Journal of Agricultural and Resource Economics 44(1):117–140 ISSN 1068-5502

A Comparison of Food Demand Estimation from Homescan and Consumer Expenditure Survey Data

Tullaya Boonsaeng and Carlos E. Carpio

This study evaluates the differences between the Exact Affine Stone Index (EASI) demand model estimates obtained using Consumer Expenditure Survey (CEX) data and Nielsen Homescan data. Results indicated that elasticities obtained from CEX and Homescan data–based demand models differ not only statistically but also economically. Own-price elasticities obtained from the CEX data–based demand model were more inelastic than those obtained from the demand model estimated using Homescan data. Further, differences between expenditure elasticities did not follow a specific pattern. We found evidence suggesting that the main source of differences is the price index used for the estimation.

Key words: food demand, food demand elasticities, Nielsen Homescan

Introduction

Demand models play an important role in the analysis and measurement of consumer preferences and the evaluation of agricultural and food policies. For example, the responsiveness of the quantity of a good demanded to a change in its price is measured by the own-price elasticity, a common output of demand model estimation. Demand models can also be used to analyze policy issues such as the welfare costs of a tax reform (e.g., Banks, Blundell, and Lewbel, 1997) or the effect of health information on demand (e.g., Tonsor, Mintert, and Schroeder, 2010).

Applied demand analysis requires two principal elements: a parametric or semiparametric demand system derived from the theory of consumer behavior (e.g., the Almost Ideal Demand System or the Rotterdam Model) and a dataset, which is used to estimate the model's parameters. An extensive literature focuses on the development of highly flexible demand forms (e.g., Piggott, 2003; Barnett and Yue, 1988). Several types of data have been used for the econometric estimation of the models, including time series, cross-sectional, and panel data; however, only a small number of demand studies have evaluated the quality, statistical properties, or the effects of the data on the final results of their analyses. Nielsen Homescan is one of the few data sources that has been evaluated (Harris and Blisard, 1995; Zhen et al., 2009; Einav, Leibtag, and Nevo, 2010). This proprietary dataset tracks consumers' grocery purchases and is collected by the Nielsen Company. Zhen et al. (2009) compared household expenditures based on Homescan data and the data from the Consumer Expenditure Survey (CEX), collected by the U.S Bureau of Labor Statistics (BLS). They concluded that the datasets report substantial differences in household food expenditures despite having comparable demographic compositions. Einav, Leibtag, and Nevo (2010) compared data recorded by a retailer through its loyalty program with Homescan data for the same group of consumers. They

---

Tullaya Boonsaeng is a research assistant professor and Carlos E. Carpio is an associate professor in the Department of Agricultural and Applied Economics at Texas Tech University.
Funding for this research was provided by a grant from the Agriculture and Food Research Initiative from the USDA National Institute of Food and Agriculture (contract 21A409-B51377-200). The authors gratefully acknowledge Jessica Todd from the USDA Economic Research Service for sharing code to calculate food price indices using Nielsen Homescan data. We also thank Kristyn Leffler for research assistance and two anonymous reviewers for comments and suggestions.
Review coordinated by Dragan Miljkovic.

found that recording errors in prices were more prevalent than errors in trip information and product and quantity information, but concluded that the degree of measurement error in those price data was comparable to that found in other datasets commonly used in social sciences. Einav, Leibtag, and Nevo also illustrated the effects of measurement error in two applications: price regressions and demand estimation. In both cases, they found significant differences between the regression results obtained using Homescan data and those based on retailer data. These findings are problematic, as a large and growing number of researchers use scanner data to investigate research questions related to food demand, diet, and health (e.g., Hausman and Leibtag, 2007; Kuchler, Tegene, and Harris, 2005).[1]

An alternative source of consumer food expenditure data at the household level is the CEX, which has been used widely in the analysis of food consumed away from home for U.S. households (Byrne, Capps, and Saha, 1996; Jensen and Yen, 1996; Stewart and Yen, 2004; Zan and Fan, 2010). One drawback of the CEX data is that they contain information only on expenditures, not on prices and quantities, which precludes their use for estimating conventional demand models, including price and income effects. However, recent advances in the econometric literature have shown that, in some cases, it is possible to overcome this limitation by constructing household-specific price indices (Stone–Lewbel [SL] prices) derived from regional price indices (Hoderlein and Mihaleva, 2008).

The main objective of this study is to evaluate the potential of publicly available datasets and state-of-the-art econometric methods in lieu of the proprietary Homescan data. A secondary objective is to cross-validate the results of food demand analyses (i.e., elasticity values) using alternative data sources. Thus, this study was designed to answer the following questions: Are there any differences between demand model estimates obtained using Homescan data and BLS data (CEX and price indices)? If there are, what are the main sources of those differences? The results of this study have important implications for researchers in the field of consumer demand analysis. If the results of demand model estimations are found to be invariant regardless of the dataset used, it will imply that demand analysis results obtained using data sources available publicly are as good (or as bad) as those obtained using Homescan data, eliminating the need for researchers to obtain access to the private, and in some cases costly, Homescan data.[2] On the other hand, if demand model estimation results differ across datasets, this will require further investigation of the sources of variation and the procedures needed to ameliorate estimation biases.

We make two main contributions to the literature. This is the first study to compare demand estimation results using two nationally representative U.S. household food expenditure datasets, one proprietary and one public. The second contribution is to evaluate the effect of alternative price indices in demand estimation results, which has also not been studied previously.

Approach

This study was conducted in two phases to enable a comparison of the differences between demand model estimates obtained using Homescan data and CEX data. Phase 1 compared the results of demand estimation using the two data sources, and Phase 2 explored the source of the differences.

---

[1] Lusk and Brooks (2011) also compared demand estimation results obtained from two household scanning panels (Nielsen Company and Information Resources, Inc.) and a random sample of the U.S. population. However, they based their analyses on choice experiments rather than reporting data on purchases or expenditures. Cornelsen et al. (2016) conducted a meta-analysis to explore the influence of various methodological aspects on demand estimates. Although the data source was one of the factors considered, it was not included in the final model specifications.

[2] There will still be cases in which CEX data cannot substitute for Homescan data, especially in studies at the level of brand or variety. With respect to the cost of the Homescan data, the Nielsen Company, in partnership with the University of Chicago Booth School of Business, now makes the Consumer Data Panel available to researchers at a cost for one faculty member of $3,000 for 3 years (see https://research.chicagobooth.edu/nielsen/datasets). CEX data are available freely at http://www.bls.gov/cex/.

Phase 1

Phase 1 was carried out in three major steps:

1. Estimate a demand system using CEX expenditure data and BLS Consumer Price Index (CPI)–based SL prices (explained in detail later).
2. Estimate a demand system for the same group of goods and the same demand system used in Step 1.1 using Homescan data with Fisher price indices.
3. Compare the elasticities obtained in Steps 1.1 and 1.2.

Phase 2

Phase 2 of the study investigated the sources of differences found in Phase 1. Because we used the same estimation procedures in both cases, there were only two main potential sources of differences: i) the data source and ii) the price indices. To explore the role of these two sources of error, we used the following multistep approach:

1. Construct artificial CEX-type data (for household expenditures) and regional price indices using Homescan data.[3] As regional price indices, we used the Quarterly Food-at-Home Price (QFAHP) Index constructed by the USDA Economic Research Service (U.S. Department of Agriculture, 2016).
2. Estimate the demand system using data constructed in Step 2.1.
3. Compare the elasticities obtained in Steps 2.2 and 1.2 (i.e., compare the elasticities estimated using Homescan data and Fisher price indices against those using constructed household SL prices).
4. Compare and evaluate the out-of-sample forecasting performance of models estimated using Fisher and SL prices.

Steps 2.1–2.4 allowed us to explore the effect of the price indices used; thus, the procedure is an approach to validate Hoderlein and Mihaleva's (2008) demand model estimation procedure using SL prices. While these authors provided evidence that using SL prices increases price variation relative to using regional price indices, which in turn results in more plausible signs of the demand coefficients and precision of parameter estimates, the procedure has not been validated using other price indices.

Data

The data used in this paper were derived from the BLS CEX and CPI and from the Nielsen Homescan Panels. The analysis was based on the annual CEX and Nielsen Homescan surveys for 2002–2006. We included only these years because the 2007 Homescan data include less information about random-weight food purchases. Using the USDA-established nutrition-based guidelines in the ERS QFAHP database (QFAHPD), we considered eight commodity groups: i) cereal and bakery products; ii) meats and eggs; iii) dairy; iv) fruits and vegetables; v) nonalcoholic beverages; vi) fats and oils; vii) sugar and other sweets, and viii) miscellaneous foods. Tables 1 and 2 provide detailed information on these food groups and their corresponding subgroups for both datasets.

[3] Because the Homescan data contain more information than do the CEX data, we assumed that Homescan is the "true" data-generating process. CEX-type data on household expenditures can be constructed easily using the quantity and price data available in the Homescan data.

Table 1. CEX Data Food Groups and Subgroups

Commodity Groups	Subgroups
Cereals and bakery	Cereals
	Bakery products
Meat and eggs	Beef and veal
	Pork
	Poultry
	Fish and seafood
	Eggs
	Other meats
Dairy	Milk
	Cheese and related products
	Ice cream and related products
	Other dairy and related products
Fruits and vegetables	Fresh fruits
	Fresh vegetables
	Processed fruits and vegetables
Nonalcoholic beverages	Juices and nonalcoholic drinks
	Beverage materials including coffee and tea
Fats and oils	Butter and margarine
	Salad dressing
	Other fats and oils including peanut butter
Sugars and other sweets	Sugar and artificial sweeteners
	Candy and chewing gum
	Other sweets
Miscellaneous foods	Soups
	Frozen and freeze dried prepared foods
	Snacks
	Spices, seasonings, condiments, sauces
	Baby food
	Other miscellaneous foods

Notes: Subgroup prices are deflated using the CPI and subsequently normalized using mean values of the price series.

The CEX data come from a 2-week diary of participating households' daily expenditures on specific food products, which are totaled to obtain bi-weekly expenditures. The survey also collects information on household characteristics. The pooled CEX cross-sectional data used in the analyses include 36,005 households and exclude households with missing values for sociodemographic variables, households that only reported expenditures for 1 week, and households reporting zero total expenditures.

The Nielsen Homescan program provides households throughout the continental United States with a handheld scanner to record all food purchases made from all outlets as they occur. We restricted this analysis to only the subset of households that also recorded items without a universal product code (UPC), such as fresh fruits or vegetables and in-store packaged breads and meats

Table 2. Homescan Data Food Groups and Subgroups

Commodity Group	Subgroups for Construction of SL Price	Subgroup Description from QFAHPD-1[a]	No. of Aggregate Products[b]
Cereal and bakery goods	1) Cereal, bread, and rolls	CB1: Whole grain bread, rolls, rice, pasta	263
		CB2: Whole grain flour and mixes	
		CB3: Whole grain frozen/ready to cook	
		CB4: Refined grain bread, rolls, rice, pasta	
		CB5: Refined grain flour and mixes	
		CB6: Refined grain frozen/ready to cook	
	2) Bakery products except bread and rolls	CB7: Baked good mixes	
		CB8: Bakery items, ready to eat	
		CB9: Packaged goods (cookies)a	
		CB10: Frozen desserts[a]	
Meat and eggs	1) Fresh/frozen low-fat meat	ME1: Fresh/Frozen low fat meat	209
	2) Fresh/frozen regular fat meat	ME2: Fresh/frozen regular fat meat	
	3) Canned meats, poultry and fish	ME3: Canned meat	
		ME5: Canned poultry	
		ME7: Canned fish	
	4) Poultry	ME4: Fresh/frozen poultry	
	5) Fish	ME6: Fresh/frozen fish	
	6) Eggs	ME8: Eggs	
Dairy and related products	1) Milk	D1: Low fat milk	137
		D4: Regular fat milk	
	2) Cheese	D2: Low fat cheese	
		D5: Low/regular fat cheese	
	3) Ice cream	D7: Ice cream[a]	
	4) Yogurt and other dairy	D3: Low fat yogurt and other dairy	
		D6: Regular fat yogurt and other dairy	
Fruits and vegetables	1) Fresh/frozen fruit	FV1: Fresh/frozen fruit	414
	2) Fresh/frozen vegetables	FV3: Fresh/frozen dark green vegetables	
		FV5: Fresh/frozen orange vegetables	
		FV7: Fresh/frozen starchy vegetables	
		FV8: Frozen/dried legumes	
		FV11: Fresh/frozen other vegetables	
		FV13: Fresh/frozen select nutrient vegetables	
		FV15: Mixed and unspecified fruit/vegetables[a]	
	3) Canned fruit and vegetables	FV2: Canned fruit	
		FV4: Canned dark green vegetables	
		FV6: Canned orange vegetables	
		FV9: Canned starchy vegetables	
		FV10: Canned legumes	
		FV12: Canned other vegetables	
		FV14: Canned select nutrient vegetables	
Nonalcoholic beverages and beverage materials	1) Juices and nonalcoholic drinks	NB1: Nonalcoholic carbonated beverages	141
		NB2: Noncarbonated caloric beverages	
		NB4: Fruit juice	
	2) Water	NB3: Water	
	3) Coffee and tea	NB5: Coffee and tea[a]	
Fats and oils	1) Oils	FO1: Oils	74
	2) Solid fats	FO2: Solid Fats	
	3) Nut butters	FO3: Nut butters[a]	
	4) Salad dressings and spreads	FO4: Salad dressings[a]	

Continued on next page...

Table 2. – continued from previous page

Commodity Group	Subgroups for Construction of SL Price	Subgroup Description from QFAHPD-1[a]	No. of Aggregate Products[b]
Sugar and other sweets	1) Raw sugars 2) Packaged sweet goods (candy) 3) Jams, jellies, preserves and other sweets	SO1: Raw sugars SO2: Packaged sweet goods (candy bars)[a] SO3: Other sweets (e.g., jams, jellies, preserves and other sweets)[a]	88
Miscellaneous foods	1) Raw and processed nuts and seeds and packaged snacks 2) Frozen entrees and sides 3) Canned soups and sauces 4) Packaged/ready-to-cook meals and sides and ready-to-eat deli items (hot and cold) 5) Baby food 6) Spices, seasonings, condiments, olives, pickles, relishes	MISC1: Raw and processed nuts and seeds MISC4: Packaged snacks MISC2: Frozen entrees and sides MISC3: Canned soups, sauces, prepared foods MISC5: Ready-to-cook meals and sides MISC6: Ready-to-eat deli items (hot and cold) MISC7: Baby food[a] MISC8: Salt, other seasoning and spices, olives, pickles, relishes[a]	458

Notes: [a]Indicates subgroups without prices in the original QFAHPD-1 dataset.
[b]An aggregate product includes all flavors and sizes for a branded product of certain type. For example, "Dannon-branded reduced-fat yogurt" includes Dannon individual-size fat-free blueberry yogurt and Dannon quart-size reduced-fat strawberry yogurt.
Source: U.S. Department of Agriculture (2016).

(i.e., the "fresh foods panel"). Failure to account for additional non–UPC purchases would bias the total expenditure of a household downward. As there is a sizeable time burden on participating households, the retention rate for households within the Homescan panel varies;[4] thus, data were treated as cross-sectional rather than panel because of variation in participation in the dataset over time. Further, to make the data comparable to the 2-week CEX data, the 2-week data from each household-specific year were selected randomly to comprise the bi-weekly dataset for Homescan data. Further sample reduction occurred because the analysis was limited to households in urban and suburban locations with purchases in at least one commodity group. This yielded a total of 35,421 year-specific household observations.

Stone–Lewbel Price Index

Following Hoderlein and Mihaleva (2008), we constructed household-level prices (SL price indices) using regional price indices. If the between-group utility function is weakly separable and the within-group subutility functions are Cobb–Douglas, then it can be shown that the SL price (v_i^l) index corresponding to the commodity group i and household l is

$$v_i^l = \frac{1}{k_i}\prod_{j=1}^{s_i}\left(\frac{p_{ij}}{w_{ij}^l}\right)^{w_{ij}^l} \tag{1}$$

with a scaling factor k_i, given by $k_i = \prod_{j=1}^{s_i} \bar{w}_{ij}^{-\bar{w}_{ij}}$, where s_i is the number of goods in commodity group i, p_{ij} is the (regional) monthly price of the jth good in commodity group i, $w_{ij}^l = p_{ij}^l q_{ij}^l / y_i^l$ is household l within group budget share of the jth good in group i, $\bar{w}_{ij}$ is the budget share of good j in group i of the reference household.[5] Equation (1) implies that household-level price indices for each commodity group can be calculated using individual goods budget shares (w_{ij}^l) and price indices (p_{ij}).

[4] In the final dataset of 11,980 households (35,421 year-specific household records), 25% of households were included for all 5 years, followed by 17%, 14%, 17%, and 27% for 4 years, 3 years, 2 years, and 1 year, respectively. While the inclusion years are consecutive, years are not necessarily the same for all households.
[5] The reference household is the household with average budget shares.

Table 3. Household Composition Variables and Characteristics

Variable	Definition
Continuous variables	
Family size*•	
Total food expenditures	Bi-weekly food expenditures
Dummy variables (yes = 1, no = 0)	
No college*•	Education level of the reference person
Some college*•	
College	
Northeast*•	Region of residence
Midwest*•	
South*•	
West	
< 25*•	Age of the reference person (in years)
≥25–30*•	
≥30–40*•	
≥40–50*•	
≥50–60*•	
>60	
White*•	Racial group of the reference person
Black*•	
Asian*•	
Other	Reference person does not self-identify as white, black or Asian
2002•	Year in which the survey was collected
2003•	
2004•	
2005•	
2006	
Hispanic*	Reference person self-identifies as Hispanic

Notes: Single asterisk (*) indicates demographic variables used in the LA/EASI model. Single bullet (•) indicates demographic variables used to regress and impute SL prices.

The SL price index is undefined when one or more of the subgroup commodity shares, w_{ij}^{l}, is equal to 0. Hoderlein and Mihaleva (2008) avoided this problem by dropping observations with w_{ij}^{l} equal to 0. Although plausible for lower levels of censoring, this solution severely restricts datasets with higher censoring levels. Therefore, this analysis adopted the regression imputation approach employed in demand studies of cross-sectional data with censored expenditures that use unit values as a proxy for prices (see Cox and Wohlgenant, 1986; Alfonzo and Peterson, 2006; Lopez, 2011). We used the estimates of SL price indices for uncensored observations obtained from equation (1), regressed their logs on a set of demographic characteristics, and used the regression results to predict the prices households face with censored observations (see Table 3).

SL prices used in Phase 1 of the study with CEX original data were constructed using BLS regional CPIs. However, it is important to note that the monthly CPIs for food subgroups that the BLS provides are reported only at the national level. Monthly and quarterly regional CPIs for the Northeast, Midwest, West, and South census regions are provided only for more aggregate good categories (i.e., CPI for all expenditure items). Thus, to account for regional food price variation, we constructed regional CPIs for the subgroups by deflating the national monthly and quarterly food subgroup CPIs using the corresponding regional (Northeast, Midwest, West, and South) CPIs for all expenditure items.[6] Although this procedure assumes constant relative price differences among all food subgroups between two regions during any period, the resulting CPIs incorporate all the price information made available by the BLS to reflect both temporal and regional price variation. To produce consistent regional monthly and quarterly CPI series over time, we used the average CPI from 2002 to 2006 as the base period (i.e., average CPI 2002–2006 = 100). The monthly CPI series used in this project were not adjusted seasonally. Table 1 shows the subgroups for the construction of SL group prices for the aggregate demand models.

SL prices used in Phase 2 of the study were constructed using subgroup shares calculated using Homescan expenditures and ERS QFAHPD prices (see Table 2), which are also constructed using Nielsen Homescan data (Todd et al., 2010). We used the QFAHPD-1 version, which contains quarterly prices for 52 food groups based on both UPC-coded and random-weight food purchases for nine divisions and four regions. We used 49 of the 52 food group prices reported in QFAHPD-1 and created prices for 11 additional food groups not included in the original using the same ERS procedures (see Table 2) (Todd et al., 2010). In contrast to the CPI prices, which are available only at the regional level, we also constructed QFAHPD prices at regional and division levels.[7] We selected these prices to provide more regional variation than is available in the BLS regional CPIs. Moreover, Todd et al. also found some evidence suggesting that regional price variation is higher than price variation over time.

Fisher Ideal Price Index

We used a two-step procedure to construct the price indices for the demand model using Homescan: i) determine the price per unit for aggregate food products and ii) construct price indices for the commodity groups.

The first step involved determining a single price for a product with relatively homogeneous quality (e.g., reduced-fat yogurt), which we refer to as an aggregate food product. The aggregate products were then identified according to food commodity subgroups and then to one of the eight food commodity groups. Table 2 lists the commodity groups and subgroups together with the number of aggregate products identified within that group. Aggregate food products also were distinguished by brand type to control for quality (e.g., Dannon reduced-fat yogurt).[8] Following Diewert (1998), we used the unit value as the elementary price at an aggregate food product level. We calculated the unit value for aggregate food product s in food commodity group i for household l (UV_{si}^{l}) as

$$UV_{si}^{l} = \frac{\sum_{g=1}^{G} p_{gsi}^{l} q_{gsi}^{l}}{\sum_{g=1}^{G} q_{gsi}^{l}}, \tag{2}$$

where p_{gsi}^{l} is household l's price of the g brand in aggregate product s within commodity group i, and q_{gsi}^{l} is household l's quantity purchased of the g brand in aggregate product s within commodity group i. For some of the brand product categories in which prices, p_{gsi}^{l}, were missing, we predicted

[6] An alternative to the CPI for all expenditure items is the CPI for food at home, which is also available at the regional level. The results were robust to the regional CPI used to deflate the national food subgroups' CPIs.

[7] The nine division levels are New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific. We also checked the sensitivity of the results to the use of three alternative QFAHPD prices: division, region, and unity. Similar to results reported by Castellón, Boonsaeng, and Carpio (2015), we found the empirical differences across models to be quite small.

[8] Using a method similar to that used by Zhen et al. (2011), we identified brands that comprised a 5% or greater market share of their respective aggregate product individually. For other methodologies used in the branded food literature that use Homescan data specifically, see Zhen et al. (2011) and Arnade, Gopinath, and Pick (2008). To address concerns about degrees of freedom in the price regressions explained in the next section, where these brand-specific aggregate products contained fewer than 3,200 observations, such products were added to the "other brands" aggregate product. In the event that an entire aggregate product (all brands and non/store brands combined) contained fewer than 3,200 observations, that aggregate product was combined with another aggregate product with similar product characteristics in the same subgroup.

prices using the methods proposed by Meghir and Robin (1992) and Zhen et al. (2011).[9] To make data across product sizes comparable (e.g., ounces, pounds, etc.), we converted all product sizes to grams following the method used by the QFAHPD and calculated price per 100 g of product (Todd et al., 2010).

In the second step, we combined unit values, UV^l_{si}, into an index that represented the commodity group price. The Fisher price index for household l's commodity group i is

$$P^l_{Fi} = \sqrt{P^l_{Pi} P^l_{Li}}, \tag{3}$$

where $P^l_{Li} = \frac{\sum UV^l_{si} q_{si}}{\sum UV_{si} q_{si}}$; $P^l_{Pi} = \frac{\sum UV^l_{si} q^l_{si}}{\sum UV_{si} q^l_{si}}$; P^l_{Li} and P^l_{Pi} are household l's Laspeyres and Paasche price indices for commodity i, respectively; q_{si} is the average quantity purchased for aggregate product s in commodity i for the average household; and q^l_{si} is the quantity purchased for aggregate product s in commodity i for household l. The Laspeyres index represents the price differential household l pays for an average quantity of commodity i relative to the average household, while the Paasche index represents the price differential household l pays for its own consumption of commodity i relative to the price the average household would pay for the same quantity of commodity i. The Fisher price index formula is viewed widely as "ideal," as it is a geometric mean of the Laspeyres and Paasche indices (Diewert, 1998). The Fisher index is also the exact price index for a quadratic, homogeneous, within-group subutility function. More importantly, the Fisher index is deemed "superlative" because it is the exact price index for a quadratic, within-group subcost function, which can provide a second-order differential approximation to an arbitrary, twice continuously differential cost function (Diewert, 1976). Therefore, the Fisher price index is theoretically a more robust alternative for calculating food commodity prices, especially relative to SL prices.

Model and Estimation Procedure

We used Lewbel and Pendakur's (2009) recently proposed Exact Affine Stone Index (EASI) demand system as the parametric demand model. Relative to the popular Almost Ideal Demand Systems, EASI allows for more flexible income expansion paths (Engel curves) and for unobserved preference heterogeneity (Lewbel and Pendakur). As these authors found little empirical difference between the exact nonlinear and the approximate linear EASI estimates, we used the approximate linear version of the model. The linear EASI budget share demand model can be written as

$$\mathbf{w}_l = \mathbf{b}_0 + \sum_{r=1}^{R} \mathbf{b}_r {y_l}^r + \mathbf{C}\mathbf{z}_l + \mathbf{D}\mathbf{z}_l y_l + \mathbf{A}\mathbf{p}_l + \mathbf{B}\mathbf{p}_l y_l + \boldsymbol{\varepsilon}_l, \tag{4}$$

where index l corresponds to a household, y_l is a measure of real total expenditure ($y_l = \ln x_l - \mathbf{p}'_l \mathbf{w}_l$), x_l is the total (nominal) expenditure on all commodities, $\mathbf{w}_l$ is an n vector of commodities' budgetary share, $\mathbf{p}_l$ is an n vector of commodities' log price faced by household l, $\mathbf{z}_l$ is an m vector of sociodemographic characteristics of household l, $\mathbf{p}_l y_l$ is the interaction term between prices and real expenditure, $\mathbf{z}_l y_l$ is the interaction term between sociodemographic characteristics and real expenditure, and $\boldsymbol{\varepsilon_l}$ is an n vector of error terms. $\mathbf{C}$, $\mathbf{D}$, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{b}_0$, and $\mathbf{b}_r$ are matrices and vectors of parameters. This model is also an $R = 5$ order polynomial in y_l, which, in turn, is a nonlinear function of prices, shares, nominal expenditures, and sociodemographic characteristics (see Lewbel and Pendakur, 2009, equation (8) for details).

The system of n equations of the form in equation (1) satisfies adding-up and homogeneity restrictions if

$$\mathbf{1}'_n \mathbf{b}_0 = 1,\ \mathbf{1}'_n \mathbf{b}_r = 0\ \forall\ r \neq 0, \tag{5}$$

[9] We predicted missing prices using the results of a regression with the nonmissing brand household unit prices (\$/g) as the dependent variable, and dummy market identification (city/region) and time variables as explanatory variables.

and

(6) $$\mathbf{1}_n^{'}\mathbf{A} = \mathbf{1}_n^{'}\mathbf{B} = 0_n,\ \mathbf{1}_m^{'}\mathbf{C} = \mathbf{1}_m^{'}\mathbf{D} = 0_m,$$

where symmetry of the Slutsky matrix is ensured by the symmetry of the $n \times n$ matrices **A** and **B**.

Elasticities

Lewbel and Pendakur (2009, pp. 835–836) provided formulas for the price and income semi-elasticities of the budget shares. It can be shown that the conventional Hicksian, Marshallian, and expenditure elasticities for good n can be calculated with the following formulas:

An $n \times n$ matrix of Hicksian price elasticities ($\boldsymbol{\varepsilon}^*$):

(7) $$\boldsymbol{\varepsilon}^* = \boldsymbol{\varpi}^{-1}(\mathbf{A} + \mathbf{B}y) + \boldsymbol{\Omega}\boldsymbol{\varpi} - \mathbf{I},$$

an $n \times n$ matrix of Marshallian price elasticities ($\boldsymbol{\varepsilon}$):

(8) $$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^* - \boldsymbol{N}\boldsymbol{\Omega}\boldsymbol{\varpi},$$

an $n \times 1$ vector of expenditure elasticities ($\boldsymbol{\eta}$):

(9) $$\boldsymbol{\eta} = \boldsymbol{\varpi}^{-1}\left(\mathbf{I} + \boldsymbol{\Theta}\mathbf{p}'\right)^{-1}\boldsymbol{\Theta} + \mathbf{1}_n^{'},$$

and an $n \times m$ matrix of marginal effects of the demographic characteristics on group expenditures:

(10) $$\mathbf{MSE} = \ln x\,(\mathbf{C} + \mathbf{D}y),$$

where $\boldsymbol{\varpi}$ is a diagonal matrix with the commodities' budget shares; $\boldsymbol{\Omega}$ is an $n \times n$ matrix of ones; $\mathbf{I}$ is an identity matrix; $\boldsymbol{N}$ is a diagonal matrix with expenditure elasticities, and $\boldsymbol{\Theta}$ is the derivative of equation (1) with respect to the real expenditures y_l, such that $\boldsymbol{\Theta} = \sum_{r=1}^{R} r\mathbf{b}_r y^{r-1} + \mathbf{Dz} + \mathbf{Bp}$. All the elasticities and marginal effects were calculated at the average values of the variables and both price indices were normalized to 1 at their mean value before estimation; thus, the index l denoting households is omitted from equations (7)–(10).

A well-documented problem in demand system estimation is the endogeneity of total expenditures in equation (4) (Lewbel and Pendakur, 2009). To test and account for the presence of expenditure endogeneity, we utilized a control function approach (Blundell and Robin, 2000; Wooldridge, 2010, p. 126–129). This approach augments equation (4) with the estimated ordinary least squares (OLS) residuals $\hat{v}_l$ from a reduced form model of y_l as a function of prices ($\mathbf{p}_l$), sociodemographic characteristics ($\mathbf{z}_l$), polynomial terms of log household income up to the R (R = 5) order, and interactions between i) log household income and prices and ii) log household income and sociodemographic characteristics. Including $\hat{v}_l$ controls for endogeneity of y_l. Moreover, the significance of the coefficients related to $\hat{v}_l$ provides a test for endogeneity.

The linear approximation of the EASI demand system in equation (4) augmented with the $\hat{v}_l$ term was estimated using seemingly unrelated regression (SUR) in SAS with the MODEL procedure.[10] Homogeneity and symmetry restrictions were imposed in the demand system. Following convention, the last equation was dropped from the demand system and its parameters were recovered from the adding-up constraint. To account for heteroskedasticity in the disturbances in the system of equations having the form in equation (4), we estimated standard errors for parameters, elasticities, and marginal effect estimates using 599 iterations of the nonparametric bootstrapping procedure outlined in Wooldridge (2010, p. 438).

[10] Several approaches have been recommended to deal with censoring issues in the context of demand models (e.g., Shonkwiler and Yen, 1999); however, as the main objective of this study was to determine the effects of prices, income, and sociodemographic characteristics on average demand, we used simple linear regression models (Deaton, 1997, p. 92).

Table 4. Mean Budget Shares of CEX and Homescan Data and Average Biweekly Household Food Expenditures, 2002–2006

Commodity Groups	Mean Budget Share		Mean Expenditures		Difference Mean Expenditures (Homescan − CEX)	Percentage Difference in Mean Expenditures (Homescan − CEX)
	CEX	Homescan	CEX	Homescan		
Cereals and bakery	0.1486	0.1391	19.83	12.93	−6.89	−34.76
Meats and eggs	0.2305	0.1690	34.93	18.39	−16.54	−47.36
Dairy	0.1152	0.1086	14.47	9.79	−4.68	−32.32
Fruit and vegetables	0.1513	0.1291	21.75	12.21	−9.54	−43.85
Nonalcoholic beverages	0.1239	0.1234	15.48	11.26	−4.22	−27.24
Fats and oils	0.0310	0.0331	4.51	3.25	−1.26	−28.03
Sugar and other sweets	0.0396	0.0483	5.25	4.03	−1.21	−23.14
Miscellaneous goods	0.1600	0.2493	22.90	24.12	1.22	5.31

Comparison of Elasticities between Models

In addition to comparing the magnitudes of the elasticities obtained using different datasets and price indices, we formally tested the null hypothesis of no differences between elasticities in the two different models (models 1 and 2) using the two-sample T^2 statistic (Gupta et al., 1996):

$$T^2 = \frac{N_1 N_2}{(N_1 + N_2)} \left(\hat{\boldsymbol{E}}_1 - \hat{\boldsymbol{E}}_2\right) \hat{\boldsymbol{W}}^{-1} \left(\hat{\boldsymbol{E}}_1 - \hat{\boldsymbol{E}}_2\right), \tag{11}$$

where $\hat{\boldsymbol{W}}$ is the pooled covariance matrix obtained from the two covariance matrices, N_1 and N_2 are the number of observations, and $\hat{\boldsymbol{E}}_1$ and $\hat{\boldsymbol{E}}_2$ are the vectors of estimated elasticities of models 1 and 2, respectively. The test has an F distribution with degrees of freedom p and $N_1 + N_2 - p - 1$:

$$F = \frac{N_1 + N_2 - p - 1}{(N_1 + N_2 - 2\}p} T^2, \tag{12}$$

where p is the size of the elasticity vectors ($\hat{\boldsymbol{E}}_1$ and $\hat{\boldsymbol{E}}_2$).

Out-of-Sample Forecasting

Although Fisher price indices are thought to be superior to SL prices from a theoretical viewpoint, models also can be evaluated based on their out-of-sample predictive ability using the mean square error criteria (Fraser and Moosa, 2002). Thus, we evaluated the out-of-sample predictive ability of demand models estimated using Homescan data with the two alternative price indices. Models were estimated using data from 2002–2005 and were used subsequently to forecast budget shares in 2006.

Results and Discussion

Table 4 presents summary statistics for food groups' expenditure shares in the two datasets (over a 2-week period). Average budget shares calculated using CEX and Homescan data were similar in magnitude for cereals and bakery, dairy, fruits and vegetables, nonalcoholic beverages, fats and oils, and sugar and other sweets. The largest differences in average budget shares were observed for meats and eggs (CEX 6% higher than Homescan) and miscellaneous goods (Homescan 11% higher than CEX). All other differences between shares were less than 1.2%.

Table 4 also shows a comparison of average household expenditures calculated using the two datasets.[11] The largest difference in expenditure values also corresponded to the meats and eggs commodity group. Average 2-week household expenditures on meats and eggs calculated using Homescan data were \$18.39, while the average expenditures using the CEX data were \$34.93. Thus, the average expenditures on meat and eggs estimated using Homescan data were 47.36% lower than the CEX data counterpart. Overall, average expenditures calculated using Homescan data were lower than their corresponding expenditures obtained from the CEX data for seven of the eight commodity groups. The miscellaneous group was the only one in which average expenditures calculated using Homescan data were slightly higher than those obtained using CEX data. These results are generally consistent with previous evaluations of the differences between Homescan and CEX expenditure data (Zhen et al., 2009); however, both the periods of comparison and the food groups differ. Zhen et al. compared only 2002–2005 data and found that Homescan mean expenditures were lower than were those of CEX in 15 of the 18 categories studied, with the exception of miscellaneous foods, other dairy products, and processed fruits. Moreover, Zhen et al. did not compare average budget shares.

[11] Mean expenditures were calculated with and without the households' sample weights. As the results were very similar, we only report and discuss unweighted expenditures.

Table 5. Representativeness of Sociodemographic Variables

Variable	Percentage of CEX[a] Sample 2002–2006	Percentage of Homescan[b] Sample 2002–2006	Percentage of Current Population Survey[c] 2003–2006
No college	13.6%	3.5%	14.9%
Some college	56.5%	51.8%	57.4%
College	29.9%	44.7%	27.7%
Northeast region	18.2%	23.8%	18.7%
Midwest region	25.6%	14.4%	23.0%
South region	32.9%	39.5%	36.3%
West region	23.3%	22.3%	22.0%
< 25 years old	6.4%	0.3%	5.9%
≥25–30 years old	7.5%	2.0%	7.9%
≥30–40 years old	19.9%	14.6%	19.0%
≥40–50 years old	22.2%	24.5%	21.8%
≥50–60 years old	24.7%	37.2%	25.0%
>60 years old	19.3%	21.4%	20.5%
White	83.9%	75.3%	82.1%
Black	10.6%	14.5%	12.2%
Asian	3.9%	3.7%	3.6%
Other	1.6%	6.5%	2.2%
Family size (persons)	2.5	2.4	2.4
Hispanic	10.6%	8.1%	10.6%
Non-Hispanic	89.4%	91.9%	89.4%

Notes: Current Population Survey data are only available beginning in 2003, so 2003–2006 summary statistics are compared against the 2002–2006 CEX and Homescan samples.
Sources: [a]U.S Bureau of Labor Statistics.
[b]Private Company: Nielsen.
[c]U.S. Census Bureau (2018).

To assess the representativeness of the CEX and Homescan samples, we also compared summary statistics of the sociodemographic characteristics of the households participating in each survey relative to data from the U.S. Census Bureau (2018) (Table 5). The CEX sample survey is clearly representative of the U.S. population. On the other hand, significant differences exist between the composition of the U.S. population and the Homescan sample, especially with respect to education and age of the household head. The Homescan sample has a significantly lower proportion of non–college educated household heads and a larger proportion of college-educated heads compared to the U.S. population. Similarly, the Homescan sample has a lower proportion of young (under 30 years old), and a larger proportion of older household heads (over 50 years). Recruitment and retention of these population subsegments has been documented as an issue in the Homescan sample (Muth, Siegel, and Zhen, 2007). The Homescan sample also under-represents the Hispanic population slightly. However, it is important to mention that there are differences in the way in which the datasets define a household head, which may account for some of the differences observed. The Homescan data do not contain a household head variable; thus, for households with a married couple, the household head was determined based on household composition, working status, and education.[12] However, the CEX survey and the U.S. Census Bureau Current Population Survey define the reference person or household head as "the person (or one of the people) in whose name the housing unit is owned or rented (maintained) or, if there is no such person, any adult member,

[12] More specifically for this research, household head was determined first by household composition (female/male living alone, with related others, with nonrelated others, or married). For households with a married couple, whether one gender was employed for pay was first considered, with the working member deemed the head. If both spouses were employed for pay, the number of hours spent working outside the home determined the head of household; that is, the partner who spent more time outside the home was deemed the "head." If both partners spent an equal time outside the home, we designated the partner with a higher level of education as the "head."

Table 6. Standard Derivation for the Log of CEX–SL Price, Homescan–Fisher Price, and Homescan–SL Price

	Standard Deviation for Mean Log of Price		
Commodity Group	**CEX–SL Price**	**Homescan–Fisher Price**	**Homescan–SL Price**
Cereals and bakery	0.2681	0.1884	0.2755
Meats and eggs	0.4724	0.1965	0.4035
Dairy	0.4240	0.1709	0.3907
Fruit and vegetables	0.3333	0.1586	0.3335
Nonalcoholic beverages	0.2638	0.2235	0.2916
Fats and oils	0.3697	0.1819	0.2869
Sugar and other sweets	0.3088	0.2578	0.2321
Miscellaneous goods	0.4708	0.1818	0.4437

excluding roomers, boarders, or paid employees. If the house is owned or rented jointly by a married couple, the householder is either the husband or the wife" (U.S. Census Bureau, 2015). This variable is included in the CEX data.

Table 4 and 5 focus on the difference in the mean values of the variables used in the demand models; however, estimation of demand model parameters and elasticities also depends on the variance and covariance of the variables. For example, Table 6 displays the standard deviation of the log prices used to estimate the demand models. The variability of SL prices tended to be larger than the variability of Homescan–Fisher prices, even when the same dataset was used for the analyses. The results of Table 6 also point to the different dimensions in which the data can differ, which has implication for estimation results.

Phase 1: CEX–SL Prices Model versus Homescan–Fisher Prices Model

Table 7 shows expenditure and own-price elasticities obtained from the demand models estimated using CEX–SL prices and Homescan–Fisher price indices.[13] To compare the estimated elasticities further, we also estimated the percentage difference in the expenditure and own-price between the datasets (Table 8). All of the own-price elasticities obtained from the model estimated using CEX–SL prices were more inelastic (i.e., lower in absolute value) than those obtained using Homescan–Fisher price indices. Percentage differences in the own-price elasticities ranged from -2.94% to -55.32%, with an average absolute difference of 25.10%.

We found slightly smaller relative differences between the estimated expenditure elasticities. Although expenditure elasticities based on CEX–SL prices tended to be more elastic than those based on were Homescan–Fisher prices, this was not always the case. The observed differences can be important from an economic point of view, as in the case of the expenditure elasticities for miscellaneous goods and for meat and eggs. Percentage differences in the expenditure elasticities ranged from -32.21% to 31.29%, with a mean absolute percentage difference of 21.95%.

Comparing the results for cross-price elasticities (see Appendix A) also revealed substantial differences. While 37 out of 56 cross-price elasticities obtained using the CEX dataset indicated complementary relationships between commodity groups, the majority (31 out of 56) of those obtained using the Homescan–Fisher prices actually suggested substitute relationships. However, this is not surprising given the reported difficulty in identifying cross-price elasticities using flexible demand systems (Zhen et al., 2014; Cornelsen et al., 2016).

[13] A Wald test rejected the null hypothesis that the coefficients related to $\hat{v}_l$ in the demand system are 0 (p-value < 0.001). Thus, there is evidence that log expenditures are endogenous. All the demand estimation results in the tables correspond to models that control for endogeneity. A previous version of the manuscript presented and discussed estimation results that did not account for the endogeneity of expenditures. Although the overall direction of the differences in the elasticities across models were similar to those presented here, the differences have varying magnitudes.

Table 7. Estimated Own-Price and Expenditure Elasticities using CEX, Homescan–SL Price, and Homescan–Fisher Price Data

Phase 1: Comparing CEX–SL prices and Homescan–Fisher prices

	Hicksian Own-Price Elasticities		Marshallian Own-Price Elasticities		Expenditure Elasticities	
Commodity Groups	**CEX–SL Prices**	**Homescan–Fisher Prices**	**CEX–SL Prices**	**Homescan–Fisher Prices**	**CEX–SL Prices**	**Homescan–Fisher Prices**
Cereals and bakery	-0.5958**	-0.7540**	-0.7572**	-0.8750**	1.0866**	0.8697**
Meats and eggs	-0.3466**	-0.8902**	-0.5773**	-1.1359**	1.0007**	1.4539**
Dairy	-0.5335**	-1.0654**	-0.6226**	-1.1742**	0.7726**	1.0021**
Fruit and vegetables	-0.4396**	-1.2331**	-0.6173**	-1.3817**	1.1747**	1.1506**
Nonalcoholic beverages	-0.8020**	-0.9820**	-0.9083**	-1.0755**	0.8580**	0.7572**
Fats and oils	-1.0403**	-1.0882**	-1.0826**	-1.1266**	1.3661**	1.1606**
Sugar and other sweets	-1.7213**	-1.7541**	-1.7465**	-1.7995**	0.6370**	0.9397**
Miscellaneous goods	-0.5466**	-0.6265**	-0.7140**	-0.8252**	1.0460**	0.7967**

Phase 2: Comparing Homescan–SL prices and Homescan–Fisher prices

	Hicksian Own-Price Elasticities		Marshallian Own-Price Elasticities		Expenditure Elasticities	
Commodity Groups	**Homescan–SL Prices**	**Homescan–Fisher Prices**	**Homescan–SL Prices**	**Homescan–Fisher Prices**	**Homescan–SL Prices**	**Homescan–Fisher Prices**
Cereals and bakery	-0.5560**	-0.7540**	-0.7086**	-0.8750**	1.0970**	0.8697**
Meats and eggs	-0.5185**	-0.8902**	-0.7453**	-1.1359**	1.3422**	1.4539**
Dairy	-0.5763**	-1.0654**	-0.6734**	-1.1742**	0.8939**	1.0021**
Fruit and vegetables	-0.4109**	-1.2331**	-0.5645**	-1.3817**	1.1896**	1.1506**
Nonalcoholic beverages	-0.7781**	-0.9820**	-0.8853**	-1.0755**	0.8685**	0.7572**
Fats and oils	-1.0732**	-1.0882**	-1.1146**	-1.1266**	1.2497**	1.1606**
Sugar and other sweets	-2.2342**	-1.7541**	-2.2508**	-1.7995**	0.3439**	0.9397**
Miscellaneous goods	-0.4907**	-0.6265**	-0.6954**	-0.8252**	0.8211**	0.7967**

Notes: Single and double asterisks (*, **) indicate statistical significance at the 10% and 5% level, respectively.

We also compared the statistical significance of the price elasticities obtained using both datasets. Price elasticities obtained using the CEX–SL prices were estimated more precisely than those obtained using the Homescan dataset and Fisher prices. Using the CEX–SL dataset, only seven Marshallian price elasticities were not significant at the 10% level. When using the Homescan data, 19 price elasticities were not significant at the 10% level.

With respect to the results of the F-tests using the two-sample T^2 statistic, the null hypothesis of no difference between price elasticities and expenditure elasticities was rejected ($p < 0.01$) in the three tests conducted: one each for Marshallian, Hicksian, and expenditure elasticities. Therefore, there were statistically significant differences among the three sets of elasticities obtained from demand models estimated using CEX–SL prices and Homescan–Fisher prices.

Elasticity values are measured at a specific point on the demand curve (i.e., at a specific value of the shares and explanatory variables). The specific point used to calculate the elasticities in both datasets was the point with the average values of the variables. Because the average values of the variables differed between the datasets (see Tables 4 and 5), the estimated elasticity values could differ even if parameter estimates were the same. To explore this issue, we recalculated own-price and expenditure elasticities using the parameters of one model and the average values of the explanatory variables of the other model. Own-price elasticities calculated using the parameters of the Homescan–Fisher prices demand model end CEX average values became less elastic, and own-price elasticities estimated using the parameters of the CEX–SL prices demand model and Homescan average values became more elastic. However, the average absolute differences do not necessarily decrease. For example, when comparing Marshallian own-price elasticities estimated using CEX–SL prices and calculated at the average value of Homescan data and Marshallian own-price elasticities estimated using the Homescan–Fisher prices dataset only, average absolute differences increased slightly—from approximately 25.10% (Table 8) to approximately 30%—but the pattern of the direction of the differences disappears (i.e., own-price elasticities estimated using CEX–SL prices at the average value of the Homescan data are not consistently more inelastic than those estimated using Homescan–Fisher prices). Differences in expenditure elasticities obtained using the parameters of the estimated CEX–SL prices demand model and average value of Homescan data and Homescan–Fisher price expenditure elasticities were similar in both sign and magnitude to those reported in Table 8 (approximately 18% difference). Therefore, part of the difference between the estimated own-price elasticity values seemed to be due to the specific point on the demand curve chosen as representative (i.e., average), especially in the case of own-price elasticities.

In short, we found significant differences, both statistically and economically, between the elasticities obtained from demand models estimated using CEX–SL prices and Homescan–Fisher prices. The differences were slightly larger in the case of own-price elasticities, and the differences could be due in part to the specific point used in the calculations. As argued above, the two main potential sources of differences are i) the data and ii) the type of price indices. To further explore the effect of prices alone, in the next section, we present the results of analyses that isolated price effects by using Homescan as the only source of expenditure data.

Phase 2: Homescan–SL Prices Model versus Homescan–Fisher Prices Model

All of the own-price elasticities estimated using Homescan–SL prices but those corresponding to Sugar and Other Sweets were smaller in absolute values (i.e., less elastic) than were those obtained using Homescan–Fisher prices (Table 8). Percentage differences ranged from −59.14% to 25.08%, with an average absolute difference of 26.84%. Although the range of the differences in the own-price elasticities was larger than that observed when comparing elasticities obtained using CEX–SL prices and Homescan–Fisher prices, the overall magnitude and direction of the differences, and their average values, were similar.

The differences between the estimated expenditure elasticities ranged from approximately −63.40% to 26.14%, with an average absolute difference of 17.11%. Therefore, the differences

Table 8. Percentage Differences in CEX, Homescan–SL Price, and Homescan–Fisher Price Data-Based Elasticities (Relative to Homescan–Fisher Price-Based Values)

Phase 1: CEX–SL prices vs. Homescan–Fisher prices

Commodity Groups	Percentage Differences in Marshallian Own-Price Elasticities between CEX and Homescan–Fisher Price Data (%)	Percentage Differences in Expenditure Elasticities between CEX and Homescan–Fisher Price Data (%)
Cereals and bakery	−13.46	24.94
Meats and eggs	−49.18	−31.17
Dairy	−46.98	−22.90
Fruit and vegetables	−55.32	2.09
Nonalcoholic beverages	−15.54	13.31
Fats and oils	−3.91	17.71
Sugar and other sweets	−2.94	−32.21
Miscellaneous goods	−13.48	31.29
Average absolute difference	25.10	21.95

Phase 2: Homescan–SL prices vs. Homescan–Fisher prices

Commodity Groups	Percentage Differences in Marshallian Own-Price Elasticities between Homescan–SL Price and Homescan–Fisher Price Data (%)	Percentage Differences in Expenditure Elasticities between Homescan–SL Price and Homescan–Fisher Price Data (%)
Cereals and bakery	−19.01	26.14
Meats and eggs	−34.39	−7.68
Dairy	−42.65	−10.79
Fruit and Vegetables	−59.14	3.38
Nonalcoholic beverages	−17.68	14.70
Fats and oils	−1.06	7.68
Sugar and other sweets	25.08	−63.40
Miscellaneous goods	−15.73	3.06
Average absolute difference	26.84	17.11

between the elasticities estimated using the same Homescan data but different prices were similar to those observed when comparing the two datasets (CEX–SL prices versus Homescan–Fisher prices). Moreover, differences in expenditure elasticities can also be substantial from an economic point of view, as they are for sugar and other sweets.

The results of the analyses of the statistical significance of price elasticities obtained using Homescan–Fisher prices and Homescan–SL prices were consistent with those related to Phase 1 of the study. Price elasticities obtained using SL prices were estimated more precisely than those obtained using the Fisher prices. Using SL prices, only four Marshallian price elasticities were not significant at the 10% level, while 20 price elasticities were not significant at the 10% level when using Fisher prices. The analyses of the signs of the cross-price elasticities also indicated that using SL prices results in more cross-price elasticities (37 out of 56) relative to using the Fisher prices (25 out of 56), indicating complementary relationships between commodity groups.

Results of the F-tests using the two-sample T^2 statistic also led to rejection of the null hypothesis of no difference between price elasticities and expenditure elasticities ($p < 0.01$) in the three tests conducted: one each for Marshallian, Hicksian, and expenditure elasticities.

Although the focus of the comparison in this section was on the elasticities obtained using Homescan–SL and Homescan–Fisher prices, it is also interesting to compare own-price elasticities obtained using CEX–SL and Homescan–SL prices. Although the sources of information for both expenditures and prices differed, the differences observed in own-price elasticities in this case were

Table 9. Root-Mean-Squared Error for Budget Shared Based on In-Sample and Out-of-Sample Forecasting

	Root Mean Square Error			
	In Sample		Out of Sample	
Commodity Groups	Homescan–Fisher Price	Homescan–SL Price	Homescan–Fisher Price	Homescan–SL Price
Cereals and bakery	0.1183	0.1149	0.1129	0.1118
Meats and eggs	0.1461	0.1395	0.1464	0.1388
Dairy	0.1039	0.0968	0.1035	0.0970
Fruit and vegetables	0.1149	0.1115	0.1223	0.1188
Nonalcoholic beverages	0.1297	0.1275	0.1356	0.1328
Fats and oils	0.0521	0.0504	0.0554	0.0536
Sugar and other sweets	0.0922	0.0883	0.0969	0.0915
Miscellaneous goods	0.1725	0.1644	0.1857	0.1773

smaller (approximately 10% average absolute difference) than those between elasticities obtained using Homescan with the alternative price indices (approximately 27% average absolute difference). Own-price elasticities estimated using CEX–SL prices were on average more inelastic than those estimated using Homescan–SL prices, but no clear pattern regarding the direction of the differences emerged. More importantly, both sets of elasticities identified the same goods as elastic or inelastic. This finding provides some evidence suggesting that the price index used for demand estimation is the main source of the observed differences in the elasticities obtained when using CEX–SL prices versus Homescan–Fisher prices. This finding also suggests that Nielsen Homescan panelists are not significantly more price-sensitive relative to households selected from the U.S. population (Lusk and Brooks, 2011).

In short, the results of the analyses suggest that differences in elasticity values obtained in Phase 1, which compared demand elasticities obtained using CEX–SL prices and Homescan–Fisher prices, are due to both the data and the price indices used in the estimation and calculation of the elasticities. Overall, SL prices resulted in own-price elasticity values that were smaller in absolute value than those obtained using Fisher price indices. Moreover, the price indices appear to be the main source of the observed differences in elasticities obtained using CEX–SL prices and Homescan–Fisher prices.

Out-of-Sample Forecasting

Results of the out-of-sample performance analyses suggested that the Homescan–SL price model had slightly better out-of-sample predictive ability than the Homescan–Fisher prices model (Table 9). Root-mean-square errors for all of the demand equations estimated using Homescan–SL prices were smaller than the corresponding values obtained from demand models estimated using Homescan–Fisher prices. In addition, the in-sample predictive performance of models estimated using Homescan–SL prices was also slightly superior to that of the model estimated using Homescan–Fisher prices. Although, in theory, the Fisher index has been shown to provide a second-order approximation to arbitrary within-group subutility function, the index is just that—an approximation. It is possible that in this specific application, SL prices are actually closer to the "true" but unknown price indices. Moreover, the higher variability of SL prices relative to Fisher prices—which appears to result in more precise estimates of cross-price elasticities—might also help improve the predictive ability of models estimated using SL prices.

Summary and Conclusions

The primary goal of this study was to evaluate the differences between demand model estimates using CEX and Nielsen Homescan data. The empirical analysis was conducted using the Exact

Affine Stone Index (EASI) demand system proposed by Lewbel and Pendakur (2009). We obtained data for the study from the BLS CEX survey, Nielsen Homescan, and monthly CPIs from 2002 to 2006.

Analyses of basic summary statistics indicated that even though Homescan expenditures are generally lower than those of CEX, the budget shares are more comparable across data sources. Moreover, comparisons between the composition of the U.S. population and the CEX and Homescan sample surveys showed that the CEX sample is more representative of the U.S. population.

Demand estimation results indicated that elasticities obtained from CEX and Homescan data-based demand models differed not only statistically but also economically when different price indices were used with each dataset. All of the own-price elasticities obtained from the CEX data-based demand model using Stone–Lewbel (SL) prices were more inelastic (i.e., lower in absolute value) than those obtained from the demand model estimated using Homescan data and Fisher prices, and differences between expenditure elasticities did not follow a specific pattern. Moreover, the differences in the estimated elasticity values were substantial: 25% average absolute difference in the case of Marshallian own-price elasticities and 22% average absolute difference in the case of expenditure elasticities.

The source of the differences observed may include differences in i) data (e.g., means, variances and covariances) and ii) the type of prices used. We found evidence to suggest that the price index used is relatively more important than the data used as the differences between own-price elasticities decreased substantially when the same price index is used in both datasets. SL prices seemed to result in own-price elasticity values that were smaller in absolute value than those obtained using Fisher price indices. While it is possible that measures of SL prices include more error than do Fisher price indices, given the theoretical properties of the price indices, we also found that models estimated using SL prices showed better in- and out-of-sample performance than did those estimated using Fisher prices. Thus, although from a theoretical perspective models estimated using Fisher price indices may be superior, the models estimated using SL prices outperformed the former with respect to forecasting ability, an important dimension in evaluating models. Clearly, more work is needed to explore these findings further.

Regarding dissimilarities in the elasticities due to differences in the data sources (CEX versus Homescan data), an important aspect identified in this study is the specific point used for the calculations. This source of observed differences has important implications for applied researchers. To eliminate this potential source of bias, average values, if available, of the explanatory variables that are deemed more representative of the population of interest should be used to calculate elasticities.

Applied demand analysts need to be aware of the implications related to using not only different parametric demand models or econometric estimation procedures but also alternative data sources for both expenditure and price data. This paper contributes to the small number of studies that have evaluated the effect of data-quality characteristics on demand estimation. More work also is needed to develop procedures that will help ameliorate potential biases in demand estimation because of measurement errors likely present in the price indices used.

[Received January 2018; final revision received July 2018.]

References

Alfonzo, L., and H. H. Peterson. "Estimating Food Demand in Paraguay from Household Survey Data." *Agricultural Economics* 34(2006):243–257. doi: 10.1111/j.1574-0864.2006.00122.x.

Arnade, C., M. Gopinath, and D. Pick. "Brand Inertia in U.S. Household Cheese Consumption." *American Journal of Agricultural Economics* 90(2008):813–826. doi: 10.1111/j.1467-8276.2008.01176.x.

Banks, J., R. Blundell, and A. Lewbel. "Quadratic Engel Curves and Consumer Demand." *Review of Economics and Statistics* 79(1997):527–539. doi: 10.1162/003465397557015.

Barnett, W. A., and P. Yue. "Semi-Parametric Estimation of the Asymptotically Ideal Model: The AIM Demand System." In G. Rhodes, ed., *Advances in Econometrics,* vol. 7. Greenwich, CT: JAI Press, 1988, 299–251.

Blundell, R., and J.-M. Robin. "Latent Separability: Grouping Goods without Weak Separability." *Econometrica* 68(2000):53–84. doi: 10.1111/1468-0262.00093.

Byrne, P. J., O. Capps, and A. Saha. "Analysis of Food-Away-from-Home Expenditure Patterns for U.S. Households, 1982-89." *American Journal of Agricultural Economics* 78(1996):614–627. doi: 10.2307/1243279.

Castellón, C. E., T. Boonsaeng, and C. E. Carpio. "Demand System Estimation in the Absence of Price Data: An Application of Stone-Lewbel Price Indices." *Applied Economics* 47(2015):553–568. doi: 10.1080/00036846.2014.975332.

Cornelsen, L., M. Mazzocchi, R. Green, A. D. Dangour, and R. D. Smith. "Estimating the Relationship between Food Prices and Food Consumption—Methods Matter." *Applied Economic Perspectives and Policy* 38(2016):546–561. doi: 10.1093/aepp/ppw010.

Cox, T. L., and M. K. Wohlgenant. "Prices and Quality Effects in Cross-Sectional Demand Analysis." *American Journal of Agricultural Economics* 68(1986):908–919. doi: 10.2307/1242137.

Deaton, A. *The Analysis of Household Surveys: A Microeconometric Approach*. Washington, DC: World Bank, 1997.

Diewert, W. E. "Exact and Superlative Index Numbers." *Journal of Econometrics* 4(1976):115–145. doi: 10.1016/0304-4076(76)90009-9.

———. "Index Number Issues in the Consumer Price Index." *Journal of Economic Perspectives* 12(1998):47–58. doi: 10.1257/jep.12.1.47.

Einav, L., E. Leibtag, and A. Nevo. "Recording Discrepancies in Nielsen Homescan Data: Are They Present and Do They Matter?" *Quantitative Marketing and Economics* 8(2010):207–239. doi: 10.1007/s11129-009-9073-0.

Fraser, I., and I. A. Moosa. "Demand Estimation in the Presence of Stochastic Trend and Seasonality: The Case of Meat Demand in the United Kingdom." *American Journal of Agricultural Economics* 84(2002):83–89. doi: 10.1111/1467-8276.00244.

Gupta, S., P. Chintagunta, A. Kaul, and D. R. Wittink. "Do Household Scanner Data Provide Representative Inferences from Brand Choices: A Comparison with Store Data." *Journal of Marketing Research* 33(1996):383. doi: 10.2307/3152210.

Harris, M., and N. Blisard. "Characteristics of the Nielsen Homescan Data." Working Paper, U.S. Department of Agriculture, Economic Research Service, Washington, DC, 1995.

Hausman, J., and E. Leibtag. "Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart." *Journal of Applied Econometrics* 22(2007):1157–1177. doi: 10.1002/jae.994.

Hoderlein, S., and S. Mihaleva. "Increasing the Price Variation in a Repeated Cross Section." *Journal of Econometrics* 147(2008):316–325. doi: 10.1016/j.jeconom.2008.09.022.

Jensen, H. H., and S. T. Yen. "Food Expenditures Away From Home by Type of Meal." *Canadian Journal of Agricultural Economics* 44(1996):67–80. doi: 10.1111/j.1744-7976.1996.tb00143.x.

Kuchler, F., A. Tegene, and J. M. Harris. "Taxing Snack Foods: Manipulating Diet Quality or Financing Information Programs?" *Review of Agricultural Economics* 27(2005):4–20. doi: 10.1111/j.1467-9353.2004.00204.x.

Lewbel, A., and K. Pendakur. "Tricks with Hicks: The EASI Demand System." *American Economic Review* 99(2009):827–863. doi: 10.1257/aer.99.3.827.

Lopez, J. A. "A Comparison of Price Imputation Methods under Large Samples and Different Levels of Censoring." 2011. Paper presented at the annual meeting of the Agricultural and Applied Economics Association, Pittsburgh, Pennsylvania, July 24-26.

Lusk, J. L., and K. Brooks. "Who Participates in Household Scanning Panels?" *American Journal of Agricultural Economics* 93(2011):226–240. doi: 10.1093/ajae/aaq123.

Meghir, C., and J.-M. Robin. "Frequency of Purchase and the Estimation of Demand Systems." *Journal of Econometrics* 53(1992):53–85. doi: 10.1016/0304-4076(92)90080-B.

Muth, M. K., P. H. Siegel, and C. Zhen. "ERS Data Quality Study Design, Contract 53-3K06-06-1301." RTI Project Number 0210253.001, RTI International, Research Triangle Park, NC, 2007.

Piggott, N. E. "The Nested PIGLOG Model: An Application to U.S. Food Demand." *American Journal of Agricultural Economics* 85(2003):1–15. doi: 10.1111/1467-8276.00099.

Shonkwiler, J. S., and S. T. Yen. "Two-Step Estimation of a Censored System of Equations." *American Journal of Agricultural Economics* 81(1999):972. doi: 10.2307/1244339.

Stewart, H., and S. T. Yen. "Changing Household Characteristics and the Away-from-Home Food Market: A Censored Equation System Approach." *Food Policy* 29(2004):643–658. doi: 10.1016/j.foodpol.2004.10.003.

Todd, J. E., L. Mancino, E. Leibtag, and C. Tripodo. "Methodology behind the Quarterly Food-at-Home Price Database." Technical Bulletin 1926, U.S. Department of Agriculture, Economic Research Service, Washington, DC, 2010. Available online at https://www.ers.usda.gov/webdocs/publications/47564/8516_tb1926_1_.pdf.

Tonsor, G. T., J. R. Mintert, and T. C. Schroeder. "U.S. Meat Demand: Household Dynamics and Media Information Impacts." *Journal of Agricultural and Resource Economics* 35(2010):1–17.

U.S Bureau of Labor Statistics. "Consumer Expenditure Survey." Available online at https://www.bls.gov/cex/.

U.S. Census Bureau. "Current Population Survey (CPS): Subject Definitions." 2015. Available online at https://www.census.gov/programs-surveys/cps/technical-documentation/subject-definitions.html.

———. "Current Population Survey (CPS): Data." 2018. Available online at https://www.census.gov/programs-surveys/cps/data-detail.html.

U.S. Department of Agriculture. "Quarterly Food-at-Home Price Database." 2016. Available online at https://www.ers.usda.gov/data-products/quarterly-food-at-home-price-database/.

Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2010, 2nd ed.

Zan, H., and J. X. Fan. "Cohort Effects of Household Expenditures on Food Away from Home." *Journal of Consumer Affairs* 44(2010):213–233. doi: 10.1111/j.1745-6606.2010.01163.x.

Zhen, C., E. A. Finkelstein, J. M. Nonnemaker, S. A. Karns, and J. E. Todd. "Predicting the Effects of Sugar-Sweetened Beverage Taxes on Food and Beverage Demand in a Large Demand System." *American Journal of Agricultural Economics* 96(2014):1–25. doi: 10.1093/ajae/aat049.

Zhen, C., J. L. Taylor, M. K. Muth, and E. Leibtag. "Understanding Differences in Self-Reported Expenditures between Household Scanner Data and Diary Survey Data: A Comparison of Homescan and Consumer Expenditure Survey." *Review of Agricultural Economics* 31(2009):470–492. doi: 10.1111/j.1467-9353.2009.01449.x.

Zhen, C., M. K. Wohlgenant, S. Karns, and P. Kaufman. "Habit Formation and Demand for Sugar-Sweetened Beverages." *American Journal of Agricultural Economics* 93(2011):175–193. doi: 10.1093/ajae/aaq155.

Table A1. Estimated Uncompensated and Expenditure Elasticities Using the CEX Dataset and SL Prices

	Prices								
Quantity Demanded	Cereal and Bakery	Meats and Eggs	Dairy	Fruits and Vegetables	Nonalcoholic Beverages	Fats and Oils	Sugar and Other Sweets	Misc. Goods	Expenditure
Cereal and	−0.7572**	−0.1450**	−0.0414**	−0.1103**	−0.0437**	−0.0056	0.0375**	−0.0208**	1.0866**
bakery	(0.0144)	(0.0093)	(0.0084)	(0.0095)	(0.0099)	(0.0048)	(0.0072)	(0.0089)	(0.0175)
Meats and	−0.0807**	−0.5773**	−0.1044**	−0.0567**	−0.0354**	−0.0030	−0.0134**	−0.1298**	1.0007**
eggs	(0.0054)	(0.0101)	(0.0050)	(0.0061)	(0.0059)	(0.0027)	(0.0035)	(0.0064)	(0.0153)
Dairy	−0.0067	−0.1564**	−0.6226**	−0.0538**	0.0479**	−0.0185**	0.0487**	−0.0114	0.7726**
	(0.0100)	(0.0112)	(0.0145)	(0.0101)	(0.0103)	(0.0052)	(0.0073)	(0.0099)	(0.0201)
Fruit and	−0.1215**	−0.1265**	−0.0873**	−0.6173**	−0.1028**	−0.0081*	0.0163**	−0.1276**	1.1747**
vegetables	(0.0087)	(0.0103)	(0.0083)	(0.0140)	(0.0091)	(0.0045)	(0.0061)	(0.0094)	(0.0180)
Nonalcoholic	−0.0184	−0.0329**	0.0347**	−0.0776**	−0.9083**	0.0363**	0.0775**	0.0307**	0.8580**
beverages	(0.0126)	(0.0131)	(0.0107)	(0.0120)	(0.0224)	(0.0068)	(0.0105)	(0.0121)	(0.0208)
Fats and oils	−0.0687**	−0.1063**	−0.1371**	−0.0686**	0.0825**	−1.0826**	0.1352**	−0.1205**	1.3661**
	(0.0234)	(0.0234)	(0.0212)	(0.0235)	(0.0270)	(0.0463)	(0.0238)	(0.0184)	(0.0349)
Sugar and	0.2075**	0.0057	0.1575**	0.1436**	0.2701**	0.1284**	−1.7465**	0.1966**	0.6370**
other sweets	(0.0281)	(0.0260)	(0.0234)	(0.0255)	(0.0325)	(0.0187)	(0.0561)	(0.0251)	(0.0457)
Miscellaneous	−0.0133*	−0.1975**	−0.0397**	−0.1012**	0.0005	−0.0134**	0.0325**	−0.7140**	1.0460**
goods	(0.0075)	(0.0102)	(0.0071)	(0.0083)	(0.0082)	(0.0033)	(0.0054)	(0.0123)	(0.0178)

Notes: Standard errors are shown in parentheses. Single and double asterisks (*, **) indicate statistical significance at the 10% and 5% level, respectively.

Table A2. Estimated Uncompensated and Expenditure Elasticities Using Homescan Dataset and Fisher Prices

	Prices								
Quantity Demanded	Cereal and Bakery	Meats and Eggs	Dairy	Fruits and Vegetables	Nonalcoholic Beverages	Fats and Oils	Sugar and Other Sweets	Misc. Goods	Expenditure
Cereal and	−0.8750**	0.0313	−0.0535**	0.0818**	−0.0272*	0.0134	−0.0236*	−0.0169	0.8697**
bakery	(0.0317)	(0.0206)	(0.0204)	(0.0204)	(0.0168)	(0.0123)	(0.0130)	(0.0234)	(0.0287)
Meats and	−0.0555**	−1.1359**	−0.0202	−0.0584**	−0.0684**	−0.0263**	0.0531**	−0.1422**	1.4539**
eggs	(0.0168)	(0.0357)	(0.0162)	(0.0187)	(0.0165)	(0.0100)	(0.0114)	(0.0234)	(0.0296)
Dairy	−0.0870**	0.0448*	−1.1742**	−0.0192	0.0673**	0.0420**	0.0765**	0.0477*	1.0021**
	(0.0264)	(0.0254)	(0.0411)	(0.0272)	(0.0235)	(0.0154)	(0.0157)	(0.0266)	(0.0318)
Fruit and	0.0491**	−0.0252	−0.0323	−1.3817**	0.1568**	0.0010	0.1252**	−0.0436	1.1506**
vegetables	(0.0226)	(0.0243)	(0.0228)	(0.0436)	(0.0192)	(0.0156)	(0.0148)	(0.0283)	(0.0330)
Nonalcoholic	−0.0150	0.0241	0.0858**	0.2148**	−1.0755**	0.0115	0.0501**	−0.0531*	0.7572**
beverages	(0.0208)	(0.0243)	(0.0223)	(0.0217)	(0.0368)	(0.0113)	(0.0161)	(0.0275)	(0.0371)
Fats and oils	0.0159	−0.0846*	0.1203**	0.0026	−0.0071	−1.1266**	0.1220**	−0.2030**	1.1606**
	(0.0522)	(0.0515)	(0.0504)	(0.0608)	(0.0402)	(0.0991)	(0.0325)	(0.0510)	(0.0574)
Sugar and	−0.0777**	0.2726**	0.1787**	0.3616**	0.1055**	0.0910**	−1.7995**	−0.0718	0.9397**
other sweets	(0.0384)	(0.0432)	(0.0360)	(0.0422)	(0.0389)	(0.0229)	(0.0938)	(0.0491)	(0.0728)
Miscellaneous	0.0007	0.0146	0.0431**	0.0231	−0.0312**	−0.0149**	−0.0070	−0.8252**	0.7967**
goods	(0.0132)	(0.0163)	(0.0116)	(0.0150)	(0.0130)	(0.0068)	(0.0094)	(0.0272)	(0.0232)

Notes: Standard errors are shown in parentheses. Single and double asterisks (*, **) indicate statistical significance at the 10% and 5% level, respectively.

Table A3. Estimated Uncompensated and Expenditure Elasticities Using Homescan Dataset and SL Prices

	Prices								
Quantity Demanded	Cereal and Bakery	Meats and Eggs	Dairy	Fruits and Vegetables	Nonalcoholic Beverages	Fats and Oils	Sugar and Other Sweets	Misc. Goods	Expenditure
Cereal and	−0.7086**	−0.1148**	−0.0704**	−0.1304**	−0.0331**	−0.0319**	0.1021**	−0.1098**	1.0970**
bakery	(0.0184)	(0.0111)	(0.0092)	(0.0103)	(0.0110)	(0.0065)	(0.0111)	(0.0138)	(0.0193)
Meats and	−0.1286**	−0.7453**	−0.1005**	−0.0876**	−0.0934**	−0.0161**	0.0184**	−0.1891**	1.3422**
eggs	(0.0087)	(0.0149)	(0.0075)	(0.0083)	(0.0082)	(0.0043)	(0.0066)	(0.0120)	(0.0207)
Dairy	−0.0619**	−0.0807**	−0.6734**	−0.0438**	−0.0261**	−0.0136*	0.0953**	−0.0899**	0.8939**
	(0.0116)	(0.0112)	(0.0184)	(0.0103)	(0.0110)	(0.0068)	(0.0110)	(0.0138)	(0.0210)
Fruit and	−0.1534**	−0.0888**	−0.0689**	−0.5645**	−0.0657**	−0.0271**	0.0167*	−0.2378**	1.1896**
vegetables	(0.0110)	(0.0111)	(0.0093)	(0.0157)	(0.0108)	(0.0066)	(0.0088)	(0.0145)	(0.0219)
Nonalcoholic	−0.0055	−0.0478**	−0.0202*	−0.0273**	−0.8853**	0.0188**	0.0780**	0.0208	0.8685**
beverages	(0.0133)	(0.0126)	(0.0108)	(0.0121)	(0.0240)	(0.0071)	(0.0118)	(0.0171)	(0.0247)
Fats and oils	−0.1552**	−0.0663**	−0.0831**	−0.1133**	0.0231	−1.1146**	0.3950**	−0.1352**	1.2497**
	(0.0295)	(0.0247)	(0.0246)	(0.0273)	(0.0263)	(0.0722)	(0.0366)	(0.0333)	(0.0440)
Sugar and	0.3986**	0.2331**	0.2738**	0.1537**	0.2639**	0.3009**	−2.2508**	0.2829**	0.3439**
other sweets	(0.0341)	(0.0280)	(0.0273)	(0.0263)	(0.0300)	(0.0253)	(0.0743)	(0.0389)	(0.0594)
Miscellaneous	−0.0229**	−0.0402**	−0.0312**	−0.0756**	0.0161**	−0.0038	0.0318**	−0.6954**	0.8211**
goods	(0.0062)	(0.0070)	(0.0054)	(0.0061)	(0.0068)	(0.0034)	(0.0055)	(0.0143)	(0.0179)

Notes: Standard errors are shown in parentheses. Single and double asterisks (*, **) indicate statistical significance at the 10% and 5% level, respectively.