Journal of Agricultural and Applied Economics (2025), 1–10
doi:10.1017/aae.2025.16

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Unit Value Imputation Methods Using Household Scanner Data: A Case Study of Milk Purchases

Lingxiao Wang[1] and Oral Capps Jr[1,2]

[1]Department of Agricultural Economics, Texas A&M University, College Station, TX, USA and [2]Agribusiness, Food, and Consumer Economics Research Center, College Station, TX, USA
Corresponding author: Oral Capps Jr; Email: ocapps@tamu.edu

Abstract

We compared three common unit value imputation methods using household purchase data from 2018 to 2020 concerning five milk categories. Regression-based imputation outperformed household mean and retailer mean imputations, based on root mean squared error, mean absolute error, and mean absolute percent error. In a censored QUAIDS model, retailer mean imputation yielded statistically different estimates from the other two methods concerning compensated own-price and cross-price elasticities. We demonstrated that different price imputation methods used in household demand estimation generate different results in predicted prices and estimated price elasticities, and these differences may not necessarily be trivial.

Keywords: Household scanner data; regression imputation; QUAIDS model; unit value imputation

JEL classifications: C18; D12

1. Introduction

With the increasing use of household-level data from third-party vendors such as NielsenIQ and Circana to estimate censored response models (e.g., the Tobit model (Zheng et al., 2018) and the Heckman sample selection model (Capps et al., 2023; Cheng et al., 2021a)) as well as demand systems models for various commodities, the issue of price or unit value imputation merits attention. This issue arises from the fact that households are observed to purchase zero amounts of certain products during specific periods. Hence, the ratio of expenditures to quantities purchased, often named unit values as a proxy for retail prices, is unknown. Since previous studies suggest bias associated with missing unit values may occur, apart from the inherent endogeneity issues (Deaton, 1988, 1990, 1997), it is crucial to determine how to impute these unit values when they are missing (Dong et al., 1998; Erdem et al., 1998).

The literature has extensively explored methods for imputing missing observations (Little and Rubin, 2019; Pigott, 2001; Schafer, 1997). A commonly used approach is ad hoc forward or backward extrapolation (Enders, 2022). However, in price imputation, this method has been criticized for introducing selection bias (Erdem et al., 1998), especially when missing data are not random.

Imputation methods also have been extensively explored in survey data, primarily focusing on nonresponse (Rubin, 2004). In price imputation, the challenge is most prominent in constructing price indices (Bradley, 2003), where observed data often consists of store-level prices without links to household-level characteristics such as demographics or purchase behavior.

More recently, advanced techniques such as machine learning (Zeng and Rao, 2024), Markov Chain Monte Carlo methods (Kyureghian et al., 2011), and geospatial data integration (Hill and Scholz, 2018) have gained attention. While these methods offer potential improvements, they are often criticized for their complexity in both modeling and implementation. Given that price imputation is not the primary focus of demand analysis, the choice of method should balance ease of implementation with predictive accuracy.

The most used methods for imputing missing unit values of demand analysis in the literature include regression-based imputation, household mean imputation, and retailer mean imputation. Despite the widespread reliance on imputation techniques in general, there has been limited systematic evaluation of their predictive accuracy and implications for price imputation in demand analysis. To the best of our knowledge, no prior study has rigorously compared these methods to determine which yields the most accurate unit value imputations. By filling this gap, our findings provide new insights into the trade-offs among different imputation strategies, contributing to a more robust foundation for price imputations in empirical demand analysis.

Additionally, our study sheds light on the implications of different imputation methods within a censored QUAIDS demand system framework. This aspect also has been largely unexplored in prior research, and our findings emphasize potential differences that can arise when using various imputation methods. We believe this contribution is valuable for researchers working with scanner data, where missing price information is a persistent challenge.

In a case study, we utilize household purchases of five categories of milk products from the Nielsen Homescan Panel over 2018–2020 to compare the performance of imputed unit values obtained through these three approaches. Furthermore, we assess how the three methods affect the magnitude of compensated own-price and cross-price elasticities as well as expenditure elasticities associated with the estimation of a censored Quadratic Almost Ideal Demand System (QUAIDS) model.

The milk industry serves as a valuable case study due to its widespread consumption, nutritional importance, and evolving market dynamics. As a staple food in many households, milk plays a central role in consumer purchasing behavior. In recent years, this industry has undergone notable transformations, including the rise of plant-based milk alternatives and increased product differentiation within dairy milk categories (e.g., lactose-free, organic, and flavored milk). These developments make milk a representative product for analyzing demand interrelationships using system methods.

Our findings reveal that the differences in predicted prices and estimated price elasticities via these three price imputation methods are not trivial. The predicted values from these three methods were not highly correlated. In our case study, the regression-based method outperforms the household mean and the retailer mean imputation methods for all five milk categories. The retailer mean imputation method generated statistically different estimates of own- and cross-price price elasticities from the other two imputation methods.

2. Unit value imputation methods

Using the ratio of dollar sales to quantities purchased, we derive unit values and proxies for retail prices. The construction of unit values is consistent with the methodology proposed by Deaton (1987). Indeed, as pointed out by Deaton, bias associated with the use of unit values may occur (Deaton, 1988, 1990, 1997). The bias is attributed to quality variation and reporting errors in expenditures and/or quantities (measurement errors). Deaton (1988) suggested that the bias associated with quality variation makes the demand for a commodity appear to be more elastic, overstating the response of quantity to changes in price.

Gibson and Rozelle (2011) suggested that two types of measurement error bias are evident: (1) attenuation bias because unit values are noisy measures of market prices; and (2) bias due to correlated errors in measuring expenditures and/or quantities. In the case of attenuation bias, they noted that the bias was in the opposite direction to that attributed to quality variation. If so, then the bias due to quality variation and the bias due to attenuation are offsetting to some degree. However, Gibson and Rozelle (2011) also pointed out that the bias due to correlated errors operated in the opposite direction to attenuation bias. Consequently, the bias due to correlated errors reinforces the bias due to quality effects. Importantly, Gibson and Rozelle (2011) documented that the bias associated with quality variation was relatively minor, also consistent with the finding of Deaton (1997).

2.1. Regression-based imputation

The regression-based imputation method utilizes demographic information from purchasing households to infer unit values for non-purchasing households. This method has been widely used for unit value imputation in the economic literature (Alviola and Capps, 2010; Bakhtavoryan et al., 2022; Capps et al., 2023, Cheng et al., 2021a, 2021b; Dharmasena and Capps, 2012, 2014; Kyureghian et al., 2011; Lopez et al., 2012). In Alviola and Capps (2010), Dharmasena and Capps (2012, 2014), Cheng et al. (2021a, 2021b), and Capps et al. (2023). Missing imputed values for households who did not purchase the products in question were generated via auxiliary regressions in which observed unit values for each of the respective products were regressed as a function of demographic factors, typically household income, household size, and region as well as dummy variables pertaining to time period. These instrument variables have been used in these prior studies to not only obtain values of missing prices but also to mitigate price endogeneity issues. Notably, the predicted unit values using a regression-based method are specific to the household, particularly household income, household size, geographic region, and to a particular period.

2.2. Household mean imputation

Household mean imputation, also known as group mean imputation and cell mean imputation (Lopez, 2014), replaces missing unit values of non-purchasing households with mean unit values based on purchasing households according to various criteria. For example, Ackerberg (2001) used observed unit values obtained in the same week and in the same store from purchasing households to replace missing unit values for non-purchasing households. Additionally, Dong et al. (2004) and Golan et al. (2001) replaced missing prices for non-purchasing households with the mean price of purchasing households located in the same state and in the same area of urbanization. This imputation method assumes that both non-purchasing and purchasing households face the same average price level for a specific product in a particular geographic location and during a particular time. Household income and household size do not play any role in predicting unit values based on household mean imputation.

2.3. Retailer mean imputation

Unlike the regression-based and household mean imputation methods, which use data from household purchasing records (e.g., the Nielsen Homescan Panel), the retailer mean imputation method utilizes actual retail price information based on purchases that occur at stores located in various geographic markets affiliated with third-party vendors like NielsenIQ and Circana. The respective vendors themselves impute prices using the average price of the Universal Product Code (UPC) during a particular time by retail outlet. Hence, the retailer mean imputation method relies on average prices common to the same geographic area(s) to represent the unobserved prices of products related to non-purchasing households (Zhen et al., 2014). Importantly, these price imputations do not vary across households within the same period. The variability of unit values based on the household imputation method and the retailer imputation method typically is much less than the variability of unit values based on the regression-based imputation method. Additionally, like the household mean imputation method, household income and household size do not play any role in predicting unit values based on retailer mean imputation.

3. Data

We utilize household purchase data concerning various milk products from the Nielsen Homescan Panel for price imputation using regression-based and household mean methods. These datasets are aggregated by quarter and by year.[1] Additionally, we categorize these products into five categories: traditional white milk, traditional flavored milk, lactose-free milk, organic milk, and the aggregate of plant-based milk alternatives (PBMA).[2] Our dataset contains quarterly milk purchase data of 43,310 households from 2018 to 2020.

For the regression-based method, we used an out-of-sample validation approach. Specifically, we regressed observed unit values for each of the five products for calendar years 2018 and 2019 (serving as the training period), where observed unit values for each of the five product categories were regressed on household income, household size, DMA fixed effects, and quarter and year indicators.[3] For all five categories considered, heteroscedasticity was detected using the Breusch-Pagan test in each of the regression-based imputation equations. We address heteroscedasticity by calculating robust standard errors (White, 1980). We then applied the estimated models to predict unit values for calendar year 2020 (the testing period) and evaluated the prediction accuracy by comparing imputed values against the observed 2020 values. For the household mean method,[4] we took the average of the observed unit values by DMA and quarter to obtain the predicted values for each of the five products for the calendar year 2020. For the retailer mean method, we matched households based on *retail* prices reported by Nielsen from retail outlets in the same DMA and obtained the average of observed DMA unit values per quarter for the calendar year 2020.[5]

Table 1 shows summary statistics of the observed values and the missing rates of unit values for each product category over the period 2018–2020. The missing rate for the price of a specific product is calculated as the number of observations with zero purchases divided by the total number of observations. Given the rather sizeable missing rates associated with the milk-related products, the issue of unit value imputations warrants attention.

4. Empirical results

Mean predicted unit values vary across imputation methods, as shown in Table 2. In Table 3, we examine the correlations among predicted unit values from the three imputation methods to assess their consistency. High correlations indicate similar imputed prices across methods, suggesting minimal impact on demand estimates. Lower correlations, however, highlight discrepancies that may influence price elasticity estimates. The respective predicted unit values associated with these three methods were not highly correlated. These results imply that the use of these imputations may yield different magnitudes of own-price elasticities, cross-price elasticities, and total expenditure elasticities.

[1]Monthly or weekly data will increase the missing rate in prices.

[2]Our five-category classification – traditional white milk, traditional flavored milk, organic milk, lactose-free milk, and plant-based alternatives – captures key consumer behavior patterns beyond fat content, reflecting market segmentation and health considerations.

[3]One reviewer raises the question of controlling for household fixed effects. Household income and household size are common socio-demographic factors in the regression imputation method.

[4]Suggested by a reviewer, we also have tried the weighted household mean price, taking expenditure share as the weight. The correlation between the weighted household mean price and the household mean price (equally weighted) was relatively high at 0.848. Further, the QUAIDS demand estimation results were consistent with or without the use of weighted household means, perhaps attributed to the vast sample size.

[5]The retailer mean is constructed from retailer scanner data, where stores report weekly average prices at the UPC level. To impute household missing prices, we first identify the Designated Market Area (DMA) of each store. Next, we compute the average product prices for each DMA at the quarterly level. Finally, households are matched to their respective DMAs, and missing prices are imputed using these quarterly DMA-level averages.

Table 1. Average unit values and missing rates for each milk category, 2018–2020

Product category	Average unit values ($/oz)	Missing rate 2018	Missing rate 2019	Missing rate 2020
Traditional White	0.027	14.2%	15.5%	15.6%
	(0.022)			
Traditional Flavored	0.047	76.2%	77.4%	77.6%
	(0.031)			
Lactose-Free	0.059	90.6%	90.5%	85.4%
	(0.018)			
Organic	0.059	91.0%	90.8%	90.0%
	(0.019)			
PBMA	0.051	67.6%	66.3%	65.1%
	(0.109)			

Note: Standard errors are in parentheses.

Table 2. Means of observed and predicted unit values for calendar year 2020

Product category	Observed value ($/oz)	Regression imputation ($/oz)	Household mean imputation ($/oz)	Retailer mean imputation ($/oz)
Traditional White	0.028	0.026	0.023	0.036
Traditional Flavored	0.049	0.041	0.037	0.067
Lactose-Free	0.062	0.055	0.054	0.067
Organic	0.059	0.056	0.056	0.071
PBMA	0.057	0.042	0.043	0.056

To measure the precision of the predicted unit values against the observed unit values, we used three conventional metrics associated with forecasting: root mean square error (RMSE), mean absolute error (MAE), and mean absolute percent error (MAPE). These metrics, presented in Table 4, revealed that unit values predicted via the regression-based method had the smallest RMSE, MAE, and MAPE for all five product categories. Hence, among the three methods considered, the regression-based method outperformed the household mean and the retailer mean methods regarding prediction accuracy. Notably, most MAPE values exceeded 25%, indicating disparities between predicted and observed unit values, especially for traditional flavored milk.

Finally, we compared the compensated own-price and cross-price elasticities derived from the estimation of a household-level censored QUAIDS model (Banks et al., 1997), based on imputed values using the three methods. Specifically, we adopted and re-estimated the QUAIDS model of Capps and Wang (2024) using the imputed values associated with each of the three methods in analyzing interrelationships among dairy milk and plant-based milk alternatives for U.S. households from 2018 to 2020.

In Figure 1, we show the estimates of compensated own-price and cross-price elasticities with 95% confidence intervals based on the three imputations associated with missing unit values. In Figure 2, we compare the estimates of expenditure elasticities with 95% confidence intervals based on these unit value imputations. In most cases, the compensated price elasticities estimated via the

Table 3. Correlations among predicted unit values based on the three imputation methods

Product category	Imputation method	Regression imputation	Household mean imputation	Retailer mean imputation
Traditional White Milk	Regression	1.000	0.482	0.463
	Household Mean		1.000	0.223
	Retailer Mean			1.000
Traditional Flavored Milk	Regression	1.000	0.331	0.387
	Household Mean		1.000	0.131
	Retailer Mean			1.000
Lactose-Free Milk	Regression	1.000	0.204	0.074
	Household Mean		1.000	0.009
	Retailer Mean			1.000
Organic Milk	Regression	1.000	0.248	0.210
	Household Mean		1.000	0.100
	Retailer Mean			1.000
PBMA	Regression	1.000	0.127	0.550
	Household Mean		1.000	0.064
	Retailer Mean			1.000

Table 4. Evaluations of predictions based on the three imputation methods with observed values for calendar year 2020

Product category	Imputation method	RMSE	MAE	MAPE (%)
Traditional White Milk	Regression	0.030	0.008	27.8
	Household Mean	0.031	0.009	31.0
	Retailer Mean	0.032	0.013	61.0
Traditional Flavored Milk	Regression	0.031	0.020	44.5
	Household Mean	0.033	0.022	51.1
	Retailer Mean	0.036	0.029	101.2
Lactose-Free Milk	Regression	0.020	0.012	19.2
	Household Mean	0.020	0.013	19.5
	Retailer Mean	0.020	0.014	25.1
Organic Milk	Regression	0.020	0.012	25.1
	Household Mean	0.021	0.013	28.1
	Retailer Mean	0.024	0.018	43.7
PBMA	Regression	0.158	0.021	26.9
	Household Mean	0.158	0.021	27.9
	Retailer Mean	0.158	0.027	45.8

https://doi.org/10.1017/aae.2025.16 Published online by Cambridge University Press

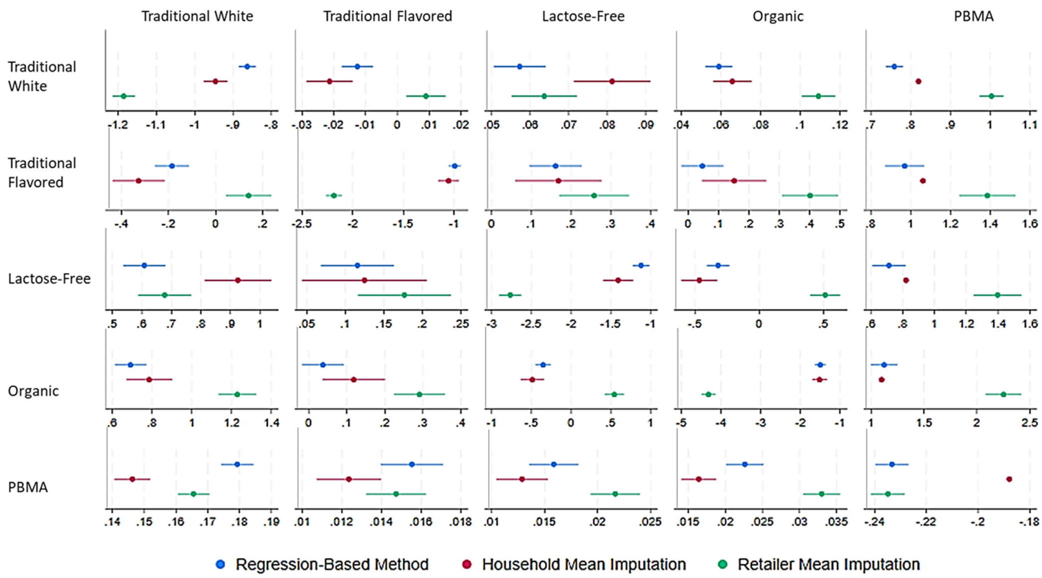

Figure 1. Compensated own-price and cross-price elasticity estimates and 95% confidence intervals using three unit value imputation methods.

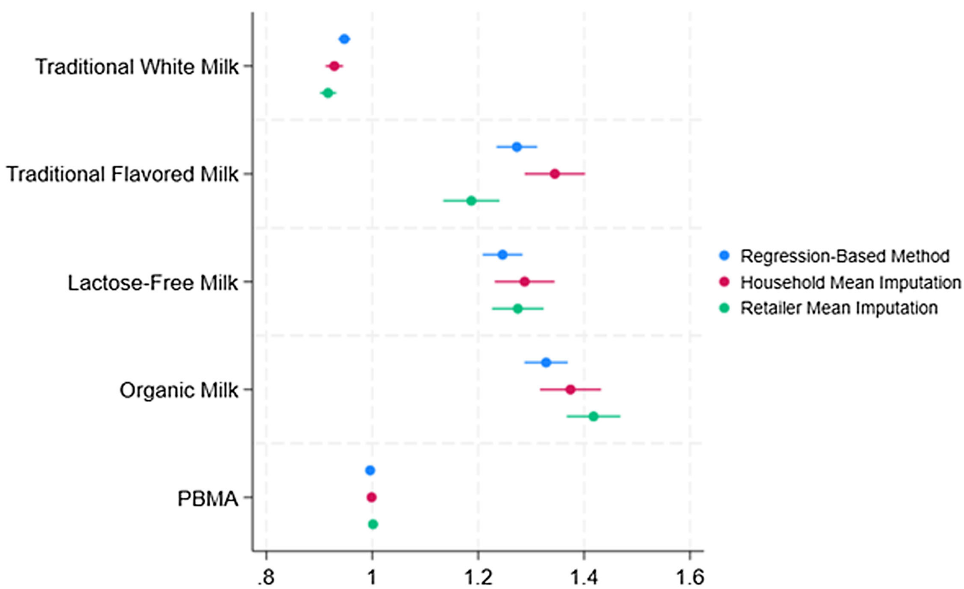

Figure 2. Total expenditure elasticity estimates and 95% confidence intervals using three unit value imputation methods.

regression-based and the household mean methods for missing unit values were relatively consistent with each other. But these compensated price elasticities were statistically different from those obtained using the retailer mean method. For example, from Figure 1, the compensated own-price elasticity for traditional white milk, calculated using the regression-based and household mean method for missing unit values, was less than 1 in absolute value, indicative of inelastic demand. In contrast, the compensated own-price elasticity for traditional white milk based on missing unit values imputed using the retailer mean method was calculated to be greater than 1 in absolute value, indicative of elastic demand.

https://doi.org/10.1017/aae.2025.16 Published online by Cambridge University Press

However, regarding total expenditure elasticities, as presented in Figure 2, the estimates from all three methods displayed relative consistency. That said, realize in demand system analysis that due to the homogeneity condition, the sum of the unconditional own-price and cross-price elasticities along with the total expenditure elasticity for each category must sum to zero. Hence, if differences across imputation methods give rise to differences in own-price and cross-price elasticities, then these differences may translate into differences in total expenditure elasticities.

5. Concluding remarks

Regression-based, household mean, and retailer mean imputation methods are commonly used to address missing unit values in estimating censored response and demand systems models. This study compared these imputation methods using data from household purchases of five milk products from 2018 to 2020, finding that predicted unit values for 2020 were not highly correlated across methods. In our case study, the regression-based method was preferred based on RMSE, MAE, and MAPE metrics. The study also assessed the impact of these imputation methods on the magnitude and significance of compensated own-price, cross-price, and expenditure elasticities from a censored QUAIDS model. While expenditure elasticities were unaffected by the imputation method, the type of imputation significantly influenced compensated price elasticities, with those from the retailer mean method differing statistically from the others. All these results suggest that the choice of price imputation method plays a non-trivial role in estimating price elasticities using household-level scanner data.

The observed differences in imputation outcomes can be attributed to how each method handles missing price data, particularly in relation to the extent and pattern of missingness. Household mean imputation assumes stable purchasing patterns within households, making it appropriate when missing prices occur among regular buyers. In contrast, regression-based imputation leverages observable household and market characteristics, which may be more effective when price variation is driven by demographics or regional differences. Retailer mean imputation, on the other hand, assumes uniform store-level pricing; however, if prices vary significantly across retailers, this method may introduce bias. Given these distinctions, selecting an imputation method that aligns with the data structure is critical, as it can influence demand estimation results.

In this study, we employ a linear model to impute missing prices using the regression-based approach, consistent with standard approaches in the literature. While this method provides a straightforward and interpretable framework, we acknowledge that alternative regression specifications, including non-linear models or additional predictor variables, could enhance imputation accuracy. In addition, as is common in studies using scanner data, if a household does not record a purchase of a particular item in a given period, it is not possible to determine whether the household chose not to buy the item (true zero demand), did not encounter the product, or failed to scan the item due to recording error (Einav et al., 2010).

Additionally, while our analysis focuses on a specific set of products and time periods, replicating this approach across different product categories and extended time frames would further assess the robustness of our findings. The primary objective of this paper is to provide a practical reference for commonly used price imputation methods in demand estimation. Future research could explore more complex models, including machine-learning techniques, to refine prediction accuracy. Going forward, we recommend replicating this analysis across different products and time periods to further validate and refine our conclusions.

Acknowledgements. Researchers' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researchers and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

Author contribution. Conceptualization, OCJ; Methodology, LW, OCJ; Formal Analysis, LW; Data Curation, LW; Writing – Original Draft, LW, Writing – Review and Editing, OCJ, LW; Supervision, OCJ; Funding Acquisition, NA.

Financial support. This research received no specific grant from any funding agency, commercial or non-profit sectors.

Data availability statement. Researcher's own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC, and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at the University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. Because of contractual stipulations, we are not at liberty to share the data publicly.

Competing interests. Authors declare no conflict of interests.

References

Ackerberg, D.A. "Empirically distinguishing informative and prestige effects of advertising." *RAND Journal of Economics* **32**,2(2001):316–33.

Alviola, P.A., and O. Capps. "Household demand analysis of organic and conventional fluid milk in the United States based on the 2004 Nielsen Homescan panel." *Agribusiness* **26**,3(2010):369–88.

Bakhtavoryan, R., O. Capps, and S. Dharmasena. "A household-level demand system analysis of nuts in the United States." *Agricultural and Resource Economics Review* **51**,2(2022):283–310.

Banks, J., R. Blundell, and A. Lewbel. "Quadratic Engel curves and consumer demand." *Review of Economics and statistics* **79**,4(1997):527–39.

Bradley, R. "Price index estimation using price imputation for unsold items." In *Scanner Data and Price Indexes*. University of Chicago Press, 2003, pp. 349–82.

Capps, O., M. Cheng, J. Kee, and S.L. Priestley. "A cross-sectional analysis of the demand for coffee in the United States." *Agribusiness* **39**,2(2023):494–514.

Capps, O., and L. Wang. "US household demand system analysis for dairy milk products and plant-based milk alternatives." *Journal of the Agricultural and Applied Economics Association* **3**,4(2024):655–72.

Cheng, G., O. Capps, and S. Dharmasena. "Demand analysis of peanuts and tree nuts in the United States: A micro-perspective." *International Food and Agribusiness Management Review* **24**,3(2021a):523–44.

Cheng, G., O. Capps, and S. Dharmasena. "Demand interrelationships of peanuts and tree nuts in the United States." *Journal of Agribusiness* **39**,345-2022-278(2021b):15–38.

Deaton, A. "Estimation of own-and cross-price elasticities from household survey data." *Journal of econometrics* **36**, 1-2(1987):7–30.

Deaton, A. "Quality, quantity, and spatial variation of price." *The American Economic Review* **78**,3(1988):418–30.

Deaton, A. "Price elasticities from survey data: extensions and Indonesian results." *Journal of econometrics* **44**,3(1990):281–309.

Deaton, A. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Washingon, D.C: The World Bank, World Bank Publications, 1997. http://documents.worldbank.org/curated/en/593871468777303124.

Dharmasena, S., and O. Capps. "Intended and unintended consequences of a proposed national tax on sugar-sweetened beverages to combat the US obesity problem." *Health economics* **21**,6(2012):669–94.

Dharmasena, S., and O. Capps. "Unraveling demand for dairy-alternative beverages in the United States: The case of soymilk." *Agricultural and Resource Economics Review* **43**,1(2014):140–57.

Dong, D., B.W. Gould, and H.M. Kaiser. "Food demand in Mexico: An application of the Amemiya-Tobin approach to the estimation of a censored food system." *American Journal of Agricultural Economics* **86**,4(2004):1094–107.

Dong, D., J. S. Shonkwiler, and O Capps. "Estimation of demand functions using cross-sectional household data: The problem revisited." *American Journal of Agricultural Economics* **80**,3(1998):466–73.

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**,2(2010):207–39.

Enders, C.K. *Applied Missing Data Analysis*. Second Edition. Guilford Publications, 2022.

Erdem, T., M.P. Keane, and B. Sun. "Missing price and coupon availability data in scanner panels: Correcting for the self-selection bias in choice model parameters." *Journal of econometrics* **89**,1-2(1998):177–96.

Gibson, J., and S. Rozelle. "The effects of price on household demand for food and calories in poor countries: Are our databases giving reliable estimates?" *Applied Economics* **43**,27(2011):4021–31.

Golan, A., J.M. Perloff, and E.Z. Shen. "Estimating a demand system with nonnegativity constraints: Mexican meat demand." *Review of Economics and statistics* **83**,3(2001):541–50.

Hill, R.J., and M. Scholz. "Can geospatial data improve house price indexes? A hedonic imputation approach with splines." *Review of Income and Wealth* **64**,4(2018):737–56.

Kyureghian, G., O. Capps, and R.M. Nayga. "A missing variable imputation methodology with an empirical application." In *Missing Data Methods: Cross-Sectional Methods and Applications*. Emerald Group Publishing Limited, 2011, pp. 313–37.

Little, R.J., and D.B. Rubin. *Statistical Analysis with Missing Data*. Third Edition. John Wiley & Sons, 2019.

Lopez, J.A. "Imputation methods and approaches: an analysis of protein sources in the Mexican diet." *International Journal of Food and Agricultural Economics (IJFAEC)*, **2**, 1128-2016-92035(2014):29–48.

Lopez, J.A., J.E. Malaga, B. Chidmi, E.J. Belasco, and J. Surles. "Mexican meat demand at the table cut level: Estimating a censored demand system in a complex survey." *Journal of Food Distribution Research* **43**,2(2012):64–90.

Pigott, T.D. "A review of methods for missing data." *Educational Research and Evaluation* **7**,4(2001):353–83.

Rubin, D.B. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, June 2004.

Schafer, J.L. *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC, July 1997. https://doi.org/10.1201/9780367803025.

White, H. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica: Journal of the Econometric Society* **48**,4(1980):817–38.

Zeng, S., and D. Rao. "Random forests with economic roots: explaining machine learning in Hedonic imputation." *Computational Economics*, (2024): 1–25.

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**,1(2014):1–25.

Zheng, W., S. Dharmasena, O. Capps, and R. Janakiraman. "Consumer demand for and effects of tax on sparkling and non-sparkling bottled water in the United States." *Journal of Agribusiness in Developing and Emerging Economies* **8**,3(2018):501–17.

Cite this article: Wang, L. and O. Capps Jr (2025). "Unit Value Imputation Methods Using Household Scanner Data: A Case Study of Milk Purchases." *Journal of Agricultural and Applied Economics*. https://doi.org/10.1017/aae.2025.16