, running, cooking, and spending time with friends. She looks forward to doing all of those things more once this paper is finished and filed.

ACKNOWLEDGMENTS

The author would like to thank her thesis adviser, Dr. Walter Thurman, for his guidance, support, and patience through this process. Additionally, the author would like to thank Dr. Chen Zhen who provided invaluable guidance, feedback, and perspective on all things price indexes. Last, but certainly not least, the author would like to thank her husband, Bryce, without whom she most surely would have lost perspective and sanity during her graduate school tenure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

Acronym or Abbreviation	Description
CPI	Consumer Price Index
COLI	Cost-of-Living Index
BLS	U.S. Department of Labor’s Bureau of Labor Statistics
CEX	Consumer Expenditure Survey
SNAP	Supplemental Nutrition Assistance Program
ERS	U.S. Department of Agriculture’s Economic Research Service
TPOPS	Telephone Point-of-Purchase Survey
ELI	Entry-Level Item
POPS	Point-of-Purchase Survey
PSU	Primary Sampling Unit
UPC	Universal Product Code
FE	Fixed-Effects
AR (1)	First-Order Autoregression
IV	Instrumental Variable
MSA	Metropolitan Statistical Area
HPI	House Price Index

1. Introduction

The Consumer Price Index (CPI) is calculated by the Bureau of Labor Statistics (BLS) to measure the change in prices paid for a given basket of goods and services. The CPI is frequently used as an economic indicator to adjust for income eligibility requirements and to adjust government assistance benefit amounts based on changes in the estimated cost of living. Currently, the CPI is used in determining the income eligibility and the level of benefits disbursed through various federal assistance programs, such as the Supplemental Nutrition Assistance Program (SNAP, formerly the Food Stamp Program). Although a substantial amount of research has been conducted testing, validating, and supporting BLS's current methods of calculating the CPI (Shultz and Mackie 2002) and the CPI's use as a tool for measuring changes in the costs of living, several studies have identified bias in the way the CPI is calculated (Hausman and Leibtag 2009; Griffith et al 2009). The method by which a market basket of goods is determined and used in the CPI may be biased in a number of ways, succinctly identified by Hausman and Leibtag who noted the following:

- Substitution bias – When the price of a product increases substantially, consumers switch to a lower-priced substitute.
- New goods bias – New products in the marketplace are not immediately added to the CPI. For instance, cell phones were not added to the CPI for 15 years after they were introduced.
- Quality bias – Product advancements over time are not factored into the CPI. An example of this is technological advancements which lengthen the useful life of a product.
- Outlet substitution bias - Shifts in shopping patterns to low-cost retail or online outlets are not adequately represented in the CPI.

Researchers at USDA's Economic Research Service (ERS) and BLS have also indicated that the CPI for food at home overstates the amount of inflation in food expenditures because of the frequency with which the market basket and store selection are updated (see ERS 2007 and MacDonald 1995). To help correct for substitution bias, the BLS, starting in 2002, changed the frequency by which the market basket of goods is updated from 10 years to 2 years (ERS 2007), making the inflation measure more relevant to current consumption patterns. However, this change does not adequately address shifts in shopping patterns in the short term, such as the shift away from food purchased at traditional grocery stores and supermarkets to food purchased at supercenters (Hausman and Leibtag 2009).

For the past decade, household food purchasing patterns have shifted substantially. In many markets, households have shifted away from purchasing food at traditional grocery stores and supermarkets and shifted towards purchasing food from supercenters, mass merchandisers and club stores. For example, in 2009, food sales at supercenters and club stores accounted for 18.1% of all U.S. retail food sales compared to 5.9% in 1999. It is estimated that prices of goods sold at supercenters are 5-11% lower than prices of comparable goods sold at traditional grocery stores and supermarkets, depending on market share (ERS 2010). However, it is unclear whether these food products' price differences have an effect on the food price indexes in markets where supercenters have a large market share.

Previous studies have shown that BLS currently uses a method of outlet selection that significantly underestimates the probability of selecting newly constructed, lower-priced supercenters from which to collect price information (Hausman and Leibtag 2009). However, to date, such studies have not examined the significance of a supercenter effect on the price index for food. Accordingly, this study seeks to measure the effect of supercenter prices on the price index using household scanner data for purchases of two products: milk, which is perishable and non-storable, and canned beans, which are non-perishable and easily storable. To measure this effect, this study examines the relationship between the proportion of products sold at supercenters and the difference in price indexes excluding and including supercenter prices on milk and canned beans. If the effect of supercenters on price indexes is substantial, as evidenced by these items, it may indicate that the CPI for consumers in areas where supercenters are prominent is overstated. Is there evidence that price indexes decrease as consumers switch to supercenters for food purchases? Is the supercenter effect of supercenters on prices more pronounced for products that are non-perishable?

This paper is organized as follows: Section 2 briefly describes BLS's current methods for calculating the CPI and research conducted testing those methods; Section 3 examines different measures of calculating the cost of living index; Section 4 outlines the data and methods used in this study; Section 5 discusses the results of the study; Section 6 briefly describes some limitations to the study; and, Section 7 discusses the implications of the study's results.

2. Review of the Consumer Price Index and Relevant Research on Gaps and Improvements

The current method of constructing the CPI employs several rounds of survey-based data collection from consumers and retailers. These survey results are then used to develop basic (item-level) and aggregate price indexes. Both the accuracy of the results and the feasibility of this approach are well researched and documented. The BLS has also taken measures in recent years to update the CPI formula to incorporate innovative methods to improve its accuracy. These updates include increasing the frequency of data collection as well as revising the calculation of price ratios. This section summarizes the current BLS methodology, based on information published in Chapter 17 of

BLS's Handbook of Methods, and discusses research on the gaps inherent in BLS methods and potential improvements. A more comprehensive description of BLS's price index methodology, including equations, can be found in Appendix A.

2.1 Current CPI Methodology

The CPI is calculated using data from two consumer surveys and one retailer survey. The first data source is the Consumer Expenditure Survey (CEX), which is conducted every two years by the U.S. Census Bureau. CEX results are used to update data on the items purchased by households, including determining the relative importance of household items, and to derive new cost weights for the market baskets. This collection frequency enables BLS to update the basket of goods biennially; however, it does not provide the requisite information to adjust the quantity of goods purchased. Nor does it provide information on how much households paid for individual items. For example, CEX participants are asked to document items and expenditures purchased on a specific date, as indicated by BLS, on a form which looks like a diary. Data collected on participants' purchases captures basic item information. For example, if a participant in the New York City area purchased two quarts of organic whole milk for $3.99 per quart on the date for which records of purchases are requested by BLS, the participant only records that they purchased milk (without the specific type, quantity, or container size) for $7.98. BLS then determines the quantity of milk purchased implicitly, once price data is gathered.

The second survey is the Telephone Point-of-Purchase Survey, or TPOPS, conducted by the U.S Census Bureau which seeks to identify where consumers purchase certain products. Borrowing from the prior milk example, after BLS identifies milk as an item in the market basket through the CEX survey, it then selects a TPOPS participant sampled from the New York City area to inquire whether they purchased milk during the a certain period (or, the recall period); in this case is the previous two weeks. If the answer is affirmative, the respondent is then asked to identify the outlet from which the milk was purchased, in addition to how much the respondent spent on the item at that outlet during the recall period. However, expenditure data collected through TPOPS is not used directly in the CPI calculation. Instead, BLS utilizes the reported outlets and expenditure amounts to determine the probability of outlet selection for gathering market prices.

After market basket items are determined (CEX) and a list of retail outlets is compiled (TPOPS), BLS contacts outlets directly to collect price data for specific items. For example, assume that BLS determines, using TPOPS data, that in New York City, milk is typically purchased from The Food Emporium on 8th Avenue in Midtown Manhattan. Also assume that BLS selects this outlet for price data collection. BLS then contacts that specific The Food Emporium location to determine which milk item is most representative (this process is explained in a later section). For this example, assume that BLS determines that private-label (also called store-brand), conventional, fat-free milk in one-gallon containers (price of $4.99) is the most representative milk product. For the next four years, BLS will collect monthly price data for this specific item at that specific The Food Emporium location. Once price data are collected, BLS then derives the quantity purchased using the milk expenditures from the CEX ($7.98) and the price information collected from The Food Emporium for a one-gallon container ($4.99). Per this example, BLS determines that the consumer purchased 1.6 gallons of milk, when, in fact, the consumer purchased 2 quarts of milk, or 0.5 gallon.

The remainder of this section further explains the procedures employed by BLS to determine item selection, outlet selection, and price selection, and how these elements are used to determine the CPI – focusing primarily on the methodology used to determine the CPI for food items[1].

2.1.1 Item selection

As mentioned previously, items selected for inclusion in the CPI are derived from responses to the CEX survey. Items are first grouped into 8 major categories. These categories include food and beverages, house, recreation, education and communication, medical commodities, apparel, transportation, and other goods and services and these categories are segmented into a total of 80 expenditure classes. An example of group segmentation and item classification is depicted in Figure 1. For example, the major category "Food and Beverages" is split into the following expenditure classes:

- Cereals and Bakery Products
- Meat, Fish, Poultry and Eggs
- Dairy and Related Products
- Fruits and Vegetables
- Nonalcoholic Beverages and Beverage Materials
- Other Food at Home
- Food Away From Home

[1] Prices and quantities for items used in the CPI such as rent and house, medical services, and vehicles are collected differently than the majority of items used in CPI. Because this study focuses on the prices for food products, mechanisms for collecting price information on non-food products is ignored.

- Alcoholic Beverages

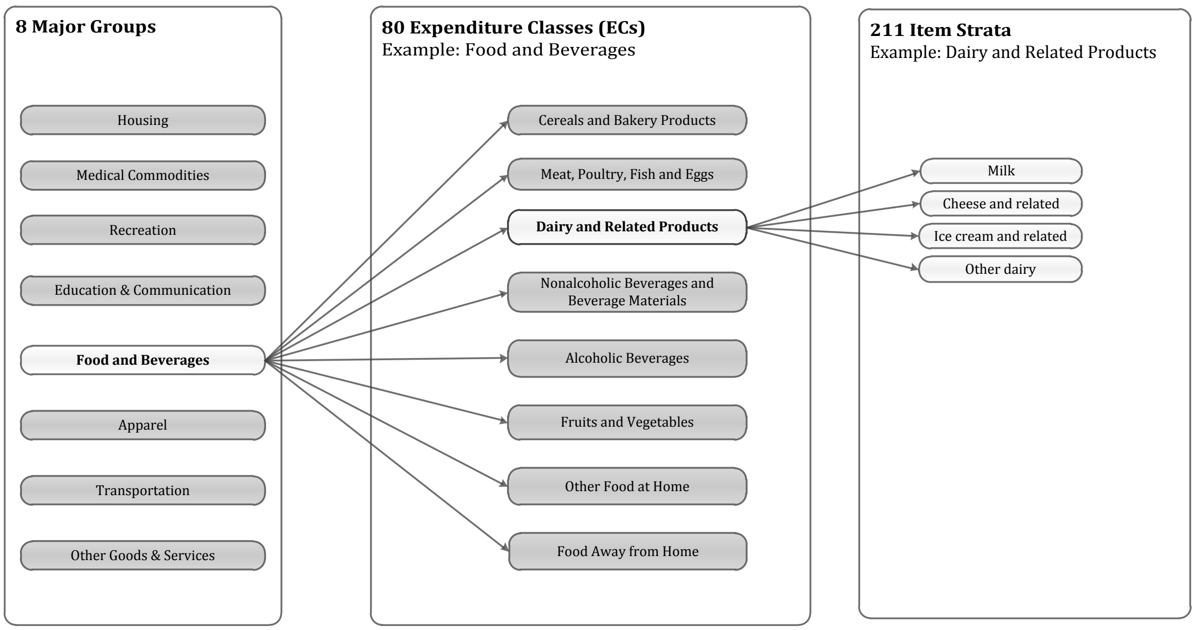

Figure 1 Item Grouping for Items in the CPI: Dairy and Related Products

Each expenditure class is then split into item strata. There are 211 total item strata defined by BLS. For example, as shown in Figure 1, the expenditure class “Dairy and Related Products” includes the following item strata:

- Milk
- Cheese and related products
- Ice cream and related products
- Other dairy

Within each item stratum, one or more substrata, called entry-level items (ELIs), are defined. In the dairy and related products example above, “Milk” is split into two ELIs - Whole milk and Other than Whole Milk. However, for the other item strata within Dairy and Related Products, there is exactly one ELI. In total, there are 305 ELIs that are the ultimate sampling units for items included in the CPI. This is the level at which data collectors for BLS collect price information and product descriptions. Products that represent each ELI are selected separately for each of the 87 primary sampling units (PSUs), which are primarily geographically defined (e.g. New York City and San Francisco), and rotated every 4 years. Expenditure weights are determined at the PSU-item strata level whereby the weights correspond to the probability of selection by CEX participants.

2.1.2 Outlet selection

Outlets selected for price data collection are chosen based on information collected through the TPOPS conducted by the U.S. Census Bureau. TPOPS captures the name and location of outlets at which survey participants purchase items in 214 broad Point-of-Purchase Survey (POPS) categories. These categories separated into two groups – 1) commodities and services, which consist of 209 categories; and, 2) rent and owners’ equivalent rent, which accounts for the remaining 5 categories. The objective of the commodities and services sample design is to group ELIs by PSU according to how consumers group the items they purchase at the same outlet. Many food items are purchased from the same outlet and in the same shopping trip. Accordingly, BLS groups food-at-home products into three POPS categories: nonmeat staples; meat, poultry, and fish; and, fruits and vegetables.

TPOPS employs random-digit dialing to select survey participants. Participants are asked about household makeup and demographics to determine whether the sample is representative of the PSU population. Next, the Census Bureau collects data on purchases for sixteen POPS categories within each PSU. POPS categories are rotated by PSU quarterly such that it takes four years to administer surveys for all POPS categories for a given PSU. The survey asks the participant whether they had expenditures in each POPS category included in the questionnaire during the recall period. The recall period ranges from 1 week to 5 years depending on the nature of the products included in the POPS category. Once the participant has confirmed expenditures in a given POPS category, data is collected about the outlet name and location where the respective items were purchased and the amount spent on items in that category. The reported outlets collected comprise the sampling frame from which to collect price data. Expenditure amounts collected in this step then are used to determine the probability of outlet selection for price data collection.

Once a sampling frame of outlets for each POPS category is obtained, outlets are weighted according to the expenditures reported in TPOPS. Outlet selection probability weights are edited such that no outlet has more than a 25% chance of being selected for price data collection for each ELI. Outlets and item samples are then merged to determine which outlets will be sampled to gather price information for products in each item strata.

2.1.3 Price selection

Price data are collected for each ELI based on the outlets selected through TPOPS. BLS employs a multi-stage probability selection technique to select the representative item for each ELI in a PSU among all items that fit into a specific description. The probability of item selection is generally proportional to the outlet's sales of the items included in each ELI group. Depending on the availability of store sales information, BLS determines the proportions of sales based on the following (in order of preference):

1. Obtaining sales proportions from the outlet's respondents

2. Ranking items by importance of sales as indicated by the outlet's respondent

3. Using shelf space to determine an item's relative importance to other items in the ELI; and,

4. Using equal probability

Based on the assigned probability weights, a specific item is randomly selected to be the representative item for a given ELI. BLS describes the unique item on a checklist in order to ensure that the item included does in fact belong to the given ELI and then obtains a price for the representative item. During subsequent months, until the sample is refreshed in TPOPS, BLS obtains prices for the exact item from the same outlet.

Examples of Price Data Collection Exceptions

In some cases, the price reported by the outlet is not the actual price paid by its customers. Circumstances in which the price recorded by BLS deviates from the "regular" price posted by the outlet include the following:

- Cents-off coupon – These are included in the price when the coupon is readily available to all customers, such as a sticker on the product or on a shelf near the item.

- Concessions – A concession is a deduction of a specific amount from the proposed selling price. BLS uses an average concession from a 30-day period in order to determine a product's price in this circumstance.

- Different-day pricing – If an item is not currently available but was in the previous seven days, the price from one of the previous seven days is collected. An example of this is fresh fish, where item selection varies from day to day.

- Discount price – These are prices experienced by customers through a loyalty or grocery-card program. BLS includes the discount price of products if the discounted price is realized by at least 50% of the sales for that product.

- Quantity discounts – Unit prices for items may decrease as quantity increases. In these cases, BLS collects data on the first multiple-unit price. For example, one ear of corn sells for $0.25, 3 ears sell for $0.60 and 7 ears sell for $1.00. BLS collects data for 3 ears of corn at $0.60.

For items that are no longer available for pricing (e.g. when the specific outlet no longer sells the specific item originally selected for price data collection), BLS uses a replacement procedure based on a previously defined checklist. This procedure seeks to ensure that the replacement item selected is closest to the item previously priced. However, in the event the replacement item is substantially different, the replacement item's price is adjusted for quality based on BLS's discretion.

2.2 Review of Previous Literature

Previous research has examined the feasibility of incorporating store and household scanner data into the price index calculations with mixed results. High frequency scanner data, or data with many billions of observations, require a substantial amount of computing resources. Accordingly, items with the same price-identifying characteristics must be aggregated (Reinsdorf 1999). Schultz and Mackie (2002) highlighted many of the advantages conferred through the use of point-of-sale scanner data in price index calculations. They also recommended that an evaluation of the accuracy of using the CEX to establish expenditure weights is needed. Reinsdorf (1999) argued that the BLS method of estimating quantity weights based on item sampling introduces errors that could be mitigated by using scanner data given that scanner data is able to capture actual elasticities of substitution whereas the BLS method assumes no substitution among products. Further, Hausman and Leibtag (2009) explored bias caused by outlet substitution in which newer outlets may take up to four years to be fully integrated into the TPOPS outlet sampling frame.

Hausman and Leibtag (2009) estimated outlet substitution bias for various food products using price and quantity data from the 1998-2001 Nielsen Homescan panel. The average prices calculated enabled them to consider both the effect of increased expenditures at supercenters, mass merchandisers, and club stores as well as the substitution effects experienced when consumers face lower prices from increased competition. Hausman and Leibtag noted that TPOPS outlet selection increases the probability of a supercenter, such as Wal-Mart, being included as more consumers shop there. However, they also noted that problems arise when identical goods are purchased at both a newly constructed supercenter and an established traditional grocery store within the same market. For example, assume that a new store, such as a supercenter, opens in a market selected by BLS but misses the window to be eligible for inclusion in the outlet sample determined by TPOPS. The same good may be cheaper at the newer supercenter, but due to lags in updates to the BLS retail outlet sampling frame, supercenter prices are excluded and BLS instead matches the item with a higher price at a traditional grocery store. This exemplifies outlet substitution bias in the CPI.

In 1999, Reinsdorf explored the feasibility of using scanner data in constructing price indexes using store-level coffee transactions in two markets, Chicago and Washington D.C. Reinsdorf found that the sheer volume of data, which exceeded 1.8 million records, suggested that a sampling of products may prove most feasible for estimating prices. Although Universal Product Codes (UPCs) are unique to a product, a product may have multiple UPCs. Accordingly, Reinsdorf aggregated units with identical price-determining characteristics (e.g. brand, size, packaging, and flavor) and demonstrated index bias when items are not aggregated per price-determining characteristics. Bias arises when manufacturers alter characteristics of the product, such as decreasing the product's size, thereby requiring a new UPC and increasing the unit value of the product. Similarly, in a prior study, Diewert (1995) argued for the need to use unit values to combine highly disaggregated data.

Ivancic, Diewert, and Fox (2011) also researched price index construction using high-frequency scanner data. Price and quantity shifts in products, which may occur when competing stores engage in "price wars," may distort price indexes that are constructed from store-level scanner data. This distortion occurs when shocks to price induce quantity shifts which result in sharp oscillations in price indexes. When these price indexes are averaged to develop a chained index, the sharp price and quantity fluctuations cause the chained indexes to fail the time-reversibility test – a desirable property of certain index numbers, which is explained later in this text. Ivancic, Diewert, and Fox proposed using the GEKS index formula (developed by Gini 1931; Elteto and Koves 1964; and, Szulc 1964) and chained Fisher's indexes to construct multilateral indexes that can be compared across geographic areas and time. This and other methods of price index construction are briefly described in a later section.

3. Means and Cost-of-Living Indexes: A General Discussion

Measures for comparing prices of goods can be calculated via multiple approaches. The simplest measures are the arithmetic mean and the arithmetic mean weighted by quantity (or, share-weighted mean). Although arithmetic price indexes may give researchers and policy makers an idea for the direction of price differences (e.g. a ratio greater than one indicates an increase in the current period relative to the base period), they are considered less desirable for measuring the magnitude of changes in prices. Geometric means are more desirable because they possess a time reversibility property (Diewert 1998). This section addresses the time reversibility property and demonstrates how simple averages, or arithmetic means, fail to exhibit it. Second, in the context of developing a cost-of-living index, it describes how these methods of averaging prices are used in constructing price indexes.

3.1 Arithmetic Means and Geometric Means: The Time Reversal Test

To calculate a price index between two commodities over time, it is useful to determine how much the price of one commodity changes over time relative to the other. To do so, one must examine the ratio of prices over two time periods. Arithmetic means are simple averages of numbers, calculated by taking the sum of a series of numbers and dividing by the number of observations. Although arithmetic means are used routinely for describing characteristics of a samples (e.g. demographics), they are less desirable when describing a relationship of characteristics over time, such as the difference between prices in one time period and another time period. Geometric means, or means calculated by taking the nth root of the product of a series of n numbers, are preferred (Diewert 1997). This is because arithmetic means fail, and geometric means pass, the time reversibility test. In order for an average price to satisfy the time reversibility test, the value of the average of the ratios of prices must equal the ratio of the average prices. For average prices calculated using arithmetic means, the ratio of average prices between time periods 1 and t is calculated:

$$P_{1t}^{RA} = \frac{\frac{1}{N}\sum_{i=1}^{N} p_{it}}{\frac{1}{N}\sum_{i=1}^{N} p_{i1}} \tag{1}$$

Using arithmetic means, the average of price ratios between time periods 1 and t is calculated:

$$P_{1t}^{AR} = \frac{1}{N}\sum_{i=1}^{N}\frac{p_{it}}{p_{i1}} \tag{2}$$

Take, for example, the calculation of the average change in the price for products A and B from time periods 1 to 2 using arithmetic means. The objective is to calculate the amount by which prices for A and B changed from period 1 to period 2. Using Equation (1), the ratio of the average prices of A and B is calculated:

$$P_{12}^{RA} = \frac{\frac{1}{N}(p_{A2} + p_{B2})}{\frac{1}{N}(p_{A1} + p_{B1})}$$

$$= \frac{(p_{A2} + p_{B2})}{(p_{A1} + p_{B1})} \tag{3}$$

$$= \frac{p_{A2}}{(p_{A1} + p_{B1})} + \frac{p_{B2}}{(p_{A1} + p_{B1})}$$

Now, using Equation (2), the average of the price ratios of A and B is calculated:

$$P_{12}^{AR} = \frac{1}{N}\left(\frac{p_{A2}}{p_{A1}} + \frac{p_{B2}}{p_{B1}}\right)$$

$$= \frac{p_{A2}}{N * p_{A1}} + \frac{p_{B2}}{N * p_{B1}} \tag{4}$$

In order to pass the time reversal test, the following must be true:

$$P_{12}^{AR} = P_{12}^{RA}$$

$$\frac{p_{A2}}{N * p_{A1}} + \frac{p_{B2}}{N * p_{B1}} = \frac{p_{A2}}{(p_{A1} + p_{B1})} + \frac{p_{B2}}{(p_{A1} + p_{B1})} \tag{5}$$

The only instance in which the above is true is where $p_{A1} = p_{B1}$, which although possible, is unlikely. This becomes less likely as more products are included in the calculation. In contrast, using geometric means, the ratio of average prices between time periods 1 and t is calculated:

$$P_{1t}^{RA} = \frac{\prod_{i=1}^{N}(p_{it})^{1/N}}{\prod_{i=1}^{N}(p_{i1})^{1/N}} \tag{6}$$

Using geometric means, the average of price ratios between time periods 1 and t is calculated:

$$P_{1t}^{AR} = \prod_{i=1}^{N}\left(\frac{p_{it}}{p_{i1}}\right)^{\frac{1}{N}} \tag{7}$$

Again, assume the objective is to calculate the average change in the price for products A and B from time periods 1 to 2 using geometric means. Equation (6) yields the ratio of the average prices of A and B:

$$P_{12}^{RA} = \frac{(p_{A2} * p_{B2})^{1/2}}{(p_{A1} * p_{B1})^{1/2}}$$

$$= \left[\frac{(p_{A2} * p_{B2})}{(p_{A1} * p_{B1})}\right]^{\frac{1}{2}} \tag{8}$$

$$= \left(\frac{p_{A2}}{p_{A1}}\right)^{\frac{1}{2}} * \left(\frac{p_{B2}}{p_{B1}}\right)^{\frac{1}{2}}$$

Now, using Equation (7), the average of the price ratios of A and B is calculated:

$$P_{12}^{AR} = \left(\frac{p_{A2}}{p_{A1}}\right)^{\frac{1}{2}} * \left(\frac{p_{B2}}{p_{B1}}\right)^{\frac{1}{2}} \tag{9}$$

As shown, $P_{12}^{RA} = P_{12}^{AR}$ in all cases, regardless of whether prices for products are the same in the first period, as was the requirement to pass the time reversal test using arithmetic means.

To demonstrate the time reversal test using actual price data, Table 1 shows the prices of flour, eggs and bacon in two time periods - December 2001 and December 2010 – and the ratio of the price change for each commodity between the two time periods. In order to satisfy the time reversibility test, the relationship between 2001 and 2010 must be equal to the reciprocal of the relationship between 2010 and 2001. In other words, the value of the mean ratios must equal the ratio of the means.

Table 1 Comparison of Arithmetic and Geometric Means of Prices and Ratios

Product	Prices		Ratios	
	2001	2010	2001 on 2010	2010 on 2001
Flour, white, all purpose, per lb.	0.282	0.441	0.639	1.564
Bacon, sliced, per lb.	3.300	4.160	0.793	1.261
Eggs, grade A, large, per doz.	0.925	1.793	0.516	1.938
Arithmetic Mean of Prices	1.502	2.131	0.705	1.419
Geometric Mean of Prices	0.951	1.487	0.640	1.563
Arithmetic Mean of Price Ratios			0.650	1.588
Geometric Mean of Price Ratios			0.640	1.563

Source: BLS, Average Price Data

In this example, the arithmetic mean of the ratios of 2001 prices on 2010 prices is 0.650. However, the ratio of the arithmetic mean of 2001 prices on the arithmetic mean of 2010 is 0.705. The ratio of the geometric mean of 2001 prices on 2010 prices is 0.640 which is equal to the geometric mean of the ratios of 2001 prices on 2010 prices thereby satisfying the time reversibility property. Another way to examine this property is to note the reciprocal of the means of the ratios of 2001 prices on 2010 prices. The geometric mean of the ratios of 2001 prices on 2010 prices is 0.640. The reciprocal of 0.640 is 1.563, which is equal to the ratio of the geometric mean of 2010 prices on 2001 prices. Conversely, the reciprocal of the arithmetic means of the ratios of 2001 prices on 2010 prices (0.705) is 1.540, which does not equal the arithmetic means of the ratios of 2010 prices on

2001 prices. Therefore, the time reversibility quality is desirable in the context of price indexes because it allows for comparisons of prices regardless of the selected base period.

3.2 Estimating the Cost-of-Living Index: The Problem of the Consumer

Generally, there are two types of price indexes – direct and chained. Direct indexes, such as Laspeyres and Paasche, employ weighted arithmetic means of price ratios utilizing a fixed amount of goods. These indexes measure the change in price over the same set of goods regardless of any changes in quantities purchased. Specifically, direct indexes simply measure the changes in prices by computing the price relatives; between two time periods – the price level in time *t* and the price level in the base period. In contrast, chained indexes are capable of measuring the change in prices based on the share-weighted price relatives in a pair of periods. Periods used in chained indexes may be adjacent or separated by other periods. These one-period changes are then combined to yield the relative price levels over the entire period under consideration.

The following subsections discuss direct and chained indexes in the context of the consumer's cost function. Specifically, they discuss how the consumer's cost function, defined as expenditures for goods in the consumer's basket necessary to reach a target utility level, is used to determine the cost-of-living index.

3.2.1 Estimating the cost-of-living index using direct indexes

The following demonstration relies on the assumption that the consumer engages in behavior that optimizes her level of satisfaction, or utility – a theory first developed and introduced by Alexander Konüs (1939). It explains how two direct index number formulae – Laspeyres and Paasche – represent the upper and lower bounds, respectively, for estimating the change in the total cost required to maintain a consumer's given level of utility. It also explains how these approximations can be averaged to yield a more accurate estimate of the cost-of-living index. Additionally, this subsection demonstrates other methods of chaining simple price ratios which are used to estimate the change in the cost of living.

The Laspeyres price index is the share-weighted (arithmetic) average of price relatives based on items in a consumer's basket in the base period. Its twin estimator, the Paasche price index, is the share-weighted average of price relatives based on items in the consumer's basket in the current period. The Laspeyres price index for between time 0 and time *t* is defined:

$$P_{0t}^{L} = \frac{\sum_{i=1}^{N}(p_{it} * q_{i0})}{\sum_{i=1}^{N}(p_{i0} * q_{i0})} \tag{10}$$

Where p_{it} is the price of item *i* in period *t;* p_{i0} is the price of item *i* in the base period; and q_{i0} is the quantity of item *i* sold in the base period. The expenditure share s_{i0} for good *i*, or, the weighted arithmetic average of good *i*, among all goods *j = 1,...,N* in the base period is defined:

$$s_{i0} = \frac{p_{i0} * q_{i0}}{\sum_{j=1}^{N} p_{j0} * q_{j0}} \tag{11}$$

To demonstrate the measurement of the change in share-weighted price relatives, after rearranging terms in equation (10) and substituting equation (11), the Laspeyres price index formula can be rewritten:

$$P_{0t}^{L} = \sum_{i=1}^{N} s_{i0}\left(\frac{p_{it}}{p_{i0}}\right) \quad (12)$$

Where s_{i0} is the expenditure share of good *i* in the base period.

The Paasche price index formula is similar to Laspeyres, except that it considers a fixed basket of goods in period *t*, providing a comparison of prices for those goods in the base period with prices in period *t*. The Paasche price index for between time *0* (base period) and time *t* is:

$$P_{0t}^{P} = \frac{\sum_{i=1}^{N}(p_{it} * q_{it})}{\sum_{i=1}^{N}(p_{i0} * q_{it})} \quad (13)$$

Again, similar to Laspeyres, Paasche yields the share-weighted change in prices. However, the share of goods is determined in period *t* rather than the base period 0. If the share weight s_{it} is defined in equation (14), the Paasche index number can be rewritten per equation (15).

$$s_{it} = \frac{p_{it} * q_{it}}{\sum_{j=1}^{N} p_{jt} * q_{jt}} \quad (14)$$

$$P_{0t}^{P} = \sum_{i=1}^{N} s_{it}\left(\frac{p_{it}}{p_{i0}}\right) \quad (15)$$

3.2.2 Measuring changes in the cost of living using consumer behavior theory

To demonstrate how Laspeyres and Paasche indexes fit into the context of consumer choice theory, consider the relationship between the consumer's choice to consume a combination of goods and services and a budget constraint. The relationship between utility maximization and cost minimization is satisfied using duality. Constrained utility maximization implies that the consumer minimizes the cost of a given level of utility while simultaneously maximizing utility at a given budget level.

Consider, as shown in Figure 2, that a consumer optimizes utility in period 0 at point A given a certain budget level m. Point A includes a bundle that contains only goods x and y at prices p_{x0} and p_{y0}, respectively.

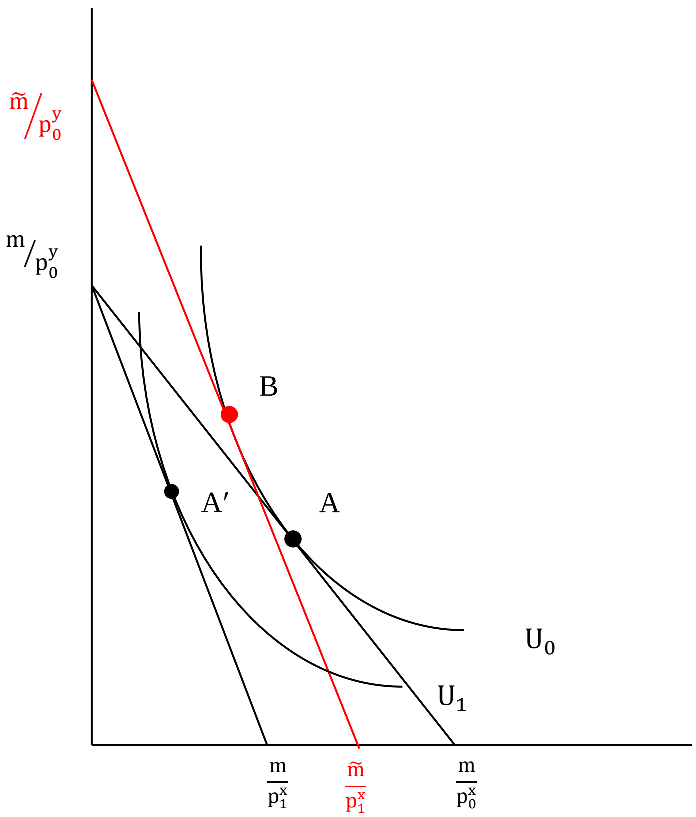

Figure 2 Effect of a Change in the Price of Good X, Previous Period Versus Current Period

Where the budget level m is defined

$$m = p_{x0}q_{x0} + p_{y0}q_{y0} \tag{16}$$

Consider that in the next period, period 1, the price of good x increases to p_{x1} and assume that the price of good y does not change. At a constant budget level m, the budget line shifts inward. Given the same level of income, the consumer must choose less of both goods x and y (shown as

Point A′) and experiences a decrease in the level of utility to U_1. For the consumer to maintain the previous level of utility U_0, income must increase to $\tilde{m}$. Therefore, the true cost of living increase, or the amount by which income must increase to maintain the same level of utility U_0, is $\tilde{m} - m$. At the $\tilde{m}$ level of income and prices p_{x1} (where $p_{x1} \neq p_{x0}$) and p_{y1} (where $p_{y1} = p_{y0}$ in this example), the consumer would likely choose a different proportional mix of goods x and y (shown as Point B) even though the income has been compensated for the increase in the price of good x. The true cost of living index considers the substitution effect of a change in the price of good x on the quantities of goods x and y, reflected by an increase in consumption of good y and a decrease in consumption in good x.

Neither the Laspeyres nor the Paasche indexes account for the substitution effects of a relative change in the prices of goods in the consumer's basket. These indexes simply calculate the amount by which income must increase to account for a change in prices assuming that the consumer chooses to consume the same exact quantities of goods at every income level. For example, the Laspeyres index is said to be a "forward-looking" index. It represents the cost of buying the same set of goods determined in a previous period; here, period 0 at new prices in period 1. In Figure 3 below, the consumer chooses to consume at Point B; however, the Laspeyres index assumes the consumer chooses to consume at Point A, which requires a greater amount of income to compensate for a price change. The estimated increase in the cost of living calculated by the Laspeyres formula is $\tilde{\tilde{m}} - m$, which is greater than the true increase in the cost of living, at $\tilde{m} - m$. Therefore, the Laspeyres index overstates the true change in the cost of living from period 0 to period 1.

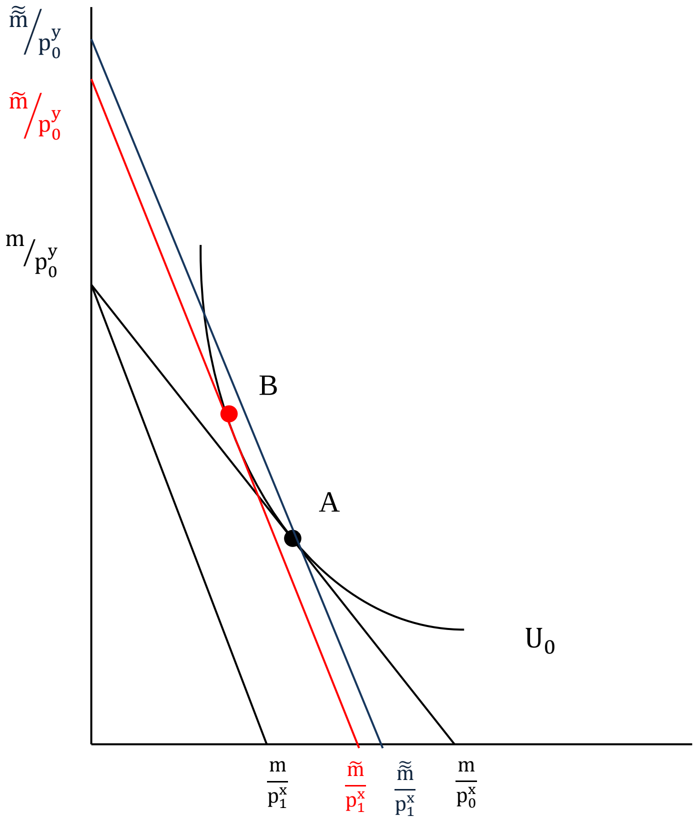

Figure 3 The Effect of a Change in Price of Good X: Laspeyres Index

Now consider, per Figure 4 below, that a consumer optimizes utility in the current period (i.e. period 1) at point A given a certain budget level m. Point A includes a bundle that contains only goods *x* and *y* at prices p_{x1} and p_{y1}, respectively. Consider that in the prior period, period 0, the price of good x was lower, at p_{x0}. Assume again that the price of good y has not changed. At a constant budget level m, the budget line shifts out such that the consumer is able to consume more of both goods x and y (shown as Point A′). The consumer previously experienced a greater level of utility at U_0. However, in this example, the consumer only needed income at $\tilde{m}$ to achieve the current period's level of utility U_1.Therefore, the true cost of living increase based on the current period's level of utility is $m - \tilde{m}$. At the m level of income and prices p_{x0} (where $p_{x0} \neq p_{x1}$) and p_{y0} (where $p_{y0} = p_{y1}$ in this example), the consumer would is likely to again choose a different proportional mix of goods x and y (shown as Point B) even though the income has been compensated for the decrease in the price of good x. Again, the true change in the cost of living considers the substitution effect of a change in the price of good x on the quantities of goods x and y, reflected by an increase in consumption of good y and a decrease in consumption of good x.

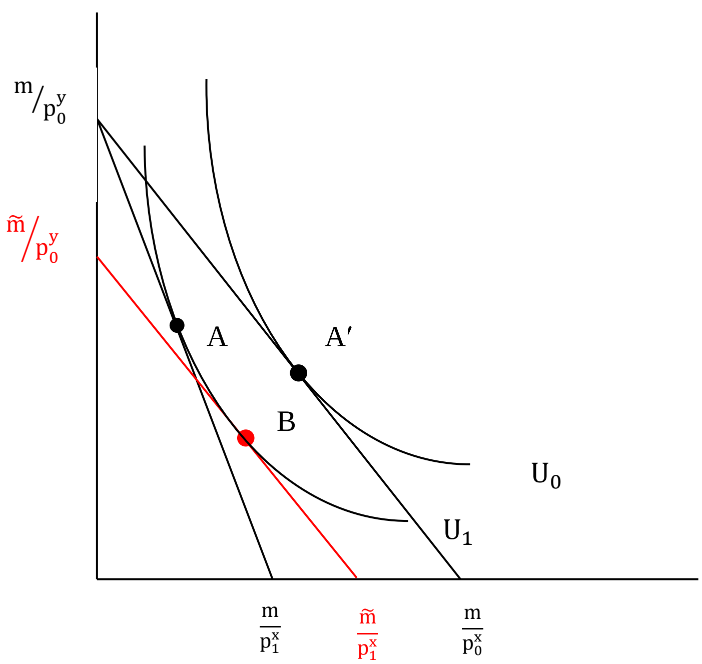

Figure 4 The Effect of a Change in Price of Good X, Current Period Versus Previous Period

The Paasche index, contrary to the Laspeyres index, is said to be a "backward-looking" index. It represents the cost of buying the set of goods determined in the current period at old prices. In Figure 5 below, the consumer chose to consume at Point B when faced with lower prices in the previous period compared to the current period. However, the Paasche index assumes the consumer chose to consume at Point A, which requires a greater amount of income in order to compensate for a price change. The estimated increase in the cost of living from period 0 to period 1 calculated by the Paasche formula is $m - \tilde{\tilde{m}}$, which is less than the true increase in the cost of living $m - \tilde{m}$. This demonstrates how the Paasche index understates the change in the cost of living.

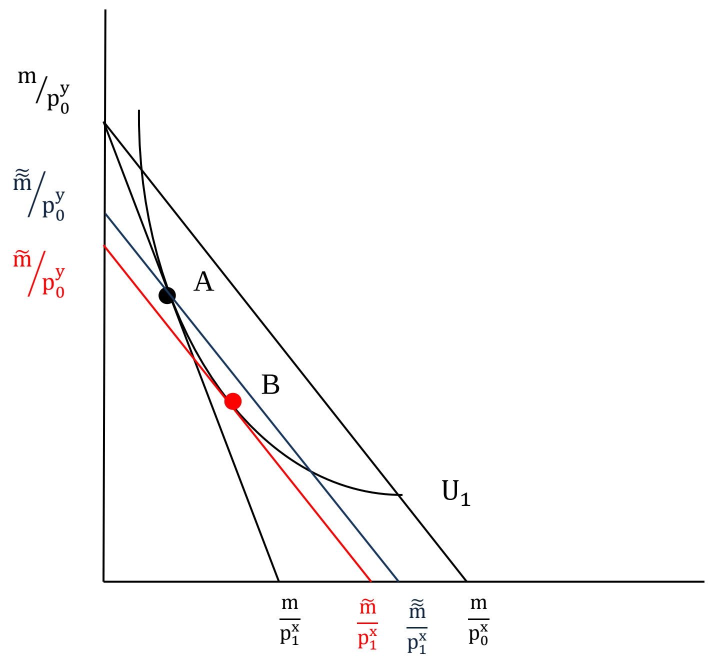

Figure 5 Effect of a Change in the Price of Good X: Paasche Index

Previous research (Bowley 1920; Pigou 1924; Fisher 1922; Konüs 1939; Diewert 1993) suggests that using the geometric mean of the Laspeyres and Paasche indexes, known as Fisher's Ideal index, yields the "exact" (term used by Diewert 1993) cost of living index. In other words, *chaining* the Laspeyres and Paasche indexes yields a closer approximation of the amount by which income must change to achieve a constant level of utility.

3.3 Chaining Direct Indexes to Obtain COLI Approximations

Chained indexes, such as Fisher's, Tornqvist, and GEKS, measure the changes in prices based on a combination of share-weighted price relatives in two adjacent or separate periods. In other words, one period's rates of change are combined, or chained, usually by taking the geometric mean of price ratios, to yield the relative levels of prices over the entire period under consideration. Fisher's index is a chained index calculated by taking the geometric mean of the Laspeyres and Paasche indexes. Under normal conditions (i.e. in the absence of price wars where quantities increase substantially and prices decrease substantially), chaining reduces the spread between the Laspeyres index and Paasche index (Diewert 1978), yielding a closer approximation of the true cost of living index.

A desirable feature of some chained indexes is the use of a flexible basket approach, which allows for consumers to substitute goods as prices and preferences change. Flexible-basket indexes differ from fixed-base indexes by accounting for the quantities consumed both before and after a price change. When exploring the feasibility of using high-frequency scanner data to develop price indexes, Ivancic, Diewert, and Fox (2011) found that estimates of price change were generally higher using fixed-basket versus flexible basket chained indexes. They postulated that this was due to new items entering the marketplace that were substitutes for other, higher priced items. These results suggested that the inclusion of new items in the price index calculation is important and that by not doing so, the calculation can bias the cost of living index estimates upwards.

3.3.1 Fisher formula

The Fisher index formula is calculated by taking the geometric mean of the Laspeyres and the Paasche. As previously described, Laspeyres is as an upper-bound of the change in prices, whereas Paasche is a lower-bound. The Fisher index between base period 0 and time t is written:

$$P_{0t}^{F} = (P_{0t}^{L} * P_{0t}^{P})^{1/2} \tag{17}$$

Thus, the Fisher index formula is considered an estimate that falls in between the upper-bound (Laspeyres) and lower-bound (Paasche) estimates of the cost of living index. Diewert (1993) indicated that the Fisher (Ideal) formula estimates the "exact" cost-of-living index. Although the Fisher index is a chained combination of the Laspeyres and Paasche indexes, these direct indexes fail to account for item substitution occurring when new goods become part of the market basket of goods, consumers switch products due to relative price changes, and consumer preferences change (as shown previously). Since these fixed-base indexes do not account for new products entering the marketplace or old products leaving over time, it becomes increasingly difficult to find goods that are in every time period. This renders the index hardly indicative of the prices of goods that are purchased by households over time.

3.3.2 Other chained indexes

Other chained indexes include Tornqvist and GEKS. The Tornqvist price index, used in calculating BLS's Chained CPI for all Urban Consumers (C-CPI-U), calculates the change in price defined as the difference in the natural log of prices weighted by the average share of the item in each time period. Similar to Fisher's formula, the Tornqvist index incorporates expenditure weights using the average share of total expenditures that each good represents in time periods 0 and t. Borrowing from equations (11) and (14), the formula for the Tornqvist price index for time t compared to time 0 is defined:

$$P_{0t}^{T} = \prod_{i=1}^{n} \left(\frac{p_{it}}{p_{i0}}\right)^{\frac{s_{i0}+s_{it}}{2}} \tag{18}$$

The natural log of the Tornqvist formula more clearly demonstrates how the average of each good's expenditure share in the two periods is multiplied by the natural log of the ratio of prices of that good between the base period *0* and period *t*:

$$ln(P_{0t}^{T}) = \frac{1}{2}\sum_{i=1}^{n}(s_{i0} + s_{it})\left(ln(p_{it}) - ln(p_{i0})\right) \tag{19}$$

An index formula proposed by Gini (1931), Elteto and Koves (1964) and Szulc (1964), referred to as GEKS, is a multilateral index which uses the geometric mean of bilateral index numbers, such as Fishers. GEKS serves as a multilateral index formula because indexes across entities or geographic areas in time can be compared directly or through their relationships with other entities, a feature lacking in bilateral indexes. This formula is often used to compare prices across countries (Kravis 1984). The GEKS formula for measuring the price index allows for new and old products to enter and leave the marketplace, while also allowing for the maximum number of product matches. Products need only be available in two concurrent periods in for inclusion in the price index calculation.

4. Estimating Prices for Selected Food Categories using Scanner Data

The previous section demonstrated the estimated upper and lower bounds of the cost of living index and how these estimates can be combined to yield a more accurate approximation of the true cost of living index. As mentioned previously, these estimates fail to account for substitution among goods and, more specifically and perhaps more importantly, they fail to account for new goods entering the marketplace. As consumers' shopping preferences shift toward lower-cost outlets, such as club stores and supercenters, they may pay lower prices for products which enable them to achieve higher levels of utility with the same levels of income. With more relative income, consumers may simply consume more of their current basket of goods, switch to products that were previously price prohibitive, such as switching from private-label (or store brand) products to branded products, or purchase new products that were unavailable before. Changes in consumer preferences due to lower prices are difficult to capture through periodic surveys, outlet sampling, and product sampling – BLS processes that are generally updated every 4 years.

Household-level scanner data collected by Nielsen and obtained by USDA's Economic Research Service (ERS) is used in this study to better capture shifts in the amount that consumers purchase, where they shop, and, to some extent, the substitution between brands of products. This section briefly describes the data and the method used to construct the price indexes which are used to measure the effect of an increased level of expenditures at supercenters and the price indexes for two food items: canned beans and milk.

4.1 About the Data

The data used for this study comes from the Nielsen Homescan panel from years 2004 through 2006. The Nielsen Homescan data comprise a national panel of 61,500 households, with approximately 39,000 households participating in the food data collection subset purchased by ERS[2]. This study uses the food subset of UPC-coded items purchased from a variety of retail outlets, such as grocery stores, supercenters, convenience stores and drug stores. Households are randomly recruited and selected based on whether they meet specific demographic characteristics to ensure that the resulting sample is nationally representative. The panel is recruited for inclusion on a permanent basis; however, adjustments due to attrition or updates to demographic targets do occur. Nielsen assigns each urban household to a specific market and projection factor. Nielsen categorizes urban households in the 2004-2006 datasets into 52 markets, which somewhat mirror Metropolitan Statistical Areas (MSAs). Households that are not located within a Nielsen market are assigned to Census Divisions based on the state where the household is located. As a result, this study includes 61 markets. A complete listing of all markets used in this study can be found in Appendix Table B-1. The projection factor assigned to each household in the panel are calculated by Nielsen based on the household's representation relative to the market and national population. These projection weights can be used to estimate population totals (Harris 2005).

4.1.1 How price and quantity data are collected

Each participant in the Nielsen Homescan panel is provided a handheld scanner which is used to collect data on individual household purchases. Following each shopping visit, the participant records the date of the purchase and store where the item was purchased. The participant then scans the barcodes for all items purchased on that trip and enters the quantity for each item purchased, whether the item was purchased at the regular price or under a promotion, and, in some cases, the price. Nielsen matches the UPC with various product characteristics, such as container size, product size, brand, flavor, and single- versus multi-pack. If the store from which the item was purchased

[2] The study author received permission to use this data via a third-party agreement with ERS.

participates in Nielsen's ScanTrack store-level data collection program, the price is imputed by Nielsen using the average price paid for that item during the week. Otherwise, the price of the item is entered by the participant (ERS 2008).

The reference period used in comparing prices and quantities consists of national average unit values for products in the first year of the dataset (2004) in order to maximize the number of product matches. Expenditure weights from the reference period are used in such a way that expenditure shares are recalculated in each subsequent period according to products available in that respective period. For example, if a product was purchased in the reference period, but not in any period thereafter, that item is dropped from the periods in which it was not purchased and expenditure shares (or share-weights) are calculated accordingly. The sample for this study was restricted to focus on refrigerated milk and canned beans in the 61 markets described earlier. Milk is chosen because it is perishable which reduces the tendency for households to purchase it in bulk quantities during periods of deep discounting. In contrast, canned beans are not as perishable and can be purchased in bulk and stored for a longer period of time.

4.1.2 Defining milk and canned bean items

Items included in this study were identified based on characteristics supplied in the Homescan dataset. For milk, all items included in the "Dairy-Milk-Refrigerated" product module were selected. These items include fat-free, lowfat, reduced-fat and whole unflavored milk but not items such as canned milk or chocolate or other flavored milk. For canned beans, items included were selected from the following product modules:

- Vegetables - Beans - Green - Canned
- Vegetables - Beans - Waxed - Canned

- Vegetables - Beans - White/Northern/Navy - Canned
- Vegetables - Beans - Vegetarian - Shelf Stable
- Vegetables - Beans - Remaining - Canned
- Vegetables - Beans - Garbanzo - Canned
- Vegetables - Beans - Lima - Canned
- Vegetables - Beans - Kidney/Red - Canned
- Vegetables - Beans - Pinto - Canned
- Vegetables - Beans - Chili - Canned

For both milk and canned beans, item quantity is based on the total number of ounces for each product purchased per household. For instance, if a one-gallon container of milk was purchased, quantity was 128 ounces. It follows that unit price is then defined as the price per ounce.

4.2 Price Index and Model Construction

As supercenters enter markets, consumers who choose to shop at supercenters experience prices that are 5-11% lower than traditional grocery stores. This represents the direct effect of supercenters on prices (ERS 2010). However, supercenters may also exhibit a secondary, or indirect, effect on price levels. Consumers who shop at supercenters are not the only ones who benefit from lower prices when supercenters enter their markets. The mere presence of supercenters increases competition such that other retailers, such as traditional grocery stores, may have to decrease prices to maintain market share. Rather than directly comparing the prices of products at supercenters with those at traditional grocery stores, which does not indicate the impact of supercenters on market-level prices, it is sensible to examine the price indexes of products across all retail outlets by geographic market (e.g. Urban NYC, Los Angeles, and Raleigh-Durham).

4.2.1 Price index construction

Two sets of price indexes for each product in each geographic market were constructed to estimate the effect of supercenters and price indexes.

- Price Index 1: Laspeyres price indexes for canned beans and milk excluding prices from supercenters. Quantities of products sold at supercenters were matched with prices for identical products at non-supercenters. Products that are unique to supercenters were excluded.

- Price Index 2: Laspeyres price indexes including supercenter prices for products. This is a truer Laspeyres index since it matches all quantities for products with the prices paid for those items. BLS does not completely exclude supercenters from price data collection, though, it is alleged that BLS under-samples supercenters for prices. Consequently, this index should be lower than the index that excludes supercenter prices to the extent that supercenters do, in fact, offer lower prices.

Not all households purchase the same items, or the same quantity, each month. For instance, some households may stockpile non-perishable items in any given month (i.e. purchase in bulk) and go an extended period of time before making a repeat purchase. To capture the average quantity purchased for each household, it is important to consider all households in the sample in each month; not just the households that purchased the goods in question. Accordingly, average quantity purchased considers both households that purchased the item in question in each month as well as households that did not.

A product's Universal Product Code (or UPC) is a unique code; however, multiple UPCs may be used for what is essentially the same product. This is because, over time, manufacturers may implement minute changes in formulation, change a product's labeling, or change a product's packaging (Reinsdorf 1999). For the purposes of this study, products with identical price-determining characteristics (e.g. brand, container size, flavor or type) are treated as the same item. This helps to ensure the maximum number of product matches between the base period and the geographic market and time combination. Also, identical products purchased at different outlets will be treated as identical products. In other words, this study assumes no differences in the quality of the shopping experience across different outlets. This is a somewhat controversial assumption. Research has indicated that a true difference in the quality of the supercenter shopping experience exists as it relates to price, such as differences in the services offered, locational convenience, and product offerings (Schultz and Mackie 2002, 168). It is difficult, however, to quantify degree by which an improved shopping experience should increase the price of a product. In contrast, as incomes rise, as they have done for much of the United States over the past decade, consumers continue to shift from traditional grocery stores to lower-priced supercenters, undermining the notion that supercenters provide a shopping experience so inferior as to completely offset any reduction in prices. Even Schultz and Mackie acknowledged that supercenter gains in popularity indicate that, even after adjustments for quality, the lower prices offered by supercenters more than offset the decrease in the shopping experience.

4.2.2 Econometric models

The difference in price indexes excluding and including supercenter prices is regressed on the percent of transactions at supercenters to examine the effect of supercenters on price indexes. The proportion of transactions occurring at supercenters serves as a proxy for the supercenters' role in the marketplace and estimates supercenters' market shares of milk and canned beans. Using the percentage of transactions at supercenters to predict the difference in price indexes, however, may introduce endogeneity. For instance, supercenters may choose to locate in markets where real estate prices are relatively low indicating that, perhaps, food input prices were already low in those markets relative to markets with high real estate values. If prices were already relatively low in areas and supercenters chose to locate in a market because of other (omitted) factors, the relationship between the presence of supercenters and the difference in price indexes may not be causal. In addition, lower supercenter prices drive more consumers to shop at supercenters and therefore increase the percentage of supercenter transactions. Supercenters may also choose to locate where prices are high because of the potential profit from undercutting the competition.

The house price index is introduced as an instrumental variable (IV) to control for some of the potential endogeneity. Two assumptions must hold for the house price index to be a relevant and valid instrument. First, the independent variable - percent of transactions at supercenters - must be significantly correlated with the instrument (relevancy). Second, the instrument must be uncorrelated with the error term, except through the variable on percentage of transactions at supercenters (exclusivity). A later section demonstrates why the house price index is both relevant and exclusive.

House price index data are compiled by the Federal Housing Finance Agency (FHFA) for the top 25 Metropolitan Statistical Areas (MSAs), states, and census divisions by quarter. Because store location decisions are typically made well in advance of the first transaction, the non-seasonally adjusted house price index variable is lagged by two years (assuming one year for site selection and one year for construction). FHFA house price index data are published quarterly. Cubic spline interpolation (PROC EXPAND using SAS® software), a commonly used technique to extrapolate economic quarterly data into monthly, is used to extrapolate monthly data values (Becker and Hurn

2004, 162). This method has been used extensively in research (McColluch and Kwon 1993, Hodrick and Prescott 1997) and extensively explored by Pavía-Miralles (2010).

First, where available, Nielsen Homescan markets were assigned to MSAs included in the FHFA dataset. Second, house price indexes were assigned to census divisions where appropriate. Third, the remaining Homescan markets were assigned to house price indexes for states where those markets are predominantly located. A list of the Homescan market-to-FHFA data crosswalk is found in Appendix Table B-1.

After constructing unit values (i.e. price per unit) for milk and canned beans at the item level, quantity variables (i.e. number of units), the instrument (house price index), and assigning non-urban households to census regions, price indexes were generated using SAS® software. The resulting dataset consists of T = 36 periods and N = 61 geographic markets, combining for a total of 2,196 observations. Because there are no missing values in any time period for any market, this is a balanced panel dataset.

For each market *i* and period *t*, the difference in price indexes excluding and including supercenter prices is the dependent variable and the percentage of transactions at supercenters is the independent variable. For each product, the restricted model, which assumes the effect is the same across all markets, to estimate this relationship can be expressed:

$$\ddot{P}_{it}^{L} - P_{it}^{L} = \alpha_i + \beta PercentSC_{it} + \varepsilon_{it} \tag{20}$$

$$i = 1,...,61 \text{ and } t = 1,\ldots,36$$

Where $\ddot{P}_{it}^{L}$ is the Laspeyres price index excluding prices from supercenters for market *i* in time *t*, P_{it}^{L} is the Laspeyres price index including prices from supercenters, α_i is the market-specific intercept term, $PercentSC_{it}$ is the percent of household transactions that took place at supercenters in market *i* in time t, and ε_{it} is the error term that includes the unobservable geographic market-specific characteristics and random price shocks. The coefficient β is the estimated effect of supercenters on the price index assumed constant across markets. In addition, an unrestricted model which allows the coefficient β to vary by market is estimated to determine whether the effect of supercenters is the same in each market. This model is expressed:

$$\ddot{P}_{it}^{L} - P_{it}^{L} = \alpha_i + \beta_i PercentSC_{it} + \varepsilon_{it} \qquad (21)$$

$$i = 1,...,61 \text{ and } t = 1,\ldots,36$$

For ease in interpretation, the coefficient β in the IV model does not vary by market and can be expressed:

$$\ddot{P}_{it}^{L} - P_{it}^{L} = \alpha_i + \beta_{IV} PercentSC_{it} + \varepsilon_{it}$$

$$\text{instrument: } HPI_{it} \qquad (21)$$

$$i = 1,...,61 \text{ and } t = 1,\ldots,36$$

Where HPI is the house price index. The model without controlling for endogeneity examines the association between supercenters and price indexes by examining a correlation between the presence of supercenters and the price indexes across geographic markets and time. Controlling for endogeneity using the instrumental variable, this model estimates the causal effect of supercenters on the price indexes for canned beans and milk.

4.2.2.1 Testing for serial correlation

To determine the appropriate model to use for estimating this relationship, it is necessary to determine whether the idiosyncratic errors ε_{it} are serially correlated over time (Wooldridge 2009). Ignoring serial correlation when it exists will yield consistent coefficient estimates although the standard errors may not be correct (Baltagi 2005, 84). For each product consistent with the general models presented in the previous section, the *xtserial* command in Stata is used to determine the presence of serial correlation while accounting for multiple time series (StataCorp 2011). This command employs Wooldridge's method using the residuals from a regression in first-differences (Drukker 2003). As shown in Table 2 below, the null hypothesis of no serial correlation in ε_{it} is not rejected for canned beans ($p = 0.457$); however, the null hypothesis of no serial correlation in ε_{it} for milk is rejected ($p = 0.003$).

Table 2 Test for Serial Correlation in the Difference in Laspeyres Indexes Excluding and Including Supercenters' Prices, Milk and Canned Beans, N=2196

Product	$\ddot{P}^L_{it} - P^L_{it}$, ρ	p-value
Canned Beans	0.1464	0.457
Milk	0.4615	0.003

This test indicates that, for canned beans, the fixed-effects model without correction of the variance-covariance matrix is the efficient model to use. A first-order autoregression, or AR(1), model is used to estimate a model for milk with robust standard errors that are corrected for serial correlation:

$$\ddot{P}^L_{it} - P^L_{it} = \delta_0 + \beta_i PercentSC_{it} + e_{it}$$
$$\text{where } e_{it} = \rho_i e_{it-1} + \xi_{it} \tag{22}$$

Where the error term e_{it} is serially correlated and ξ_{it} is serially independent. Since the correlation coefficient for milk shown in Table 2 is relatively low, for ease of interpretation, this study will not model the difference in price indexes for milk with both IV and autoregression. Results from the following models will be presented separately for each product:

- a fixed-effects model that does not correct for serial correlation or potential endogeneity;

- a fixed-effects instrumental variable model that corrects for some potential endogeneity but does not correct for serial correlation; and,
- a fixed-effects first-order auto-regression model that does not correct for potential endogeneity but corrects for serial correlation.

4.2.2.2 Testing the instrument for relevancy and exclusivity

House price index must be correlated with the presence of supercenters (relevant) and uncorrelated with error term (exclusive) to be an appropriate instrument. First, the correlation between the house price index and the percent of transactions at supercenters for canned beans and milk was examined to determine its relevance (PROC CORR, SAS®). As shown in Table 3 below, house price index was significantly, and negatively, correlated with percent of transactions at supercenters for both canned beans and milk.

Table 3 Pearson's Correlation Between House Price Index and Percent of Transactions at Supercenters, Milk and Canned Beans, N=2196

% of Transactions at Supercenters	House Price Index, ρ	p-value
Canned Beans	-0.2568	<0.001
Milk	-0.3014	<0.001

The null hypothesis of no correlation between the percent of transactions at supercenters and the house price index was rejected for both canned beans and milk ($p<0.001$ in both cases). Therefore, house price index is a relevant instrument to use to control for endogeneity in the model.

To satisfy the second assumption, house price indexes must be uncorrelated with food price indexes except through the presence of supercenters. Consider that food consumed at home is primarily sold at supercenters and grocery stores which are comprised of large national and regional chains. High-end stores may choose to locate in high-income areas with higher real estate values to be closer to their targeted customer base. Similarly, discount stores may choose to locate in areas with low real estate to be close their targeted customer base but they may also locate in low real estate areas to reduce overhead costs. Discount stores, by virtue of offering lower prices, must mitigate slim sales margins (i.e. revenue less cost of goods) by cutting costs in other areas. Additionally, supercenters may choose to locate in areas with low real estate values particularly due to the large footprint of their stores. In other words, supercenters may choose to locate in low real estate areas simply because doing so lowers the overall cost of goods and, therefore, maximizes the profit margin on the goods they sell. It is not likely that supercenters locate in areas because food prices are lower, therefore indicating that the HPI acts through the presence of supercenters to lower food prices instead of acting on food prices directly. Because the house price index only effects food price indexes through the presence of supercenters, which is partially determined by real estate values, HPI is a reasonable instrumental variable for the percentage of food sales at supercenters.

A time-series plot of the national house price index lagged two years is shown in Figure 6. Cubic spline interpolation was used to estimate the months between quarters since the data are published quarterly and food price index data are monthly. The results of the interpolation are demonstrated in Figure 6. The solid line shows the interpolated monthly house price index values and the diamond-shaped markers represent the published national quarterly house price index (Quarter 0 corresponds to Quarter 1 2002) values, as published by FHFA.

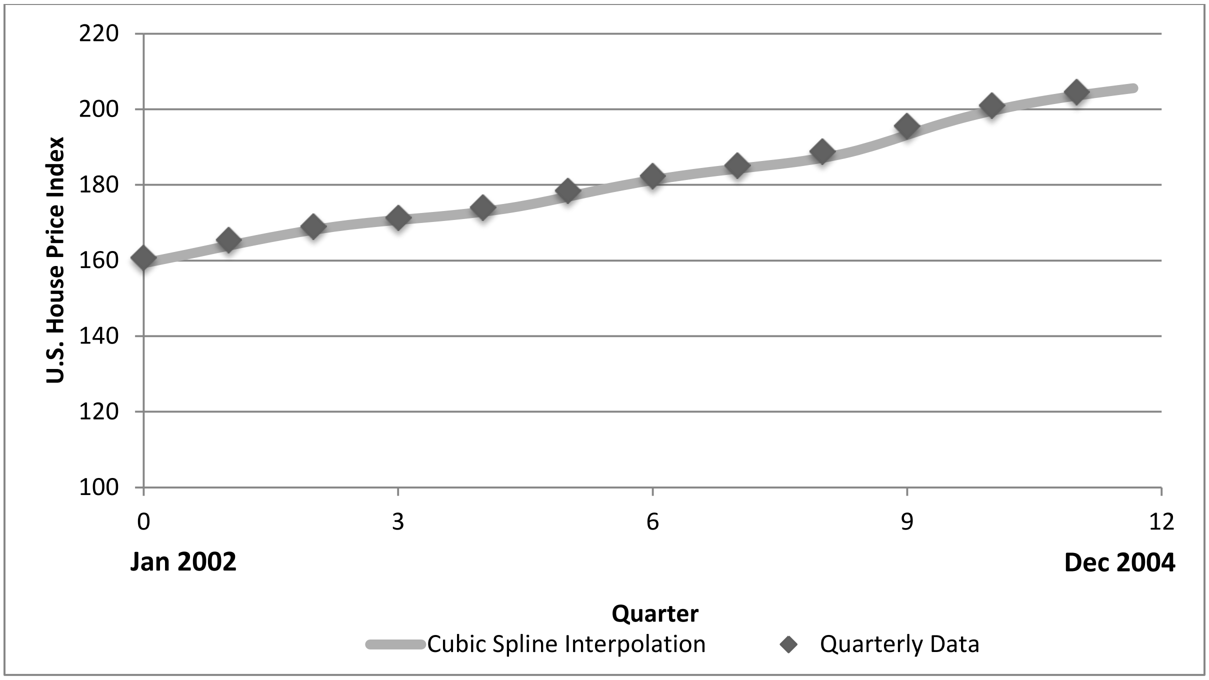

Figure 6 U.S. House Price Index: Quarterly Data and Cubic Spline Interpolation, January 2002 – December 2004

The results from the FE models with and without IV and the AR(1) models (without IV) are presented in the results section to determine whether these identification strategies yield different results. As mentioned previously, the relationship between the percentage of transactions at supercenters and the difference in price indexes excluding supercenter prices and including supercenter prices for both canned beans and milk is expected to be positive and statistically significant.

5. Results

During the period between January 2004 and December 2006, the overall (national) proportions of transactions at supercenters for milk and canned beans increased; however, this increase varied substantially by geographic market. For example, two markets had no supercenter transactions for canned beans during the 36 month period and one market had no supercenter transactions for milk. Other markets had less than 1% of canned beans or milk transactions at supercenters. In contrast, some markets exhibited significantly higher proportions of transactions at supercenters (see Appendix Table B-3 for a complete listing of overall proportions of supercenter transactions for canned beans and milk).

This study focuses primarily on changes in the presence of supercenters over time. Figure 7 and Figure 8 depict the trends in the proportion of supercenter transactions nationally and for select markets for canned beans and milk, respectively. The markets selected in these figures were chosen because they represented the extremes in the data: the largest increase in proportion of supercenter transactions; the largest decrease in proportion of supercenter transactions; the largest difference in Laspeyres indexes excluding and including supercenters; and, the smallest difference in Laspeyres indexes excluding and including supercenters.

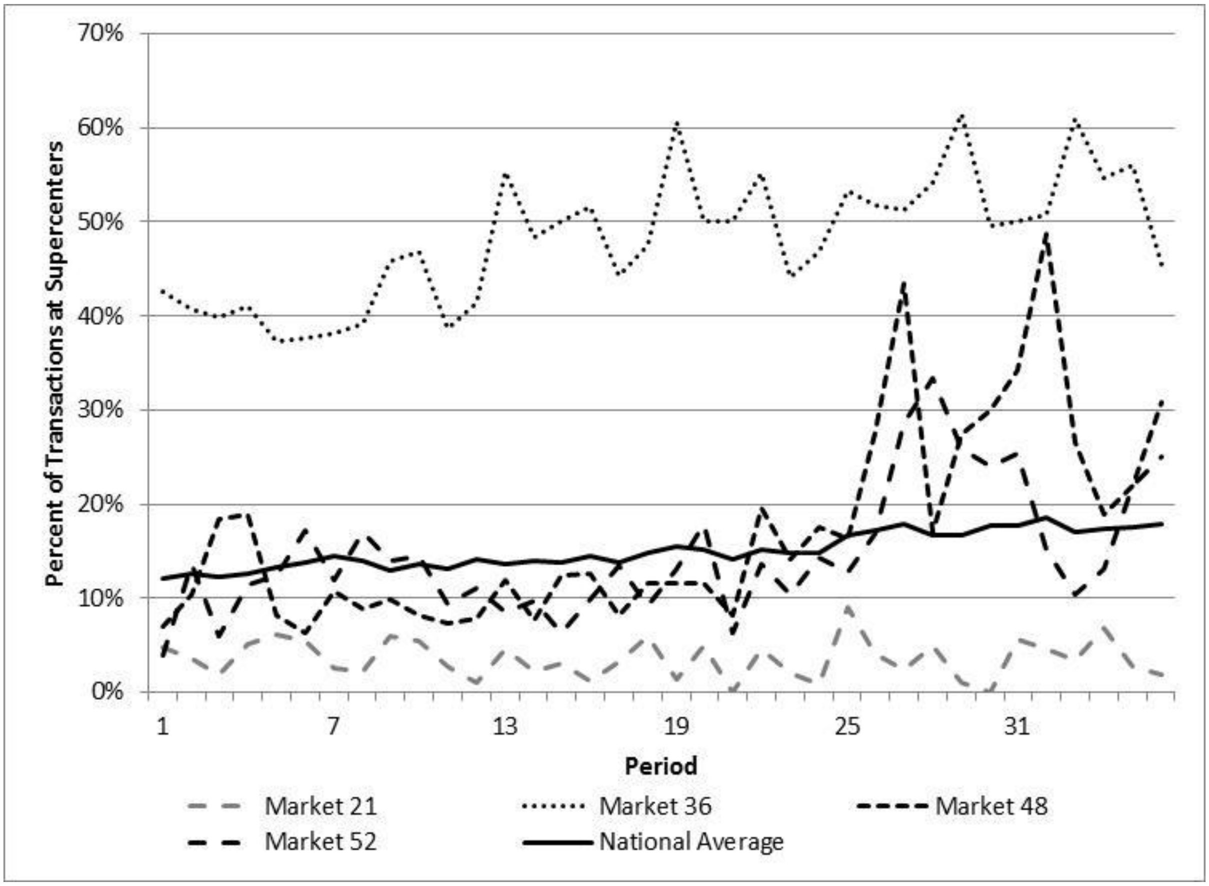

Figure 7 Percent of Canned Beans Transactions at Supercenters, National Average and Select Markets

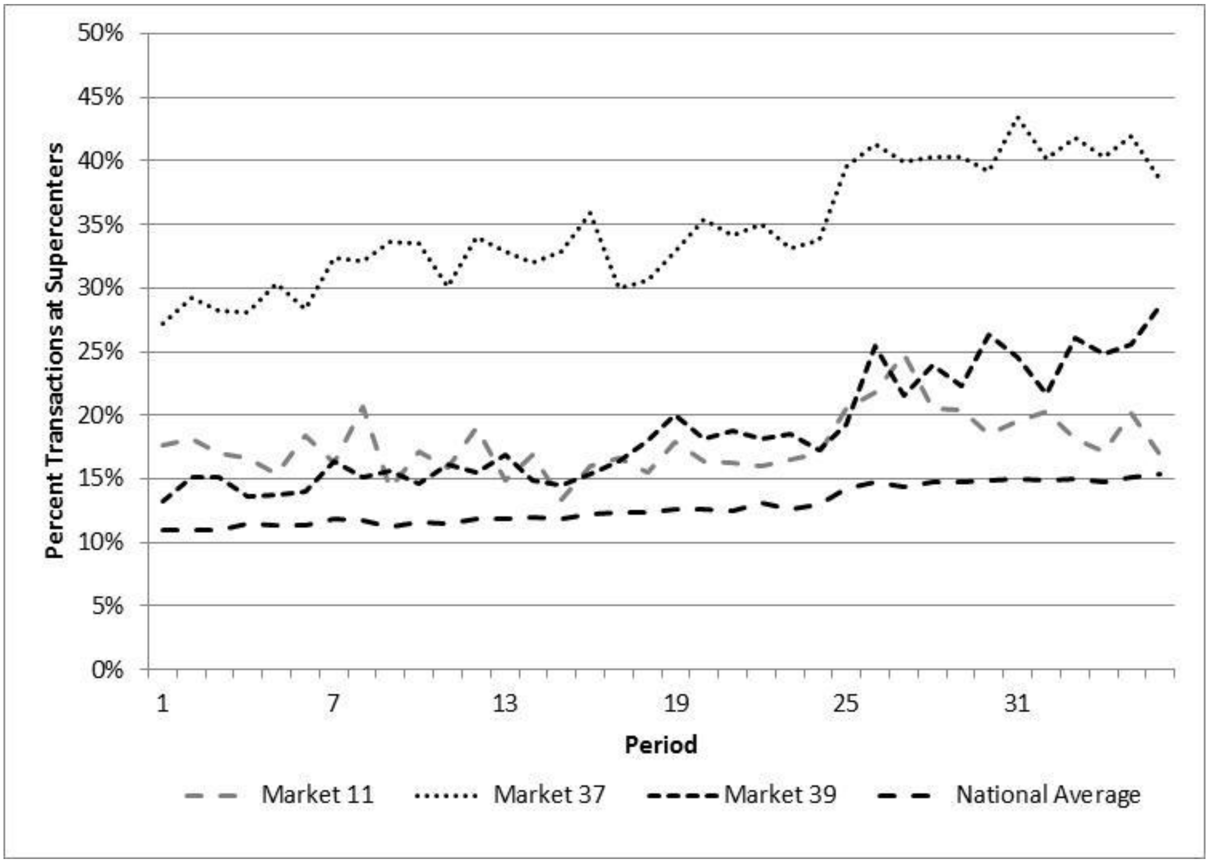

Figure 8 Percent of Milk Transactions at Supercenters, National Average and Select Markets

This section discusses the results of the fixed-effects and first-order autoregressive models described earlier. To put the results of the models into perspective, it is important to first examine some descriptive statistics. These include the change in the proportion of supercenter transactions at supercenters from the first month of the data (January 2004) to the last month of the data (December 2006), and the range in the difference of Laspeyres price indexes excluding and including supercenter prices.

5.1 Percentage of Household Purchases from Supercenters

The percentages of transactions that took place at supercenters generally increased over time; however, this increase varied by market and in some cases was nonexistent. In the first time period, January 2004, supercenters sold 12.8% of all canned bean items (see Table 4). This figure increased 5.8 percentage points (an increase of 45%) to 18.6% in the last time period, December 2006, as shown in Table 4. Milk transactions at supercenters also increased over the same time frame. In January 2004, 10% of all milk items were purchased at supercenters. By December 2006, this figure increased 4 percentages points (a 40% increase) to 14%. Several markets also demonstrated a substantial increase in the proportion of household transactions taking place at supercenters for canned beans and milk[3].

Table 4 depicts the markets exhibiting the largest increases in the proportion of supercenter transactions for each product and the markets exhibiting the largest decreases in the proportion of supercenter transactions for each item from January 2004 to December 2006. In Market 48, for example, 6.9% of household canned beans purchases took place at supercenters in January 2004. By December 2006, this percentage increased 23.8 percentage points to 30.8%. In Market 39, 13.2% of household milk purchases took place at supercenters in January 2004. In December 2006, this percentage increased 15.4 percentage points to 28.6%. Some markets experienced no growth in supercenter transactions over time and some markets actually demonstrated a decrease in the proportion of supercenter transactions for products. For example, 4.8% of canned beans purchases in Market 21 took place at supercenters in January 2004. This percentage decreased 2.8 percentage points in December 2006 to just 2%. In Market 11, 17.7% of milk purchases took place at supercenters in January 2004. This decreased 0.7 percentage points to 17% in December 2006. This

[3] Geographic region identifiers are disassociated from the results to avoid disclosing market shares for specific retailers.

decrease is not substantial and likely represents a steadier presence of supercenters over time. Appendix Table B-2 contains a list of all markets and the proportion of supercenter transactions in January 2004, including the relative percentage point increase (decrease) in December 2006 for canned beans and milk.

Table 4 Percent of Transactions at Supercenters in Period 1 (January 2004) and the Difference at Period 36 (December 2006) for Select Markets, Canned Beans and Milk

Market	% of Transactions at Supercenters, January 2004	P.P. Increase (Decrease) at December 2006
Canned Beans		
National Average	**12.8%**	**5.8**
Market 48	6.9%	23.8
Market 21	4.8%	(2.8)
Milk		
National Average	**10.0%**	**4.0**
Market 39	13.2%	15.4
Market 11	17.7%	(0.7)

5.2 Difference in Laspeyres Price Indexes Excluding and Including Supercenters' Prices

As described earlier, the difference in the Laspeyres price indexes excluding and including supercenter prices represents the direct effect of supercenters on price indexes. The overall Laspeyres price index including supercenters' prices for canned bean items, shown in Table 5, was 110.0 and the overall Laspeyres price index excluding supercenter prices for canned bean items was 109.3 – a difference of 0.7 percentage points. The largest observed difference in canned bean price indexes was in Market 36, at 2.8 percentage points. Additionally, in Market 52, the price index for canned beans including supercenter prices was surprisingly lower than the index excluding supercenter prices; however, this difference was not substantial. This observation is likely attributable to milk transactions at local dairies, given the prevalence of dairies in the region. Overall, the variation in the difference in price indexes for milk items was not substantial nationally or across markets, indicating that it may be difficult to estimate any effect of supercenters on the price index for milk. The Laspeyres price index including supercenters' prices for milk items is 100.4, and the Laspeyres price index excluding supercenters' prices for milk items is 100.3 – a difference of only 0.1 percentage point. The largest observed difference in price indexes for milk was in Market 39, where the price index excluding supercenter prices was 1.0 percentage point higher than the price index including supercenter prices. Similar to canned beans, the price index for milk including supercenter prices was lower than the index excluding supercenter prices in Market 37. Again, this difference was not substantial. The indexes for all markets can be found in Appendix Table B-4 (canned beans) and Appendix Table B-5.

Table 5 Laspeyres Price Indexes (Base=100) Excluding and Including Supercenters' Prices for Select Markets, Canned Beans and Milk, 2004-2006

Market	P^L Excluding Supercenters' Prices	P^L Including Supercenters' Prices	Difference
Canned Beans			
National Average	**110.0**	**109.3**	**0.7**
Market 36	111.5	108.8	2.8
Market 52	105.3	106.3	(1.0)
Milk			
National Average	**100.4**	**100.3**	**0.1**
Market 39	103.7	102.7	1.0
Market 37	100.1	100.8	(0.7)

5.3 Econometric Models on the Effect of Supercenter Transactions on the Difference in Price Indexes Excluding and Including Supercenters' Prices

Several models were estimated for each product to determine the association and causation between the percent of household transactions at supercenters and the difference in estimated price indexes excluding and including supercenter prices. Models were estimated using *xtreg*, a Stata command to estimate fixed effects between time and within markets, the *xtivreg* command to estimate IV models with fixed effects between time and within markets, and the *xtregar* command for estimating a first-order autoregression (AR(1)) to examine fixed effects in the presence of autocorrelation (StataCorp, 2011). Additionally, models were estimated for both products to estimate individual market-level effects, and to test whether the effect of supercenters was the same for each market. The results of the fixed-effects (with and without IV) and AR (1) models for canned beans and milk are described in the following subsections.

Models with market-specific coefficients are discussed in this section; however, due to size, they are listed in Appendix Table B-6 and Appendix Table B-7, respectively. Per the presence of correlation in the idiosyncratic errors, the preferred model for milk is the first-order autoregressive model. Conversely, per the lack of evidence of serial correlation in the difference in price indexes for canned beans, the preferred model is the fixed-effects (FE) IV model without the correction of the variance/covariance matrix.

5.3.1 Effect of supercenters on the price index of canned beans

As shown in Table 6, the FE models with and without IV for canned beans are similar, although the magnitude of the coefficient in the IV model is greater. The fixed-effects model with IV estimated a positive and significant effect of supercenters on the difference in the price index for canned beans ($\beta = 0.16; p < 0.001$), indicating that the increased presence of supercenters in a market causes a larger gap in the price index excluding and including supercenters' prices. To put this coefficient into context, 15% of all canned beans were purchased at supercenters during the 36 month period. The results indicate that, for each percentage point increase in the supercenter market share for canned beans, the gap between the price index excluding and including supercenter prices for canned beans increases 0.16 percentage points.

Table 6 Regression of Difference in Laspeyres Price Indexes Excluding and Including Supercenter Prices on Percent of Transactions at Supercenters for Canned Beans, Fixed-Effects (FE) With and Without IV and First-Order Autoregression (AR1)

Variable	$\ddot{P}^L_{it} - P^L_{it}$, Canned Beans		
	FE	**FE (IV)**	**AR(1)**
% of transactions at supercenters	0.1446* (0.0091)	0.1637* (0.0315)	0. 1464* (0.0097)
R-Squared	0.0987	0.0987	0.0985
N	2196	2196	2135
F (or Wald χ^2 for IV)	253.89	474.91	226.78

Standard errors are reported in parentheses.
* Denotes significance at the p<0.01 level.

Individual market-level coefficients (without IV) were estimated to determine whether the effect of supercenters on price indexes for canned beans is the same across all markets. After testing the restricted versus the unrestricted models (e.g. $\beta_1 = \beta_2 =, \dots, = \beta_{61}$), the null hypothesis that all coefficients are the same is rejected at the 1% level, $F(61, 2134) = 3.05$, $p < 0.01$. In addition, the results of the model examining market-level effects revealed statistically significant effects of supercenters on price indexes for canned beans. Out of 59 geographic markets[4], 12 markets had a positive and significant effect of supercenters on the price index at the 0.1% level; 8 markets at the

[4] Two markets were omitted from the model estimating market-level effects because there were insufficient supercenter transactions involving canned beans over the 36 month period.

1% level; and, 4 markets at the 5% level. The effect of supercenters on the difference in the price indexes excluding and including supercenter prices for canned beans varied substantially across markets. For example, the effect of supercenters on the difference in price indexes for canned beans in Market 44 was 0.395 – the largest effect across all 61 markets ($p < 0.01$). Considering supercenters in Market 44 increased market share over the 36 month period from 10.9% to 15.0%, this gap in price indexes is substantial. Market 18 also experienced a large effect of supercenters on price indexes ($\beta = 0.36, p < 0.01$). Supercenters in Market 18 increased market share from 15.2% to 24.0% over the 36 month period. These gaps are substantial if, in fact, BLS does under-sample supercenters when it chooses outlets from which to collect price data for goods. This relationship could indicate that, for non-perishable products such as canned beans, the correlation between the market share held by supercenters and the price index for those goods is strong.

Surprisingly, one market's (Market 34) coefficient was negative and significant at the 5% level. However, this market had a substantial and constant supercenter presence throughout the duration of the study; therefore, it is unlikely that this coefficient represents causality. Instead, this difference could be due to the issue of item substitution bias that arises from using fixed-base price indexes discussed extensively in a previous section. Also, since supercenters comprise a substantial market share in this particular market, perhaps other retailers in this market lowered their prices to compete with supercenters. The results from the model with individual market-level coefficients for canned beans can be found in Appendix Table B-6.

5.3.2 Effect of supercenters on the price index of milk

The effect of supercenters on the price index of milk was substantially smaller than that observed for canned beans. For milk (shown in Table 7), the models found a substantial difference not only with the magnitude of the estimator β, but also with the estimator's level of significance relative to the results for canned beans. The IV model found no significant effect of supercenters on the price index. Results from the AR(1) model estimated a positive and significant relationship between an increased presence of supercenters and the difference in the price index excluding and including supercenters prices for milk ($\beta = 0.02; p < 0.01$). The overall difference in the price indexes for milk excluding and including supercenters' prices was minimal, at 0.1 percentage points, and there was little variation in the price index across markets. This result is not surprising; however, it is worth noting that the correlation between the presence of supercenters and the price index for milk, a perishable good, is significant. This may indicate that while the price for milk is not as strongly correlated with the market share held by supercenters, a relationship does exist.

Table 7 Regression of Difference in Laspeyres Price Indexes Excluding and Including Supercenter Prices on Percent of Transactions at Supercenters for Milk, Fixed-Effects (FE) and First-Order Autoregression (AR1)

Variable	$\ddot{P}_{it}^{L} - P_{it}^{L}$, Milk		
	FE	**FE (IV)**	**AR(1)**
% of transactions at supercenters	0.0073 (0.0059)	0.0024 (0.0150)	0.0242* (0.0075)
R-Squared	0.0148	0.0148	0.0161
N	2196	2196	2135
F (or Wald χ^2 for IV)	1.53	1.67	10.32

Standard errors are reported in parentheses.
* Denotes significance at the p<0.01 level.

In the model for milk, it is possible that the chosen instrument is poor and that the estimated coefficient is biased (Wooldridge 2009, 519). However, since the correlation between the independent variable – percent of transactions at supercenters – is moderately correlated with the instrument, this bias is likely not substantial. It is also possible that the model identification strategy for milk used in this study is not suitable for detecting the effect of supercenters on price indexes. However, it may be that supercenters have no effect on the price index for milk.

Similar to canned beans, individual market-level coefficients (without IV) were estimated to determine whether the effect of supercenters on price indexes for milk is the same across all markets. After testing the restricted versus the unrestricted models as before, the null hypothesis that all coefficients are the same is rejected at the 1% level, $F(61, 2134) = 3.13, p < 0.01$. However, the appropriateness of the model used to estimate the effect of supercenters on the price indexes for milk is questionable due to a low F value. Also, neither the dependent nor the independent variables in the milk model varied substantially across markets or across time. Nevertheless, the results of the model examining market-level effects revealed statistically significant effects of supercenters on price indexes for milk. Out of 61 geographic markets, 3 markets had a positive and significant effect of supercenters on the price index at the 0.1% level; 5 markets at the 1% level; and, 4 markets at the 5% level.

The effect of supercenters on the difference in the price indexes excluding and including supercenters' prices for milk also varied widely across markets. For example, the effect of supercenters on the difference in price indexes for milk in Market 40 was 0.23 – the largest effect across all 61 markets ($p < 0.001$). Supercenters in Market 40 increased market share over the 36 month period from 18.4% to 24.2%. Market 58 also experienced a large effect of supercenters on price indexes, but this effect was surprisingly negative ($\beta = -0.21, p < 0.001$). Considering the difference in price indexes for milk excluding and including supercenter prices was less than 0.1% in Market 58, the negative coefficient is likely the result of an improper model rather than an indication of causality. There were other market-level coefficients that were significant and negative: Market 18, Market 36, and Market 37. In all three markets, the differences in price indexes excluding and including supercenter prices were small, but negative. These negative coefficients are likely the result of an improper model. The results from the model with individual market-level coefficients for milk can be found in Appendix Table B-7.

6. Limitations of the Study

Several limitations are important to highlight before interpreting the possible implications of this study's results. Potential limitations may include the products used in this study, data used, price index method, and the model specification. A broad limitation relates to using the percent of transactions at supercenters as an indicator of the presence of supercenters. In some markets, such as Sacramento and San Francisco, municipalities have placed development restrictions to prohibit big-box stores from locating within municipal boundaries. Such endogeneity is difficult to control in the context of broad geographic markets, as planning decisions are typically made at the local level. Additional and more specific limitations are detailed in this section.

6.1 Using Milk to Detect Differences in Price Indexes

Using milk to detect a difference in price indexes as the proportion of supercenter transactions increase may understate the effect supercenters have on price indexes. This study found that the variation in milk prices over time and across markets was minimal. Consequently, it is difficult to draw any conclusion of the effect of supercenters on prices since the prices do not differ. This lack of price variation may result from infrequent discounting of milk by retail outlets. Also, the wholesale pricing structure for milk is unique, which may impact the retail price and the regional variation in prices (or lack thereof). Additionally, state and federal laws may distort the price for milk by setting minimum prices received by dairy farmers.

6.2 Limitations in Using Homescan Data

A further limitation of this study exists in the data corresponding to gaps in coverage and possible measurement error. Although Nielsen Homescan data contains information on food purchased from a variety of retail channels, it does not contain data on food purchased for consumption away from home. Also, the prices for products included in the data are self-reported, which may lead to measurement error. ERS (2008) conducted a study to measure the accuracy of Homescan data and found that measurement error associated with Nielsen Homescan data seemed to be in line with other surveys. However, the study did find errors in recording price information, especially when price data are recorded by Nielsen by matching transactions with store-level scanner data. This is largely attributable to price data not accounting for coupons or loyalty card discounts. Such measurement error could be problematic when examining relationships involving prices paid.

6.3 Limitations in Using a Fixed-Basket Index to Determine Price Levels

This study employed a fixed-basket approach to calculating price indexes. This is a reasonable approach for short time periods, as fewer new items are introduced and consumer preferences do not change substantially. However, as consumer preferences change, so should the expenditure weights used in determining the extent to which price levels change. If supercenters introduced new products which were not available in the base period, those items were excluded from the index. A better price index to detect differences due to supercenters would allow for expenditure weights to change and introduction of new products, such as GEKS.

6.4 Limitations in House Price Index as an Instrumental Variable

The use of house price index as an instrumental variable to control for endogeneity in the model presents some challenges. First, the house price index data are published quarterly rather than monthly. Second, it is difficult to estimate how far in advance real estate prices affecting site selection proceeds the first supercenter transaction. Also, since the house price index data are not published for the same geographic market areas as the Homescan data, it is difficult to completely align the HPI values from MSAs and states to Homescan markets. Further, real estate valuation differences are highly localized. A difference in the price of real estate in one location relative to another may be attributable to road access, crime, or other variables which are not related to spatial competition or potential profitability. Also, because the HPI data are at broad MSA, state or census division levels, it is difficult to detect a difference such that correlations can be established between site selection and market prominence. Examining the data at a micro level, such as census tract, ZIP code, or some other trading market cluster, and knowing the point at which chains make site selection decisions, may provide a better indication of the relationship between market selection and site selection.

7. Discussion

In 2006, total U.S. Social Security disbursements were $546.2 billion – 4.2% of U.S. Gross Domestic Product (Trustees Report, 2007). Based on the CPI calculated by BLS, the Social Security Administration implemented a cost-of-living-adjustment of 3.3% for benefits dispersed in 2007. Assuming the same beneficiaries year-over-year, the cost-of-living adjustment increased the payments disbursed by $1.8 billion dollars. With U.S. federal expenditures growing in recent years and, in each year, exceeding revenues by a substantial amount, the methodologies behind determining these cost-of-living adjustments should be further examined, especially in the light of studies such as this one, which examined the effect of the growing market share of supercenters on the price index for foods.

This study found further evidence that supercenters have substantial market shares for food products: 15% of all canned beans and 13% of all milk sold in the United States were sold at supercenters. Further, the results of this study indicate that failing to account for supercenters in regional price indexes can bias the cost-of-living index, and that this bias may be much larger at individual market levels. As previously established, probability weights assigned to retail outlets that are used by BLS to select outlets for price data collection are edited such that each outlet has no more than a 25% chance of being selected. However, this study found that supercenters held more than 25% market share for milk and/or canned beans in several markets. Although the focus of this study was not to single out specific supercenter chains, it stands to reason that if there is only one chain of supercenters in those markets, BLS may fail to fully account for the presence of supercenters.

Further, this study found a significant effect of supercenter presence and price indexes for canned beans, but not for milk. Therefore, the effect may be more pronounced for non-perishable products relative to perishable products. A possible explanation is that consumers may acquire more non-perishable goods when prices are low, purchase items in bulk, or otherwise stockpile products that do not expire quickly. This study also found evidence of a relationship between supercenters' share of milk sales and the difference in price indexes for milk excluding and including supercenter prices. Although the effect of supercenters on the price index for canned beans was somewhat small – 0.16 percentage points for every percent increase in supercenter market share – these seemingly small variations are substantial when multiplied by billions of dollars in the context of determining the relative level of federal benefits across various geographic markets per estimated cost-of-living. In addition, several individual markets experienced a substantially greater effect of supercenters on the price index for canned beans, indicating that the gap in price indexes excluding and including supercenter prices at the market level may be substantially higher than the gap at the national level.

Further research should examine the effect of supercenters on price indexes by constructing a better instrument for controlling for endogeneity from supercenter location. The differences between the models for canned beans and milk, in the context of IV, are intriguing. The lack of variation in the price indexes for milk may make it difficult to estimate causality in the context of the effect of competition on the price for milk. This hypothesis could be strengthened by including more food products over a longer period of time. It would also be helpful to determine when supercenters enter particular markets, and to construct a model which accounts for the physical presence (i.e. number of stores per capita) of supercenters. In addition, the use of household scanner data could better estimate the typical consumer's basket, as well as inter-item substitutability in the context of varying prices, but only when the employed price index method accounts for said item substitution. For instance, when prices for canned beans increase, do consumers switch to dried beans? This study did not account for such item substitutability, which is a problem with direct price indexes. Research of this type could further estimate the true cost-of-living index faced by the consumer as preferences, prices, and items change over time in the face of changing incomes and attitudes towards certain types of foods.

References

Arnold, S.J., Handelman, J., Tigert, D.J. (1998). The impact of a market spoiler on consumer preference structures (or, what happens when Wal-Mart comes to town), *Journal of Retailing and Consumer Services*, Volume 5 (1), 1-13.

Becker, R., Hurn, A. (eds.) (2004). Advances in Economics and Econometrics: Theory and Applications. Edward Elgar Publishing. Northampton, Massachusetts.

Board of Trustees. (2007). 2007 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds, March 23, 2007. Accessed: http://www.ssa.gov/history/pdf/tr07summary.pdf.

Bowley, A.L. (1928). Notes on Index Numbers. *Economic Journal* 38, 216-237.

Diewert, W.E. (1978). "Superlative Index Numbers and Consistency in Aggregation," *Econometrica* 46, pp. 883–900.

Diewert, W.E. (1993). ""The Early History of Price Index Research", pp. 33-65 in *Essays in Index Number Theory*, Volume 1, W.E. Diewert and A.O. Nakamura (eds.),Amsterdam: North-Holland.

Diewert, W.E. (1998). Index Number Issues in the Consumer Price Index. *Journal of Economic Perspectives*, 12(1), 47-58.

Diewert, W.E. (1999). Axiomatic and economic approaches to international comparisons, in: A. Heston, R.E. Lipsey (Eds.), International and Interarea Comparisons of Income, Output and Prices, Studies in Income and Wealth, vol. 61, The University of Chicago Press, Chicago (1999), pp. 13–87

Drukker, D. (2003). Testing for serial correlation in linear panel-data models. *The Stata Journal* 3(2), 168–177.

Eltetö, Ö. and P. Köves. (1964), "On a Problem of Index Number Computation Relating to International Comparisons" (in Hungarian), *Statisztikai Szemle* 42, 507-518.

Federal House and Finance Agency (FHFA). House Price Indexes: MSA, State, and Census Division, 1999-2011. Washington, DC: FHFA. Accessed: http://www.fhfa.gov/Default.aspx?Page=87.

Griffith, R., Leibtag, E., Leicester, A., and Nevo, A. (2009). Consumer Shopping Behavior: How Much Do Consumers Save?. *Journal of Economic Perspectives*, vol. 23(2), 99–120.

Gini, C. (1931). "On the Circular Test of Index Numbers", *Metron* 9:9, 3-24.

Harris, J. M., (2005). Using Nielsen Homescan Data and Complex Design Techniques to Analyze Convenience Food Expenditures, No 19344, 2005 Annual meeting, July 24-27, Providence, RI, American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association).

Hausman, J. (2003). Sources of Bias and Solutions to Bias in the Consumer Price Index. *Journal Of Economic Perspectives*, 17(1), 23-44. doi:http://dx.doi.org.prox.lib.ncsu.edu/10.1257/089533003321164930

Hausman, J. and Leibtag, E. (2009). CPI Bias from Supercenters: Does the BLS Know that Wal-Mart Exists?. *NBER Working Paper*.

Hill, R.J. (2004). "Constructing Price Indexes across Space and Time: The Case of the European Union," *American Economic Review*, American Economic Association, vol. 94(5), pages 1379-1410, December.

Hodrick, R.J., Prescott, E.C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, vol. 29(1). pp. 1-16.

Ivancic, L., Diewert, W. E., and Fox, K. J. (2011). Scanner Data, Time Aggregation and the Construction of Price Indexes. *Journal Of Econometrics*, 161(1), 24-35. doi:http://dx.doi.org.prox.lib.ncsu.edu/10.1016/j.jeconom.2010.09.003

Kelley, T.L. (1947). Fundamentals of statistics. Harvard University Press.

Konüs, A.A. (1939). “The Problem of the True Index of the Cost of Living”, *Econometrica* 7, 10-29. [Originally published in 1924]

Kravis, I.B. (1984). “Comparative Studies of National Incomes and Prices”, *Journal of Economic Literature* 22, 1-39.

Kinnucan, H.W., Y. Miao, H. Xiao, and H.M. Kaiser (2001). Effects of advertising on U.S. non-alcoholic beverage demand: Evidence from a two-stage Rotterdam model, in Michael R. Baye, Jon P. Nelson (ed.) Advertising and Differentiated Products (Advances in Applied Microeconomics, Volume 10), Emerald Group Publishing Limited, pp.1-29.

MacDonald, J.M. (1995). “Consumer Price Index Overstates Food Price Inflation”, *Food Review*, 18(3), 28-32.

Pavía-Miralles, J. (2010) "A Survey of Methods to Interpolate, Distribute and Extra- polate Time Series," *Journal of Service Science and Management*, Vol. 3(4). 449-463.

Pigou, A.C. (1920). The Economics of Welfare. Macmillan: London.

Poirier, D.J. (1973). Piecewise Regression Using Cubic Spline. *Journal of the American Statistical Association*, 68(343), 515-524.

Reinsdorf, M. B. (1999). Using Scanner Data to Construct CPI Basic Component Indexes. *Journal Of Business And Economic Statistics*, 17(2), 152-160.

Schultze, C. L., Mackie, C. (eds.) (2002). At What Price? Conceptualizing and Measuring Cost-Of-Living and Price Indexes. Washington, D.C.: National Academy of Press.

Schaffer, M.E., (2010). *xtivreg2*: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models. http://ideas.repec.org/c/boc/bocode/s456501.html

Singh, V.P., Hansen, K.T., Blattbere, R.C. (2006). Market Entry and Consumer Behavior: An Investigation of a Wal-Mart Supercenter. *Marketing Science*, Vol. 25(5), 457-476.

StataCorp. 2011. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.

Swanson D., Hauge S., and Schmidt ML. (1999) "Evaluation of Composite Estimation Methods for Cost Weights in the CPI," *Proceedings of the Section on Survey Research Methods, American Statistical Association*.

Szulc, B. (1964), 'Indices for Multiregional Comparisons,' (in Polish), *Przeglad Statystyczny* 3, 239-254.

United States Department of Labor (USDOL) Bureau of Labor Statistics (BLS) (2007). Handbook of Methods, Chapter 17: The Consumer Price Index. Washington, DC: USDOL. Accessed: http://www.bls.gov/opub/hom/pdf/homch17.pdf.

United States Department of Agriculture (USDA) Economic Research Service (ERS) (2007). Food CPI and Expenditures: CPI Bias. Washington, DC: USDA. Accessed: http://www.ers.usda.gov/Briefing/CPIFoodAndExpenditures/cpibias.htm.

United States Department of Agriculture (USDA) Economic Research Service (ERS) (2008). On the Accuracy of Nielsen Homescan Data. Washington, DC: USDA. Accessed: http://www.ers.usda.gov/Publications/ERR69/ERR69_ReportSummary.pdf.

United States Department of Agriculture (USDA) Economic Research Service (ERS) (2009). Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences. Washington, DC: USDA. Accessed: http://www.ers.usda.gov/publications/ap/ap036/ap036.pdf

United States Department of Agriculture (USDA) Economic Research Service (ERS) (2010). Food Marketing System in the U.S.: Food Retailing. Washington, DC: USDA. Accessed: http://www.ers.usda.gov/Briefing/FoodMarketingSystem/foodretailing.htm.

Walsh, C.M. (1901). The Measurement of General Exchange Value, New York: Macmillan.

Wooldridge, J. M. (2009). Introductory Econometrics: A Modern Approach, South Western, Cengage Learning.

Appendices

Appendix A: BLS Methodology for Calculating the CPI: Extended

BLS employs a multi-step approach to calculating the CPI by conducting several rounds of survey data collection among consumers and retailers. The results from these surveys are used to develop basic (item-level) and aggregate price indexes. This appendix briefly explains how items are aggregated to construct basic indexes and how basic indexes are aggregated across items and geographic areas to construct aggregate indexes.

Item Aggregation

Commodities and services comprise 209 of the total 211 item strata. Of these, 185 are priced. The remaining 24 item strata are difficult to price (and have small weights); therefore, they are priced through imputation. BLS calculates the CPI using a two-stage process. First, item-level elementary indexes are calculated at the geographic area (i.e. PSU group) and item-strata level. These elementary indexes are then aggregated through an extensive weighting process to develop the national CPI.

Basic Indexes

For each ELI within a POPS category, BLS constructs an elementary index using the geometric mean of price ratios (between previous and current periods) weighted by the item's relative importance in the POPS category. The weighting for each POPS category is defined using a series of coefficients against the estimated daily expenditure for the POPS category. The coefficients adjust each item according to its proportion of sales in the broader POPS category, the reciprocal of the proportion of expenditures the ELI represents of the item stratum, the relative importance of the entry-level item to the item stratum and various data irregularities. The formula for the POPS category weights for each item *j* is:

$$W_{j\,POPS} = \left(\hat{P}\hat{Q}\right)_{POPS} \frac{\alpha_j * b_j * \prod Y_j}{M_j * B_j} \qquad \text{(A-1)}$$

where α is the ELI's proportion to total dollar volume of sales in the POPS category for the outlet sampled; $\hat{P}\hat{Q}$ is an estimate of the total daily expenditure for the POPS category derived from the TPOPS survey; *b* is a factor adjusting for the relative importance of the ELI to the item stratum; *Y* is a series of factors used to adjust for data irregularities such as differences in geographic coverage and any special subsampling of the outlets; *M* is the number of usable prices in months *t* and *t-1* for the ELI; and *B* is the proportion of the item stratum's expenditure accounted for by the ELI in the region.

The basic index, commonly called the elementary index, is the average price change in each item-area (think of an area as a city) combination. For each item-area (*a*,*i*) combination, the basic index is constructed using the following general formula[5]:

$$R_{t,t-1}^{a,i} = \prod_{j \in a,i} \left(\frac{P_{j,t}^{a}}{P_{j,t-1}^{a}} \right)^{\left(W_{j\,POPS}^{a} / \sum_{k \in a,i} W_{k\,POPS}^{a} \right)} \qquad \text{(A-2)}$$

where $P_{j,t}$ is the price of item *j* in time *t*. The ratio of prices in adjacent periods, *t* and *t-1*, are "linked" rather than comparing the price in each time *t* with the price in the base period.

Calculating Aggregate Weights

Once elementary indexes are constructed, BLS constructs aggregate weights for each item-area index to construct the aggregate indexes. Using the prices collected for each item strata and the expenditures reported in the CE, BLS estimates the quantity of each item strata to determine the

[5] Prior to 1999, BLS used a Laspeyres formula to construct elementary indexes. For most products, this was revised to use geometric average of price ratios. In some cases, BLS still uses a Laspeyres formula to construct elementary indexes; however, for most products, especially those within the scope of this study, the geometric average of the price ratios is used.

aggregate weights for each item-area combination which are applied to the elementary indexes described earlier. BLS pools the expenditure totals over a 24-month period to minimize variance in quantity estimates. The total expenditures (*TotExp*) for each item-area combination are defined using the following formula:

$$TotExp_{\beta_n}^{a,i} = \left(\hat{P}_{\alpha}\hat{Q}_{\beta_n}\right) \qquad \text{(A-3)}$$

Where $\hat{P}_{\alpha}$ is the estimated price of item *i* in area *a* in the base period (*α*) obtained from a selected outlet in the process described earlier; $\hat{Q}_{\beta}$ is the estimated quantity of item *i* in area *a* in the reference period (*β*) imputed using the total expenditure on item *i* and dividing it by the estimated price. The share of total expenditure $s_{\beta_n}^{a,i}$ for each item *i* is calculated by dividing the total expenditure for item *i* by the sum of expenditures for all items *j* in area *a* in the reference period (*β*)[6].

$$s_{\beta_n}^{a,i} = \frac{TotExp_{\beta_n}^{a,i}}{\sum_{j\in a} TotExp_{\beta_n}^{a,j}} \qquad \text{(A-4)}$$

$$\text{such that} \sum_{a,i} s_{\beta_n}^{a,i} = 1$$

The expenditure totals and expenditure shares, respectively, are then aggregated into larger area groups (*m*):

$$TotExp_{\beta_n}^{m,i} = \sum_{a,i\in m} TotExp_{\beta_n}^{a,i} \qquad \text{(A-5)}$$

[6] When the reference period consists of more than one year, it receives the *n* subscript to denote the year in the reference period. This will be used and further explained later in this paper.

$$s_{\beta_n}^{m,i} = \frac{TotExp_{\beta_n}^{m,i}}{\sum_{a,j\in m} TotExp_{\beta_n}^{m,j}}$$

$$\text{such that} \sum_{m,i}\left(s_{\beta_n}^{m,i}\right) = 1 \qquad \text{(A-6)}$$

A second-stage weight, used to stabilize the expenditure share estimates, is calculated as a weighted average of shares between a larger geographic area *m* and a smaller geographic area within *m*¸ a. This weight *δ* is chosen at the point which minimizes the mean squared error of the expenditure share estimate for the smaller area a. The thought behind this is that estimates from larger areas (i.e. larger samples) have less variability than estimates from smaller areas (i.e. smaller samples) (Swanson, Hauge and Schmidt, 1999) and the share-estimates which are more stable should be given a higher weight, according to the relative efficiency of the estimate for the smaller area. A composite-estimated share of the total expenditures for item *i* in area *a* for the reference period (*β*) is calculated as follows:

$$\hat{s}_{\beta_n}^{a,i} = \delta\left(s_{\beta_n}^{m,i}\right) + (1-\delta)\left(s_{\beta_n}^{a,i}\right) \qquad \text{(A-7)}$$

To obtain the revised expenditure for item *i* in elementary area *a*, the resulting share $\hat{s}_{\beta}^{a,i}$ is multiplied by the sum of all expenditures in the elementary area *a* in the reference period *β:*

$$\widetilde{TotExp}_{\beta_n}^{a,i} = \left(\sum_{a,i\in a} TotExp_{\beta_n}^{a,i}\right)\hat{s}_{\beta_n}^{a,i} \qquad \text{(A-8)}$$

BLS uses a process called "raking" to adjust item-level expenditures in such a way that when summed, they equal the total unadjusted expenditure level for each large region (m) and expenditure class (*e*) level.

$$\left(\hat{P}\hat{Q}\right)_{\beta_n}^{a,i} = \widetilde{TotExp}_{\beta_n}^{a,i} * \left(\frac{\sum_{a,i\in e,m} TotExp_{\beta_n}^{a,i}}{\sum_{a,i\in e,m} \widetilde{TotExp}_{\beta_n}^{a,i}}\right) \qquad \text{(A-9)}$$

The total estimated expenditure level in period β is then determined by averaging the elementary item-area expenditures in each year of the reference period (i.e. when β comprises more than one year):

$$\left(\hat{P}\hat{Q}\right)_{\beta}^{a,i} = \frac{1}{N}\sum_{n=1}^{N}\left(\hat{P}\hat{Q}\right)_{\beta_n}^{a,i} \qquad \text{(A-10)}$$

The next step BLS uses to determine the final aggregation weight for each elementary index is to determine the cost weight in the pivot month. The pivot month, period ν, is the period preceding the month when expenditure weights from reference period β are first used in the CPI. This is calculated by adjusting $\left(\hat{P}\hat{Q}\right)_{\beta}^{a,i}$ by a ratio of the price change from period α to period β and the price change from period α to period ν as follows:

$$\left(\hat{P}_{\nu}\hat{Q}_{\beta}\right)^{a,i} = \left(\hat{P}\hat{Q}\right)_{\beta}^{a,i} * \left(\frac{P_{\nu}^{a,i}}{P_{\alpha}^{a,i}}\right) * \left(\frac{P_{\beta}^{a,i}}{P_{\alpha}^{a,i}}\right)^{-1} \tag{A-11}$$

Finally, the cost weight in the pivot month is adjusted by ratio of the price in period α and the price in period ν to obtain an aggregation weight such that all weights can be summed to one in the base period. The result, $\hat{P}_{\alpha}\hat{Q}_{\beta}$, is an expenditure value with an implicit price in time α and an implicit quantity in time β. This is the aggregation weight AW for item *i* in area *a* in the reference period *β*.

$$AW_{\beta}^{a,i} = \left(\hat{P}_{\alpha}\hat{Q}_{\beta}\right)^{a,i} = \left(\hat{P}_{\nu}\hat{Q}_{\beta}\right)^{a,i} * \left(\frac{P_{\nu}^{a,i}}{P_{\alpha}^{a,i}}\right)^{-1} \tag{A-12}$$

Upper-Level Aggregate Indexes

Once the aggregate weights for each geographic area-item combination are determined, BLS constructs aggregated indexes across subsets across areas and nationally across items. In other words, the national CPI is the weighted average of all basic indexes for all items over all areas[7]. A Laspeyres price index formula uses aggregation weight $AW_{\beta}^{a,i}$ from reference period *β* for item *i* in area *a*, which remains fixed for 2 years, and a ratio of the elementary index from period α to period *t* and elementary index from period ν to period α (taken from equation (1)):

$$L_{0,t}^{A,I} = \sum_{a,i\,\in A,I} \left\{ AW_{\beta}^{a,i} \left(\frac{R_{\alpha,t}^{a,i}}{R_{\alpha,\nu}^{a,i}}\right) \right\} \tag{A-13}$$

[7] For this study, it is important to note that food and beverage indexes are available only at the census region and national levels.

where *A* is an aggregation of price indexes across all areas (e.g. “U.S. city average”) and *I* is an aggregation of price indexes across all items.

Appendix B: Tables

Appendix Table B-1 List of All Markets and Corresponding House Price Index Matches

Market Name	House Price Index Mapping
Albany	State: NY
Atlanta	MSA: Atlanta-Sandy Springs-Marietta, GA
Baltimore	MSA: Baltimore-Towson, MD
Birmingham	State: AL
Boston	State: MA
Buffalo-Rochester	State: NY
Census Division 1: New England	Census Division 1: New England
Census Division 2: Mid Atlantic	Census Division 2: Mid Atlantic
Census Division 3: East North Central	Census Division 3: East North Central
Census Division 4: West North Central	Census Division 4: West North Central
Census Division 5: South Atlantic	Census Division 5: South Atlantic
Census Division 6: East South Central	Census Division 6: East South Central
Census Division 7: West South Central	Census Division 7: West South Central
Census Division 8: Mountain	Census Division 8: Mountain
Census Division 9: Pacific	Census Division 9: Pacific
Charlotte	State: NC
Chicago	MSA: Chicago-Joliet-Naperville, IL (MSAD)
Cincinnati	State: OH
Cleveland	MSA: Cleveland-Elyria-Mentor, OH
Columbus	State: OH
Dallas	MSA: Dallas-Plano-Irving, TX (MSAD)
Denver	MSA: Denver-Aurora-Broomfield, CO
Des Moines	State: IA
Detroit	MSA: Warren-Troy-Farmington Hills, MI (MSAD)
Exurban NY	MSA: Edison-New Brunswick, NJ (MSAD)
Grand Rapids	State: MI
Hartford-New Haven	State: CT
Houston	Houston-Sugar Land-Baytown, TX
Indianapolis	State: IN
Jacksonville	State: FL
Kansas City	State: MO
Little Rock	State: AR

Appendix Table B-1 List of All Markets and Corresponding House Price Index Matches, Continued

Market Name	House Price Index Mapping
Los Angeles	MSA: Los Angeles-Long Beach-Glendale, CA (MSAD)
Louisville	State: KY
Memphis	State: TN
Miami	MSA: Miami-Miami Beach-Kendall, FL (MSAD)
Milwaukee	State: WI
Minneapolis	MSA: Minneapolis-St. Paul-Bloomington, MN-WI
Nashville	State: TN
New Orleans-Mobile	State: LA
Oklahoma City-Tulsa	State: OK
Omaha	State: NE
Orlando	State: FL
Philadelphia	MSA: Philadelphia, PA (MSAD)
Phoenix	MSA: Phoenix-Mesa-Glendale, AZ
Pittsburgh	MSA: Pittsburgh, PA
Portland, Or	State: OR
Raleigh-Durham	State: NC
Richmond	State: VA
Sacramento	State: CA
Salt Lake City	State: UT
San Antonio	State: TX
San Diego	MSA: San Diego-Carlsbad-San Marcos, CA
San Francisco	MSA: Oakland-Fremont-Hayward, CA (MSAD)
Seattle	MSA: Seattle-Bellevue-Everett, WA (MSAD)
St. Louis	MSA: St. Louis, MO-IL
Surburban NY	MSA: Nassau-Suffolk, NY (MSAD)
Syracuse	State: NY
Tampa	MSA: Tampa-St. Petersburg-Clearwater, FL
Urban NY	MSA: New York-White Plains-Wayne, NY-NJ (MSAD)
Washington, DC	MSA: Washington-Arlington-Alexandria, DC-VA-MD-WV (MSAD)

Appendix Table B-2 Percent of Supercenter Transactions at Period 1 (January 2004) and the Difference at Period 36 (December 2006), Canned Beans and Milk

Market	Canned Beans		Milk	
	% Supercenter Transactions, Period 1	PP Increase (Decrease) at Period 36	% Supercenter Transactions, Period 1	PP Increase (Decrease) at Period 36
Market 1	2.8%	0.6	1.8%	0.7
Market 2	1.5%	1.7	1.2%	1.0
Market 3	19.7%	8.8	18.2%	2.6
Market 4	16.8%	4.0	14.2%	2.2
Market 5	7.6%	15.0	13.4%	2.9
Market 6	18.3%	9.3	17.6%	4.7
Market 7	0.0%	1.0	0.0%	1.5
Market 8	0.5%	0.0	0.0%	0.1
Market 9	0.0%	0.0	0.2%	(0.2)
Market 10	0.6%	0.5	0.6%	1.1
Market 11	29.8%	(0.5)	17.7%	(0.7)
Market 12	0.0%	0.0	0.0%	0.0
Market 13	0.0%	3.9	0.0%	1.6
Market 14	11.6%	7.1	14.4%	4.7
Market 15	6.4%	7.8	4.6%	4.7
Market 16	9.2%	(0.6)	6.0%	5.3
Market 17	25.0%	10.6	24.7%	4.1
Market 18	15.2%	8.8	12.7%	6.7
Market 19	2.2%	4.9	0.8%	2.9
Market 20	8.4%	2.1	5.4%	0.6
Market 21	4.8%	(2.8)	5.1%	2.2
Market 22	2.9%	6.3	4.3%	5.0
Market 23	21.4%	11.5	23.1%	9.2
Market 24	1.9%	1.8	2.3%	1.4
Market 25	13.5%	3.9	13.8%	1.5
Market 26	0.7%	4.9	0.6%	3.3
Market 27	0.0%	0.0	0.3%	(0.3)
Market 28	8.3%	5.0	9.9%	3.2
Market 29	14.7%	3.2	12.9%	2.6
Market 30	0.9%	2.2	0.4%	1.6
Market 31	27.0%	12.8	29.8%	13.5
Market 32	2.8%	2.5	1.5%	4.0
Market 33	0.0%	4.3	3.2%	3.4
Market 34	33.7%	21.3	33.9%	14.1
Market 35	26.8%	9.6	27.3%	7.9
Market 36	42.6%	2.9	33.4%	10.5
Market 37	27.8%	12.1	27.2%	11.4

Appendix Table B-2 Percent of Supercenter Transactions at Period 1 (January 2004) and the Difference at Period 36 (December 2006), Canned Beans and Milk, Continued

Market	Canned Beans		Milk	
	% Supercenter Transactions, Period 1	PP Increase (Decrease) at Period 36	% Supercenter Transactions, Period 1	PP Increase (Decrease) at Period 36
Market 38	12.8%	4.7	9.0%	0.2
Market 39	13.9%	12.6	13.2%	15.4
Market 40	20.4%	8.2	18.4%	5.8
Market 41	10.9%	2.4	6.9%	4.3
Market 42	10.2%	3.4	7.6%	4.3
Market 43	15.1%	14.9	6.0%	3.8
Market 44	10.9%	4.1	10.3%	5.0
Market 45	8.8%	4.0	7.0%	4.8
Market 46	5.3%	6.6	5.9%	0.0
Market 47	17.6%	2.7	21.0%	1.7
Market 48	6.9%	23.8	12.6%	9.7
Market 49	14.4%	6.2	13.8%	2.6
Market 50	0.0%	1.0	0.0%	0.9
Market 51	12.6%	8.1	11.2%	5.4
Market 52	4.0%	21.0	8.0%	6.5
Market 53	14.3%	(1.4)	6.1%	2.8
Market 54	7.5%	5.5	8.9%	6.1
Market 55	14.2%	9.3	10.7%	6.9
Market 56	20.5%	5.3	15.6%	7.4
Market 57	23.2%	0.1	18.6%	5.5
Market 58	27.7%	5.5	25.9%	6.0
Market 59	32.0%	5.1	27.6%	6.8
Market 60	23.9%	12.7	20.3%	8.8
Market 61	1.6%	1.2	3.3%	3.7

Appendix Table B-3 Percent of Transactions at Supercenters, Canned Beans and Milk

Market	Canned Beans	Milk
Market 1	3.4	2.3
Market 2	1.7	1.5
Market 3	24.5	19.5
Market 4	18.3	14.9
Market 5	16.9	14.4
Market 6	19.6	20.2
Market 7	0.8	0.6
Market 8	0.1	0.0
Market 9	0.4	0.1
Market 10	1.8	2.1
Market 11	27.1	17.6
Market 12	0.0	0.0
Market 13	1.4	1.1
Market 14	17.6	17.4
Market 15	8.4	6.1
Market 16	9.8	7.1
Market 17	30.2	26.3
Market 18	19.2	14.2
Market 19	3.4	2.2
Market 20	9.6	6.0
Market 21	3.6	6.3
Market 22	6.5	5.4
Market 23	26.4	25.3
Market 24	3.2	2.8
Market 25	16.9	13.9
Market 26	2.0	1.5
Market 27	0.0	0.1
Market 28	10.0	11.0
Market 29	18.1	14.4
Market 30	1.2	1.0
Market 31	34.8	35.6
Market 32	3.9	3.1
Market 33	3.4	3.7
Market 34	44.2	39.6
Market 35	29.3	29.0
Market 36	47.1	39.6
Market 37	37.2	34.6
Market 38	16.0	9.2

Appendix Table B-3 Percent of Transactions at Supercenters, Canned Beans and Milk, Continued

Market	Canned Beans	Milk
Market 39	22.4	18.3
Market 40	18.6	19.8
Market 41	13.0	9.2
Market 42	11.7	10.1
Market 43	13.3	9.1
Market 44	12.1	12.4
Market 45	13.6	9.0
Market 46	6.6	5.6
Market 47	19.2	21.0
Market 48	15.5	14.8
Market 49	20.0	14.1
Market 50	0.7	0.4
Market 51	16.8	14.0
Market 52	14.0	8.8
Market 53	12.1	8.6
Market 54	14.6	12.8
Market 55	18.0	13.5
Market 56	22.7	18.3
Market 57	25.8	22.2
Market 58	28.5	28.7
Market 59	34.8	30.2
Market 60	30.1	22.6
Market 61	3.7	4.7

Appendix Table B-4 Geometric Average of Laspeyres Price Indexes (Base=100) Excluding and Including Supercenters' Prices, Canned Beans, January 2004 – December 2006

Market	P^L Excluding Supercenters' Prices	P^L Including Supercenters' Prices	Difference
Market 1	108.8	108.7	0.0
Market 2	106.6	106.5	0.1
Market 3	111.3	110.8	0.5
Market 4	109.7	108.7	1.0
Market 5	113.9	111.9	2.0
Market 6	108.5	107.6	0.9
Market 7	109.6	109.5	0.1
Market 8	111.0	111.1	(0.1)
Market 9	105.9	105.9	0.0
Market 10	114.5	114.7	(0.2)
Market 11	112.8	112.0	0.8
Market 12	101.2	101.2	0.0
Market 13	110.2	110.2	0.1
Market 14	113.2	111.6	1.6
Market 15	116.9	115.8	1.1
Market 16	107.9	107.3	0.7
Market 17	108.6	107.9	0.7
Market 18	110.8	108.6	2.2
Market 19	113.5	113.3	0.2
Market 20	110.8	110.2	0.6
Market 21	108.5	108.2	0.2
Market 22	112.0	111.3	0.7
Market 23	113.1	111.7	1.3
Market 24	108.0	107.8	0.2
Market 25	107.3	106.7	0.6
Market 26	105.8	105.8	0.0
Market 27	106.9	106.9	0.0
Market 28	110.4	109.8	0.6
Market 29	109.4	108.4	0.9
Market 30	107.7	107.6	0.1
Market 31	111.1	109.6	1.5
Market 32	105.6	105.3	0.3
Market 33	110.1	110.4	(0.3)
Market 34	108.4	109.3	(0.9)
Market 35	113.0	111.2	1.8
Market 36	111.5	108.8	2.8
Market 37	108.2	107.1	1.1
Market 38	106.4	106.4	0.0

Appendix Table B-4 Geometric Average of Laspeyres Price Indexes (Base=100) Excluding and Including Supercenters' Prices, Canned Beans, January 2004 – December 2006, Continued

Market	P^L Excluding Supercenters' Prices	P^L Including Supercenters' Prices	Difference
Market 39	110.3	109.1	1.2
Market 40	108.7	107.4	1.3
Market 41	111.0	110.6	0.4
Market 42	108.7	108.2	0.5
Market 43	110.3	110.0	0.4
Market 44	113.9	111.7	2.2
Market 45	109.9	109.3	0.6
Market 46	120.2	119.3	0.9
Market 47	116.8	114.1	2.7
Market 48	109.9	108.9	1.0
Market 49	115.7	113.3	2.3
Market 50	105.8	106.0	(0.1)
Market 51	110.5	109.7	0.8
Market 52	105.3	106.3	(1.0)
Market 53	106.9	107.2	(0.3)
Market 54	109.9	109.4	0.5
Market 55	110.1	109.6	0.6
Market 56	109.2	108.8	0.3
Market 57	113.8	111.8	2.1
Market 58	110.9	108.8	2.0
Market 59	109.0	108.1	0.9
Market 60	109.3	107.7	1.6
Market 61	106.0	106.1	(0.1)

Appendix Table B-5 Geometric Average of Laspeyres Price Indexes (Base=100) Excluding and Including Supercenters' Prices, Milk, January 2004 – December 2006

Market	P^L Excluding Supercenters' Prices	P^L Including Supercenters' Prices	Difference
Market 1	99.5	99.5	0.1
Market 2	97.2	97.2	0.0
Market 3	102.9	103.0	(0.2)
Market 4	100.4	100.6	(0.1)
Market 5	102.0	102.1	(0.1)
Market 6	101.7	101.5	0.2
Market 7	96.0	95.9	0.1
Market 8	101.1	101.0	0.0
Market 9	100.7	100.7	0.0
Market 10	100.0	99.9	0.1
Market 11	100.5	100.9	(0.4)
Market 12	97.5	97.5	0.0
Market 13	97.2	97.2	(0.0)
Market 14	104.1	103.7	0.3
Market 15	97.2	97.1	0.0
Market 16	100.3	100.2	0.2
Market 17	105.1	105.1	(0.0)
Market 18	102.3	102.9	(0.5)
Market 19	100.3	100.3	0.0
Market 20	100.8	100.8	0.0
Market 21	99.7	99.6	0.1
Market 22	100.9	100.8	0.2
Market 23	104.3	103.4	0.9
Market 24	101.1	101.0	0.0
Market 25	100.5	100.4	0.1
Market 26	99.7	99.8	(0.1)
Market 27	97.2	97.2	(0.0)
Market 28	98.6	98.4	0.2
Market 29	101.9	102.1	(0.1)
Market 30	102.4	102.4	0.0
Market 31	102.3	102.6	(0.3)
Market 32	90.1	90.3	(0.2)
Market 33	101.6	101.4	0.1
Market 34	99.5	100.0	(0.5)
Market 35	100.9	101.0	(0.1)
Market 36	101.0	101.7	(0.6)
Market 37	100.1	100.8	(0.7)
Market 38	98.6	98.7	(0.0)

Appendix Table B-5 Geometric Average of Laspeyres Price Indexes (Base=100) Excluding and Including Supercenters' Prices, Milk, January 2004 – December 2006, Continued

Market	P^L Excluding Supercenters' Prices	P^L Including Supercenters' Prices	Difference
Market 39	103.7	102.7	1.0
Market 40	97.5	97.7	(0.3)
Market 41	98.5	98.3	0.2
Market 42	102.0	101.7	0.3
Market 43	98.1	98.3	(0.2)
Market 44	102.0	101.8	0.2
Market 45	100.6	100.3	0.3
Market 46	99.6	99.7	(0.1)
Market 47	104.1	104.3	(0.2)
Market 48	101.7	101.1	0.6
Market 49	105.8	105.0	0.8
Market 50	97.1	97.1	0.0
Market 51	101.7	101.6	0.2
Market 52	98.6	98.4	0.1
Market 53	97.0	97.0	0.0
Market 54	99.9	99.7	0.2
Market 55	102.2	102.3	(0.1)
Market 56	101.9	102.5	(0.6)
Market 57	102.4	102.4	0.1
Market 58	100.1	100.0	0.1
Market 59	99.8	99.8	(0.0)
Market 60	100.7	100.6	0.0
Market 61	101.7	101.9	(0.2)

Appendix Table B-6 Regression of Difference in Laspeyres Price Indexes Excluding and Including Supercenter Prices on the Market Level Percent of Transactions at Supercenters for Canned Beans, Fixed-Effects (FE) and First-Order Autoregression (AR1)

Variable	$\ddot{P}_{it}^{L} - P_{it}^{L}$, Canned Beans			
	FE		AR(1)	
% Supercenters *Market 1	0.0818	(0.3500)	0.0982	(0.3510)
% Supercenters *Market 2	0.1820	(0.2780)	0.1700	(0.2930)
% Supercenters *Market 3	0.0857	(0.1200)	0.0468	(0.1330)
% Supercenters *Market 4	0.0550	(0.0701)	0.0783	(0.0716)
% Supercenters *Market 5	**0.1630*** **	**(0.0348)**	**0.1540*** **	**(0.0369)**
% Supercenters *Market 6	**0.1200** **	**(0.0460)**	**0.1380** **	**(0.0465)**
% Supercenters *Market 7	0.1240	(0.3180)	0.1120	(0.3430)
% Supercenters *Market 8	-0.0605	(0.9760)	-0.0492	(1.0410)
% Supercenters *Market 9	0.1460	(0.5850)	0.1370	(0.6130)
% Supercenters *Market 10	0.1300	(0.1010)	0.1140	(0.1030)
% Supercenters *Market 11	0.0561	(0.0510)	0.0476	(0.0520)
% Supercenters *Market 12	omitted		omitted	
% Supercenters *Market 13	0.2380	(0.2830)	0.2540	(0.3090)
% Supercenters *Market 14	**0.2840** **	**(0.1080)**	**0.2790* **	**(0.1230)**
% Supercenters *Market 15	**0.2360** **	**(0.0857)**	**0.2260* **	**(0.0890)**
% Supercenters *Market 16	0.1430	(0.0751)	0.1420	(0.0793)
% Supercenters *Market 17	-0.0309	(0.0658)	-0.0081	(0.0714)
% Supercenters *Market 18	**0.3350*** **	**(0.0775)**	**0.3550*** **	**(0.0849)**
% Supercenters *Market 19	0.1980	(0.1780)	0.1890	(0.1910)
% Supercenters *Market 20	0.1460	(0.1440)	0.1440	(0.1520)
% Supercenters *Market 21	0.2030	(0.1350)	0.1590	(0.1330)
% Supercenters *Market 22	0.1290	(0.0996)	0.1190	(0.1100)
% Supercenters *Market 23	0.1380	(0.0867)	0.1570	(0.0943)

Appendix Table B-6 Regression of Difference in Laspeyres Price Indexes Excluding and Including Supercenter Prices on the Market Level Percent of Transactions at Supercenters for Canned Beans, Fixed-Effects (FE) and First-Order Autoregression (AR1), Continued

Variable	$\ddot{P}_{it}^{L} - P_{it}^{L}$, Canned Beans			
	FE		AR(1)	
% Supercenters *Market 24	0.1880	(0.2330)	0.1850	(0.2430)
% Supercenters *Market 25	**0.1870***	**(0.0889)**	**0.1660**	**(0.0929)**
% Supercenters *Market 26	0.0899	(0.1580)	0.0901	(0.1600)
% Supercenters *Market 27	omitted		omitted	
% Supercenters *Market 28	0.0483	(0.1580)	0.0813	(0.1650)
% Supercenters *Market 29	**0.2830****	**(0.1060)**	**0.2540***	**(0.1160)**
% Supercenters *Market 30	0.0985	(0.297)	0.0928	(0.305)
% Supercenters *Market 31	**0.2200****	**(0.0743)**	**0.2720****	**(0.0849)**
% Supercenters *Market 32	-0.0010	(0.2100)	-0.00278	(0.2250)
% Supercenters *Market 33	-0.0845	(0.0956)	-0.0798	(0.0995)
% Supercenters *Market 34	**-0.0972***	**(0.0424)**	**-0.1100***	**(0.0476)**
% Supercenters *Market 35	**0.1350****	**(0.0475)**	**0.1320****	**(0.0506)**
% Supercenters *Market 36	**0.2100*****	**(0.0411)**	**0.2290*****	**(0.0444)**
% Supercenters *Market 37	0.0569	(0.0557)	0.0515	(0.0625)
% Supercenters *Market 38	-0.0506	(0.0861)	-0.0452	(0.0927)
% Supercenters *Market 39	-0.0172	(0.0455)	-0.0364	(0.0502)
% Supercenters *Market 40	**0.2120*****	**(0.0433)**	**0.1970*****	**(0.0436)**
% Supercenters *Market 41	0.0036	(0.1110)	0.0171	(0.1190)
% Supercenters *Market 42	0.1290	(0.1060)	0.1150	(0.1140)
% Supercenters *Market 43	**0.1870*****	**(0.0430)**	**0.2080*****	**(0.0449)**
% Supercenters *Market 44	**0.3950*****	**(0.1060)**	**0.3780****	**(0.1150)**
% Supercenters *Market 45	**0.2410****	**(0.0813)**	**0.2190****	**(0.0829)**
% Supercenters *Market 46	**0.1880***	**(0.0898)**	**0.1900***	**(0.0929)**
% Supercenters *Market 47	**0.2720****	**(0.1020)**	**0.2240***	**(0.1020)**

Appendix Table B-6 Regression of Difference in Laspeyres Price Indexes Excluding and Including Supercenter Prices on the Market Level Percent of Transactions at Supercenters for Canned Beans, Fixed-Effects (FE) and First-Order Autoregression (AR1), Continued

	$\ddot{P}^L_{it} - P^L_{it}$, Canned Beans			
Variable	**FE**		**AR(1)**	
% Supercenters *Market 48	**0.1820*** **	**(0.0266)**	**0.1690*** **	**(0.0292)**
% Supercenters *Market 49	**0.2480*** **	**(0.0537)**	**0.2230*** **	**(0.0539)**
% Supercenters *Market 50	0.1960	(0.3330)	0.2250	(0.3560)
% Supercenters *Market 51	0.1220	(0.0771)	0.1120	(0.0844)
% Supercenters *Market 52	**0.0934* **	**(0.0411)**	**0.1240** **	**(0.0462)**
% Supercenters *Market 53	0.0124	(0.1430)	0.0184	(0.1530)
% Supercenters *Market 54	**0.3120*** **	**(0.0500)**	**0.3310*** **	**(0.0517)**
% Supercenters *Market 55	**0.2580*** **	**(0.0667)**	**0.2430** **	**(0.0740)**
% Supercenters *Market 56	**0.3330*** **	**(0.0876)**	**0.3000** **	**(0.0942)**
% Supercenters *Market 57	**0.3930*** **	**(0.0709)**	**0.3950*** **	**(0.0760)**
% Supercenters *Market 58	0.1250	(0.0795)	0.1270	(0.0853)
% Supercenters *Market 59	0.0592	(0.0695)	0.0650	(0.0770)
% Supercenters *Market 60	0.0738	(0.0633)	0.0962	(0.0681)
% Supercenters *Market 61	**0.1300* **	**(0.0549)**	**0.1440** **	**(0.0550)**
Intercept	-1.534***	(0.1740)	-1.530***	(0.1690)
N	2196		2135	
F	6.894		6.122	
R^2 (overall)	0.171		0.162	

Standard errors are reported in parentheses.
* Denotes significance at the p<0.05 level.
** Denotes significance at the p<0.01 level.
*** Denotes significance at the p<0.001 level.

Appendix Table B-7 Regression of Difference in Laspeyres Price Indexes Excluding and Including Supercenter Prices on the Market Level Percent of Transactions at Supercenters for Milk, Fixed-Effects (FE) and First-Order Autoregression (AR1)

Variable	$\ddot{P}_{it}^{L} - P_{it}^{L}$, Milk			
	FE		AR(1)	
% Supercenters *Market 1	0.1520	(0.2000)	0.1310	(0.2220)
% Supercenters *Market 2	-0.0370	(0.1910)	-0.0801	(0.2280)
% Supercenters *Market 3	-0.1290	(0.0857)	-0.0448	(0.0935)
% Supercenters *Market 4	-0.0489	(0.0479)	-0.0308	(0.0454)
% Supercenters *Market 5	0.0655	(0.0421)	0.0497	(0.0422)
% Supercenters *Market 6	**0.0702**	**(0.0358)**	**0.0858***	**(0.0428)**
% Supercenters *Market 7	0.2020	(0.1680)	0.2040	(0.2350)
% Supercenters *Market 8	0.0869	(1.4030)	0.0801	(1.1710)
% Supercenters *Market 9	0.0246	(0.8850)	0.0271	(0.7920)
% Supercenters *Market 10	0.0519	(0.0913)	0.0654	(0.1120)
% Supercenters *Market 11	-0.0261	(0.0412)	0.0224	(0.0429)
% Supercenters *Market 12	0.0098	(6.6010)	0.0097	(7.0270)
% Supercenters *Market 13	0.0304	(0.2120)	0.0320	(0.3130)
% Supercenters *Market 14	0.0052	(0.0533)	0.0643	(0.0699)
% Supercenters *Market 15	-0.0195	(0.0453)	0.0011	(0.0592)
% Supercenters *Market 16	0.0783	(0.0586)	0.0814	(0.0682)
% Supercenters *Market 17	-0.1380***	(0.0387)	-0.0903	(0.0470)
% Supercenters *Market 18	**-0.2070*****	**(0.0380)**	**-0.1590****	**(0.0506)**
% Supercenters *Market 19	0.1040	(0.0872)	0.1680	(0.1170)
% Supercenters *Market 20	0.0615	(0.1800)	0.0580	(0.1750)
% Supercenters *Market 21	0.0332	(0.0860)	0.0544	(0.0984)
% Supercenters *Market 22	**0.1580*****	**(0.0444)**	**0.1600***	**(0.0622)**
% Supercenters *Market 23	**0.1570*****	**(0.0279)**	**0.1040****	**(0.0357)**

Appendix Table B-7 Regression of Difference in Laspeyres Price Indexes Excluding and Including Supercenter Prices on the Market Level Percent of Transactions at Supercenters for Milk, Fixed-Effects (FE) and First-Order Autoregression (AR1), Continued

Variable	$\ddot{P}_{it}^{L} - P_{it}^{L}$, Milk			
	FE		AR(1)	
% Supercenters *Market 24	0.0245	(0.1730)	0.0231	(0.2060)
% Supercenters *Market 25	0.0900	(0.0748)	0.0847	(0.0704)
% Supercenters *Market 26	-0.1110	(0.0649)	-0.1310	(0.0844)
% Supercenters *Market 27	0.1360	(0.4440)	0.1250	(0.4120)
% Supercenters *Market 28	0.1710*	(0.0731)	0.1850	(0.0986)
% Supercenters *Market 29	-0.0549	(0.0817)	-0.0103	(0.0984)
% Supercenters *Market 30	0.0061	(0.1140)	0.0795	(0.1340)
% Supercenters *Market 31	-0.0556*	(0.0228)	-0.0120	(0.0315)
% Supercenters *Market 32	-0.1090	(0.0681)	-0.0956	(0.0976)
% Supercenters *Market 33	0.1560*	(0.0723)	0.1500	(0.0807)
% Supercenters *Market 34	0.0166	(0.0201)	0.0505	(0.0275)
% Supercenters *Market 35	-0.0217	(0.0242)	-0.0324	(0.0308)
% Supercenters *Market 36	**-0.0931*****	**(0.0211)**	**-0.0782***	**(0.0306)**
% Supercenters *Market 37	**-0.0921*****	**(0.0205)**	**-0.0741***	**(0.0288)**
% Supercenters *Market 38	-0.1010	(0.1060)	-0.0144	(0.1230)
% Supercenters *Market 39	**0.1590*****	**(0.0218)**	**0.1500*****	**(0.0295)**
% Supercenters *Market 40	**0.3060*****	**(0.0525)**	**0.2270*****	**(0.0550)**
% Supercenters *Market 41	0.0545	(0.0689)	0.0783	(0.0863)
% Supercenters *Market 42	**0.1730****	**(0.0583)**	**0.1740***	**(0.0768)**
% Supercenters *Market 43	-0.0585	(0.0486)	-0.0106	(0.0494)
% Supercenters *Market 44	**0.1500*****	**(0.0371)**	**0.1380****	**(0.0527)**
% Supercenters *Market 45	0.0689	(0.0700)	-0.0389	(0.0636)
% Supercenters *Market 46	-0.0966	(0.0770)	-0.0684	(0.0816)
% Supercenters *Market 47	**0.2370*****	**(0.0667)**	**0.2080****	**(0.0779)**

Appendix Table B-7 Regression of Difference in Laspeyres Price Indexes Excluding and Including Supercenter Prices on the Market Level Percent of Transactions at Supercenters for Milk, Fixed-Effects (FE) and First-Order Autoregression (AR1), Continued

	$\ddot{P}_{it}^{L} - P_{it}^{L}$, Milk			
Variable	**FE**		**AR(1)**	
% Supercenters *Market 48	**0.0720****	**(0.0229)**	**0.0790****	**(0.0275)**
% Supercenters *Market 49	0.1980***	(0.0501)	0.0628	(0.0519)
% Supercenters *Market 50	0.0562	(0.2230)	0.0328	(0.3100)
% Supercenters *Market 51	0.0194	(0.0398)	0.0006	(0.0551)
% Supercenters *Market 52	-0.0258	(0.0416)	-0.0263	(0.0489)
% Supercenters *Market 53	0.0070	(0.0471)	-0.0084	(0.0559)
% Supercenters *Market 54	-0.0077	(0.0276)	-0.0076	(0.0367)
% Supercenters *Market 55	-0.0293	(0.0556)	-0.0217	(0.0772)
% Supercenters *Market 56	-0.0057	(0.0389)	0.0066	(0.0531)
% Supercenters *Market 57	-0.0436	(0.0386)	-0.0531	(0.0489)
% Supercenters *Market 58	**-0.2730*****	**(0.0456)**	**-0.2140*****	**(0.0607)**
% Supercenters *Market 59	-0.1200***	(0.0355)	-0.0488	(0.0485)
% Supercenters *Market 60	0.0565	(0.0486)	0.0268	(0.0583)
% Supercenters *Market 61	0.0811	(0.0725)	0.1340	(0.0852)
Intercept	-0.0147	(0.0902)	-0.1710*	(0.0697)
N	2196		2135	
F	6.152		2.669	
R^2 (overall)	0.096		0.080	

Standard errors are reported in parentheses.
* Denotes significance at the p<0.05 level.
** Denotes significance at the p<0.01 level.
*** Denotes significance at the p<0.001 level.