Issue # 18: Numerator Quantity Imputation Method Validation

Ron Huang

January 23, 2026

Contents

1 Methodology

In this soda quantity imputation method, we manually create crosswalk between Nielsen and Numerator data at the brand-channel level.[1] The procedure is as follows

1. Normalize brand names and channel types in both Numerator and Nielsen data; find exact matched brand and channel or create new ones that unify the brand/channel (for example, 'diet coke' and 'coke zero' unified to 'coke'; gas & convenience and bodega unified to convenience). The number of merged brands and channels for 2022-2023 panel is as follows:
    - Number of matched and unified channels: 29
    - Number of Numerator channels: 107. Number of Nielsen channels: 68.
    - Number of matched and unified brands: 328
    - Number of matched numerator brands: 536 out of 3216.
    - Number of matched Nielsen brands: 350 out of 537.[2]
2. In Nielsen data, calculate the average price per fluid ounces at the brand-channel-year-month by $\frac{\text{item unit price}}{\text{package weight}}$, where the item unit price is adjusted for inflation and seasonality using CPI of 'Nonalcoholic beverages and beverage materials in U.S. city average, all urban consumers, seasonally adjusted' (CPI).
3. Merged Nielsen price per fluid ounces column to Numerator data by the brand-channel-year-month crosswalk.

2 Imputation Quality

We can validate the accuracy of this imputation method by repeating the same procedure above within the Nielsen data. We compute the average price per fluid ounces for each category of the matched brand-channel-year-month crosswalk between Nielsen and Numerator and impute the total fluid ounces per purchase by $\frac{\text{item unit price}}{\text{average price per fl oz}} \times \text{quantity}$.[3] We compare the actual observed total fluid ounces per purchase to $\frac{\text{item unit price}}{\text{average price per fl oz}} \times$ quantity at the household-month level to evaluate the imputation accuracy. At the household-month level, the average absolute error is about 35% and median absolute error is about 28%. 84% of household-month errors are below 50%.

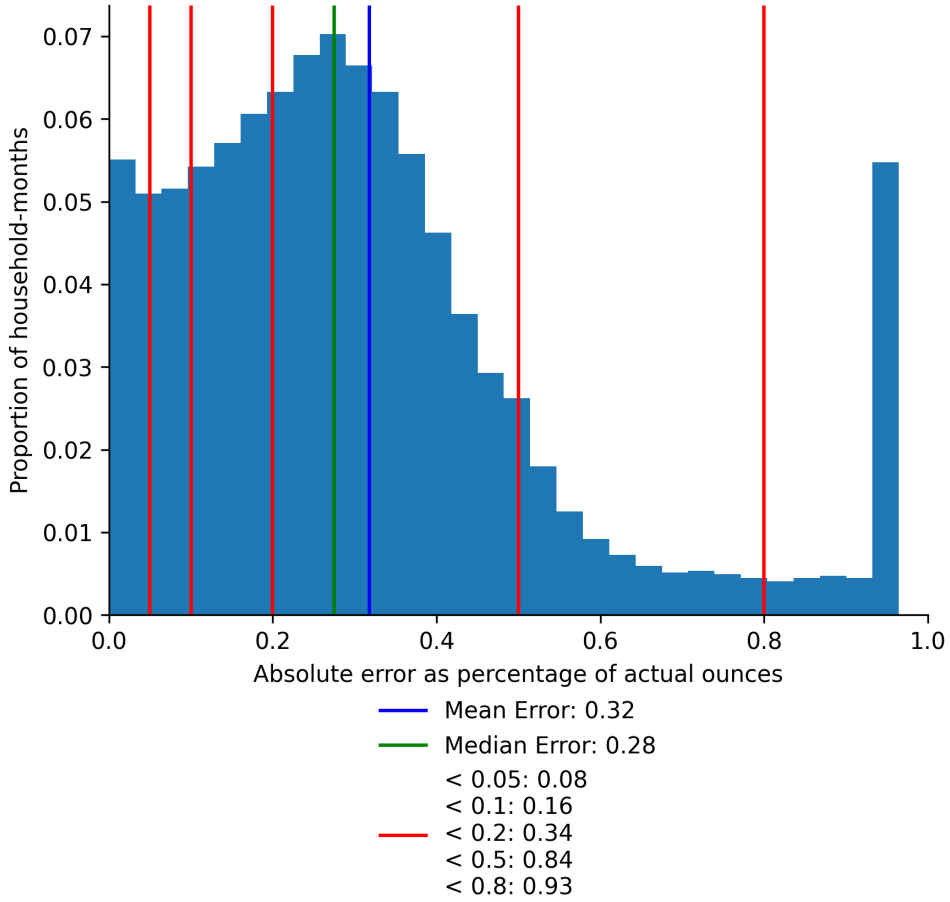
[1] We do not observe retailer's name in Nielsen data. We only see its internal retailer identifier.

[2] Before 2021, the brand names in Nielsen data may not be well cleaned and normalized so that's why we observed about 7000 unique brands in the deliverable for previous imputation method. Here the unique brands in Numerator being 3216 may also suggest that it is not well cleaned and normalized, so we did a simple normalization before

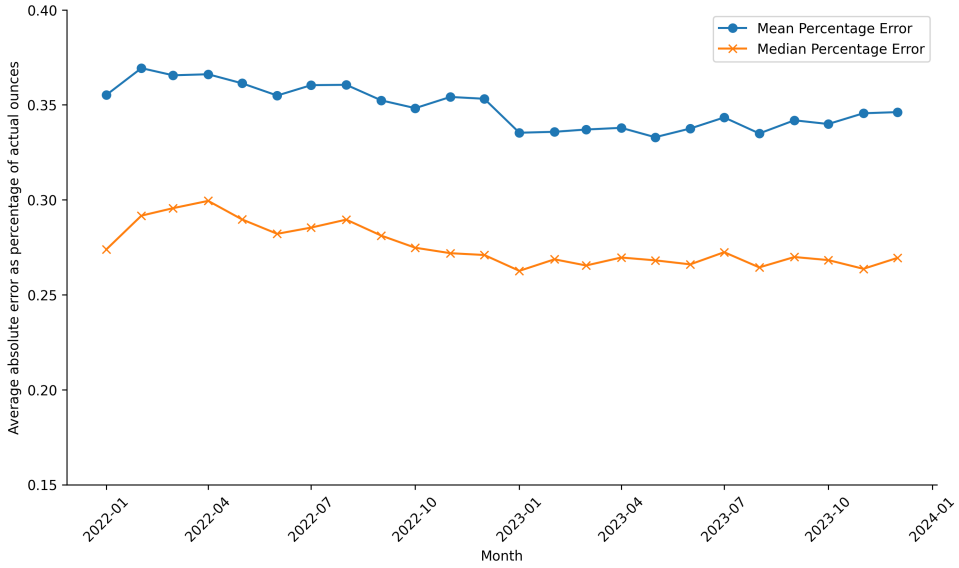
[3] For the unmatched, we do backward elimination to match on brand-year-month, and to match just on brand if still unmatched.

Figure 1

(a) Absolute error as percentage of observed total fluid ounce per household-month



(b) Monthly average of absolute error as percentage of observed total fluid ounce per household-month



Notes: There are some outliers with large absolute errors. In panel (a) we winsorize at the 95th percentile for visualization purpose. Errors in panel (b) are not winsorized.

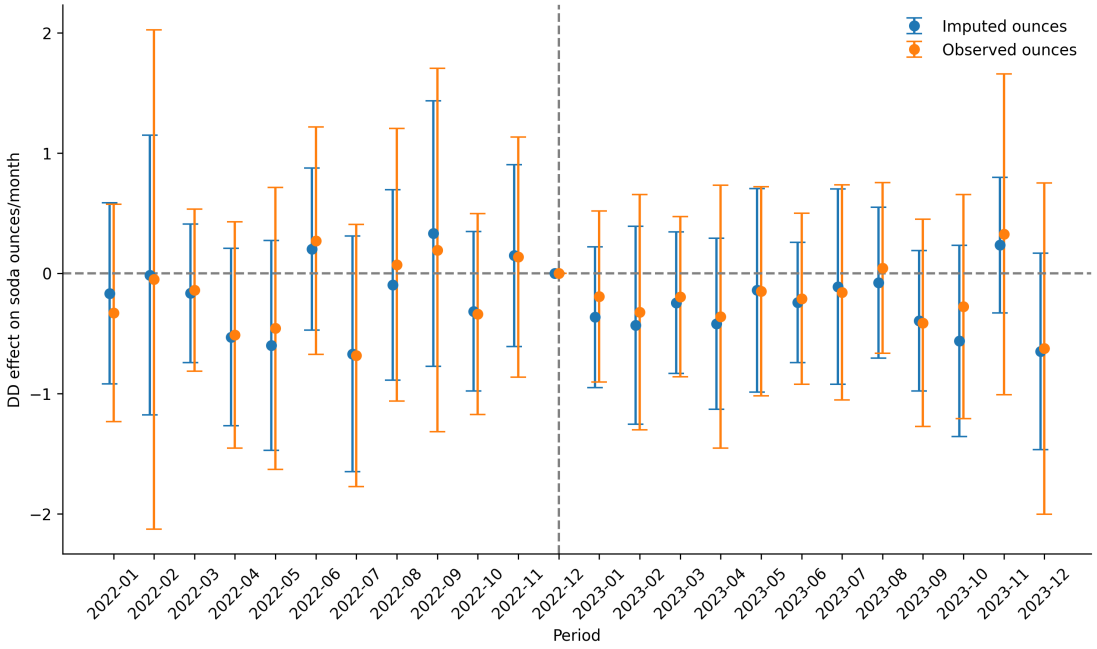


Figure 2: Event Study - Observed and Imputed Fluid Ounces

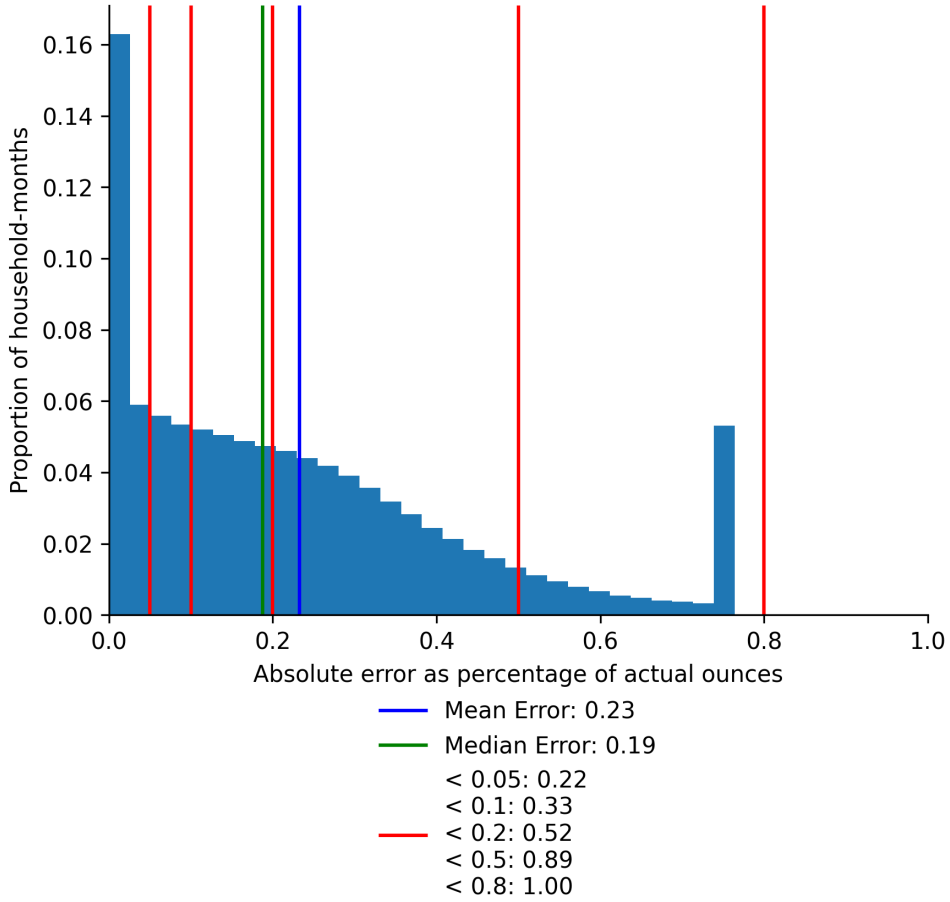
Figure 8 presents the event study using imputed and actual observed fluid ounces as left hand side outcome in diff-in-diff regression. In general, imputed quantity does not shift the estimates significantly regardless of non-negligible imputation errors. The seemingly more precise estimates stem from the fact that the quantity variation within each brand-channel-year-month cell would be reduced as we populate the average of price per fluid ounces to impute quantity.

2.1 Add State and Income

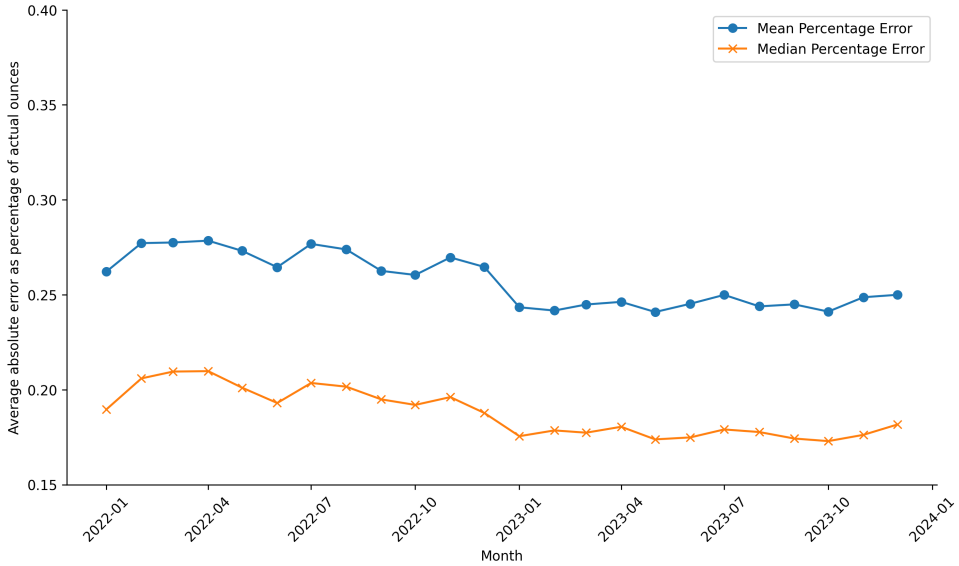
Here we try to add states and income when creating the crosswalk between Numerator and Nielsen in order to try to reduce the imputation error, e.g. we compute average price per fluid ounces at the brand-channel-state-income-year-month level, and merge Nielsen to Numerator at this level. Note that for unmatched items, we use brand-year-month average to impute; if still unmatched, we use brand average to impute. When actually merging with Numerator data, for soda away from home, we first impute using soda at home unit price; we then multiple the unit price of soda away from home by 2, according to the ratio of unit price for soda away from home to unit price for soda at home in FoodAPS data.

Figure 3

(a) Absolute error as percentage of observed total fluid ounce per household-month



(b) Monthly average of absolute error as percentage of observed total fluid ounce per household-month



Notes: There are some outliers with large absolute errors. In panel (a) we winsorize at the 95th percentile for visualization purpose. Errors in panel (b) are not winsorized.

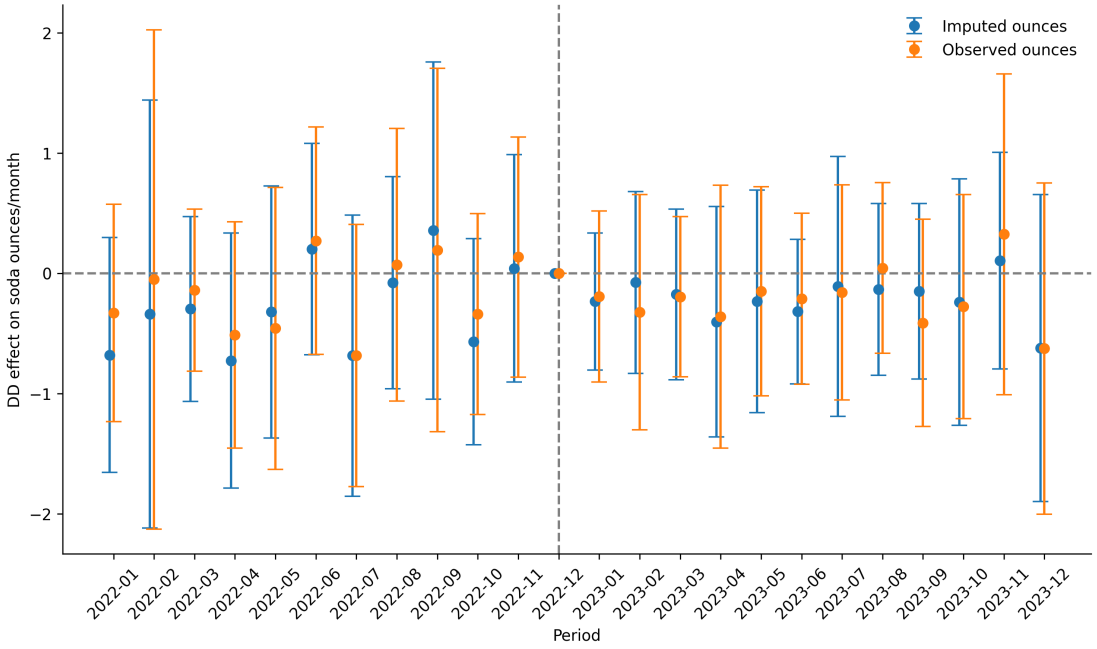


Figure 4: Event Study - Observed and Imputed Fluid Ounces

3 Cross-Validation

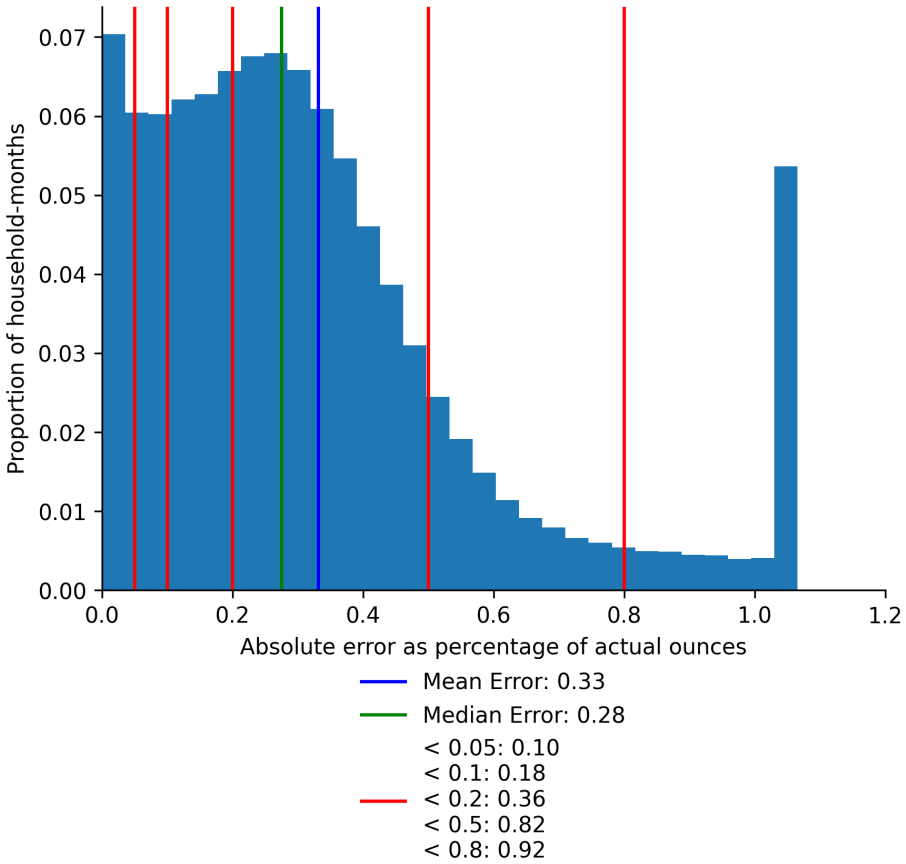
Note that the previous section does not separate a test and training set when calculating the error rate of the prediction problem. Here, we do validation set approach (one training and one test dataset) and 5-fold cross-validation. The initial crosswalk for matching is now brand-channel-year-month-region-income.[4] We do the same backward elimination for the unmatched, to match on brand-channel-year-month-region, brand-channel-year-month, brand-year-month, and brand if still on matched.

[4]Region is the 4 census region. Income is the indicator for below median income.

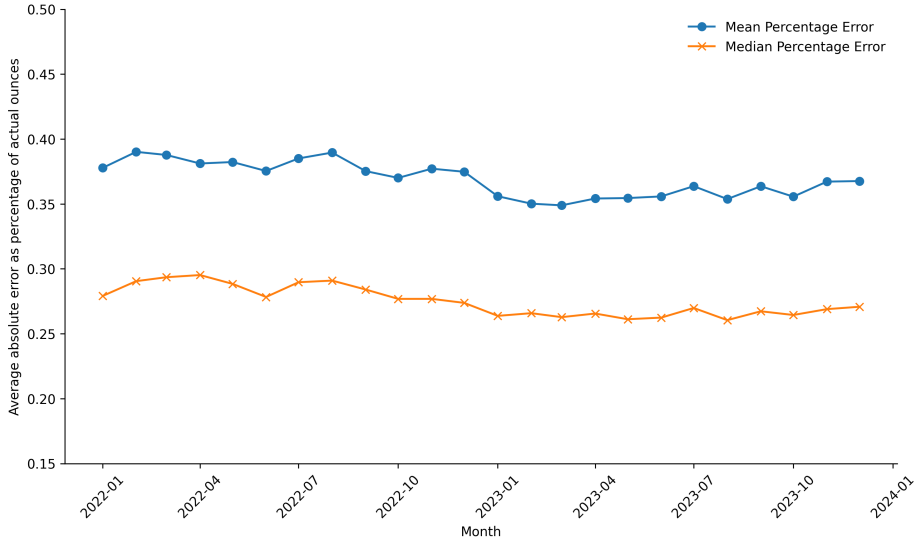
3.1 Validation Set Approach

Figure 5

(a) Absolute error as percentage of observed total fluid ounce per household-month



(b) Monthly average of absolute error as percentage of observed total fluid ounce per household-month



Notes: There are some outliers with large absolute errors. In panel (a) we winsorize at the 95th percentile for visualization purpose. Errors in panel (b) are not winsorized.

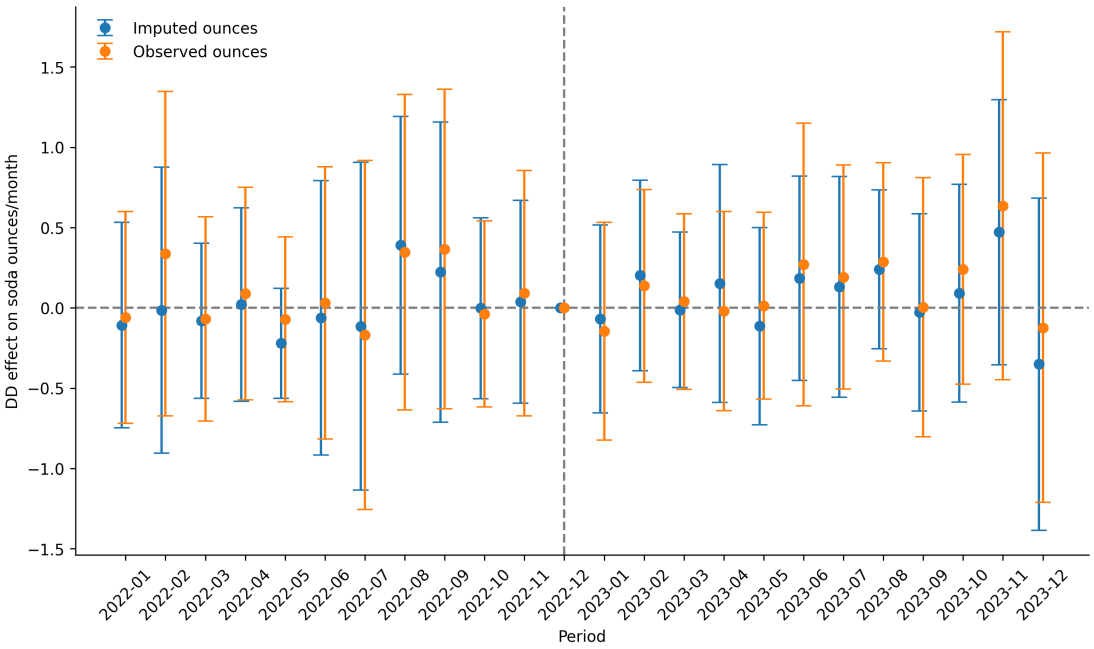
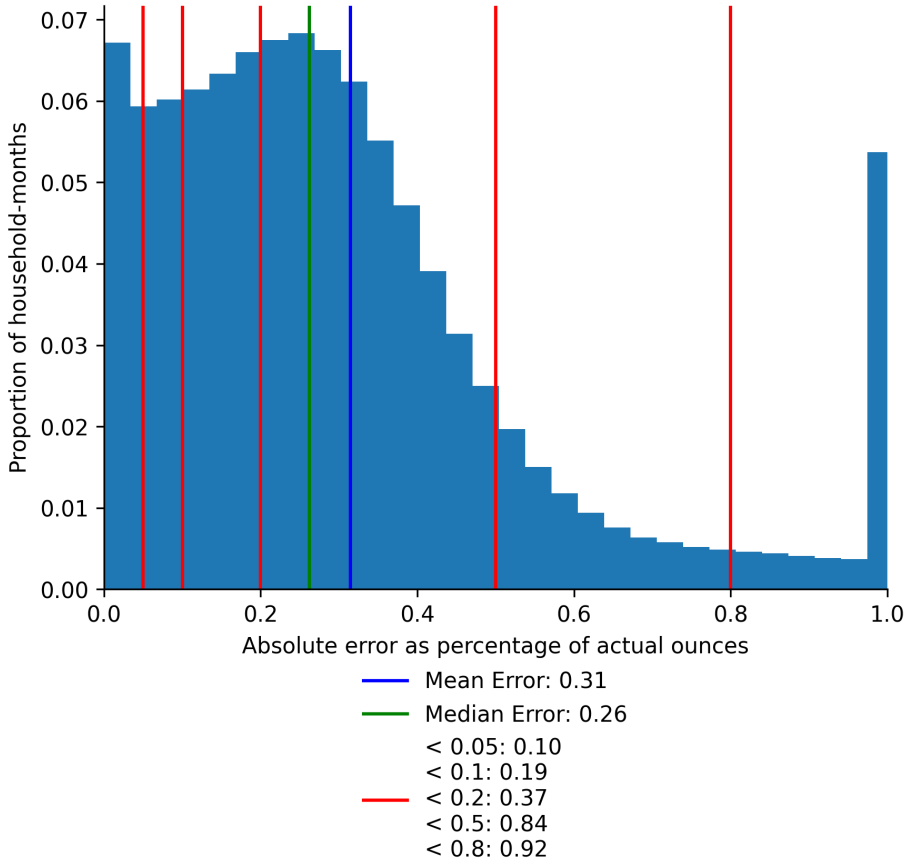


Figure 6: Event Study - Observed and Imputed Fluid Ounces

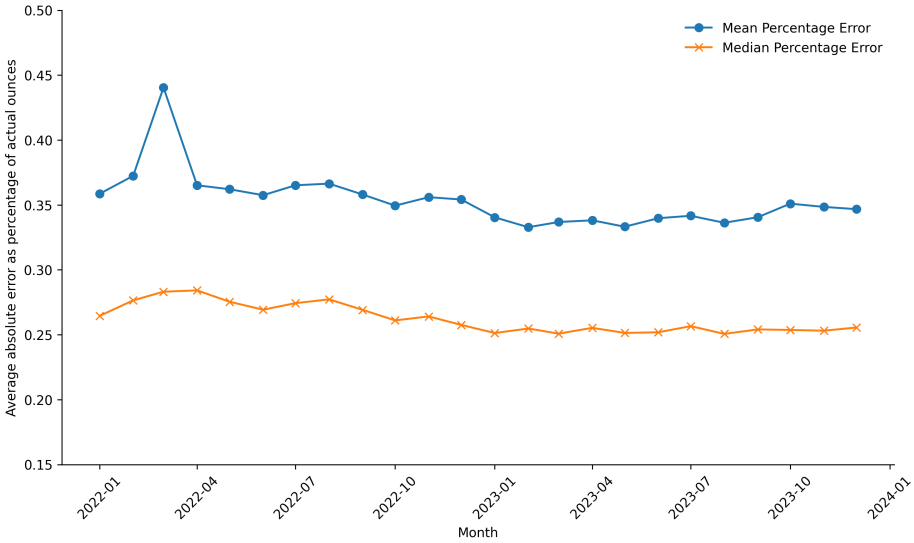
3.2 5-Fold Cross-Validation

Figure 7

(a) Absolute error as percentage of observed total fluid ounce per household-month



(b) Monthly average of absolute error as percentage of observed total fluid ounce per household-month



Notes: There are some outliers with large absolute errors. In panel (a) we winsorize at the 95th percentile for visualization purpose. Errors in panel (b) are not winsorized.

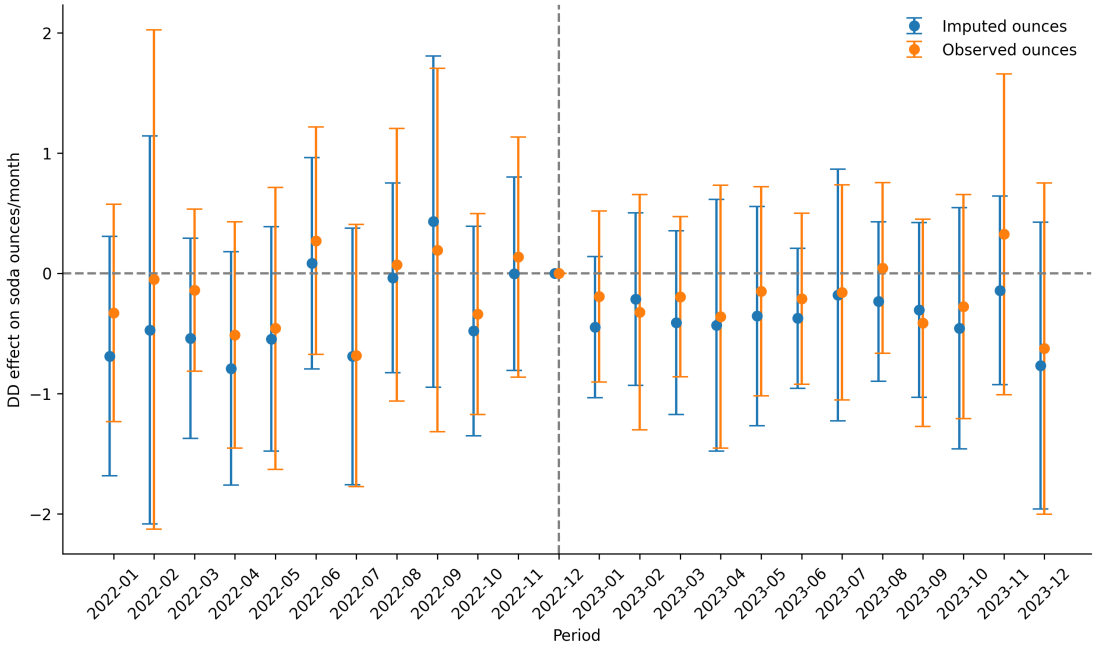


Figure 8: Event Study - Observed and Imputed Fluid Ounces

4 OLS and Post-Lasso

In this section, we conduct quantity validation by the following proposal to see how well we can predict 2023 Nielsen data by training on 2022 data:

1. Calculate price per fluid ounces in both year and adjust by 'Nonalcoholic beverages and beverage materials in U.S. city average, all urban consumers, **seasonally unadjusted**' (CPI).

2. Train a predictive model on 2022 data:

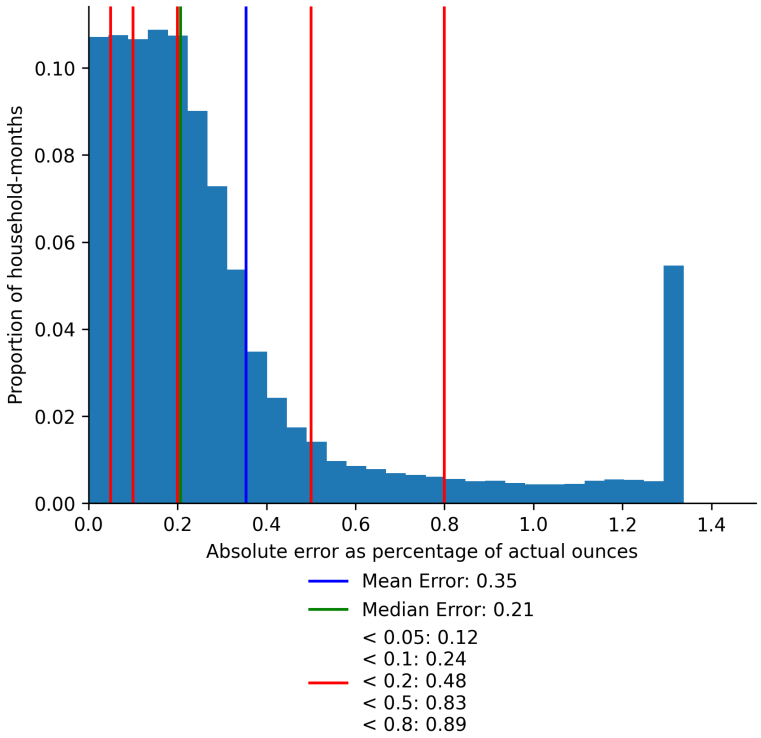
   - OLS with
     $\ln(\text{Price/ounce}) = \text{ChannelDummies} + \text{RegionDummies} + \ln(\text{HouseholdIncome}) + \ln(\text{Item Price}) + \text{MonthDummies} + \text{BrandFEs}$
   - Post-Lasso to run OLS on regressors with non-zero coefficients in first stage Lasso

3. Get predicted values for 2023 data for each observation in item table

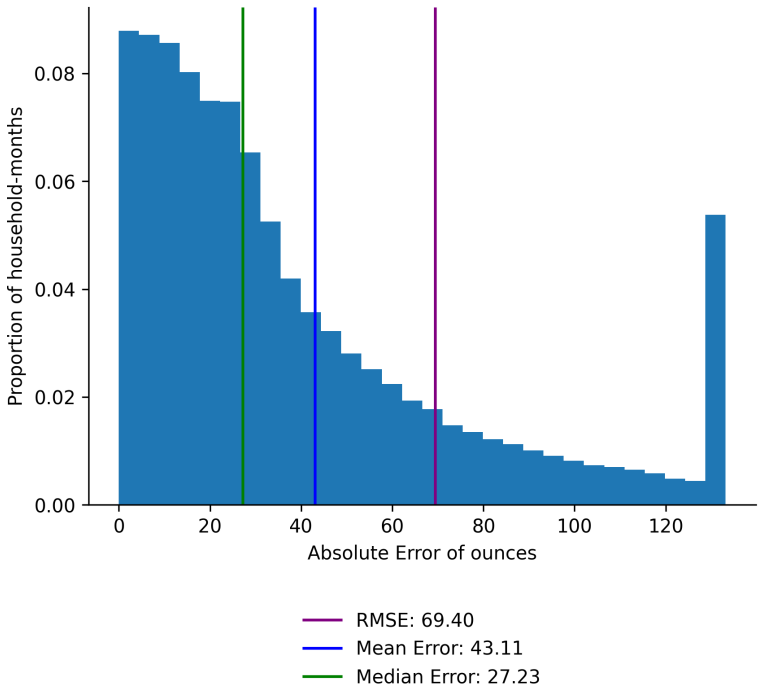
4.1 OLS

Figure 9

(a) Absolute error as percentage of observed total fluid ounce per household-month



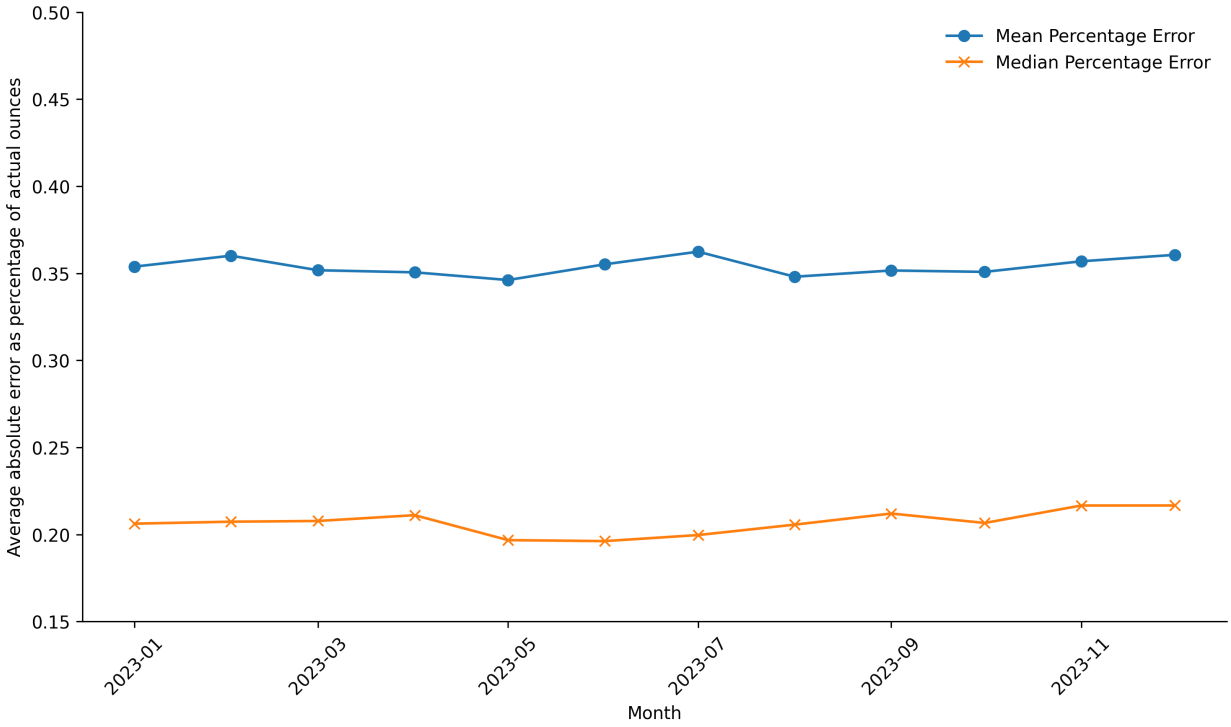
(b) Absolute error for household-month consumption



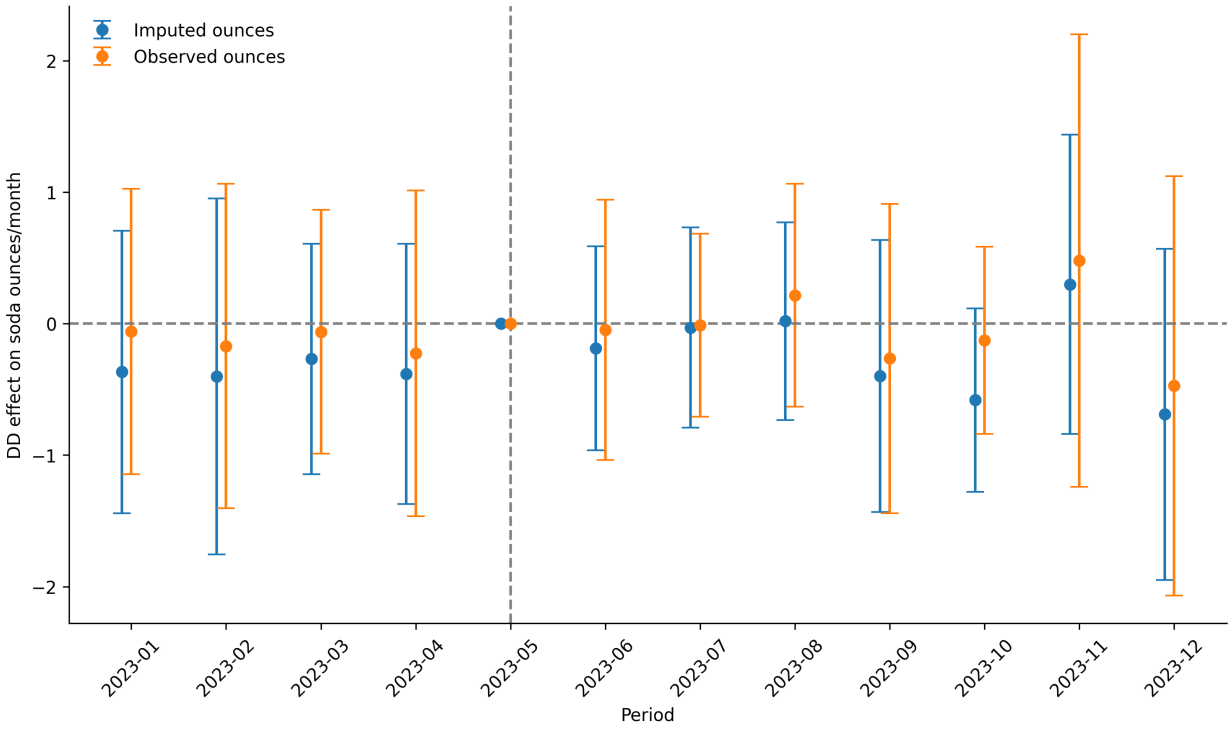
Notes: There are some outliers with large absolute errors. We winsorize at the 95th percentile for visualization purpose.

Figure 10

(a) Monthly average of absolute error as percentage of observed total fluid ounce per household-month

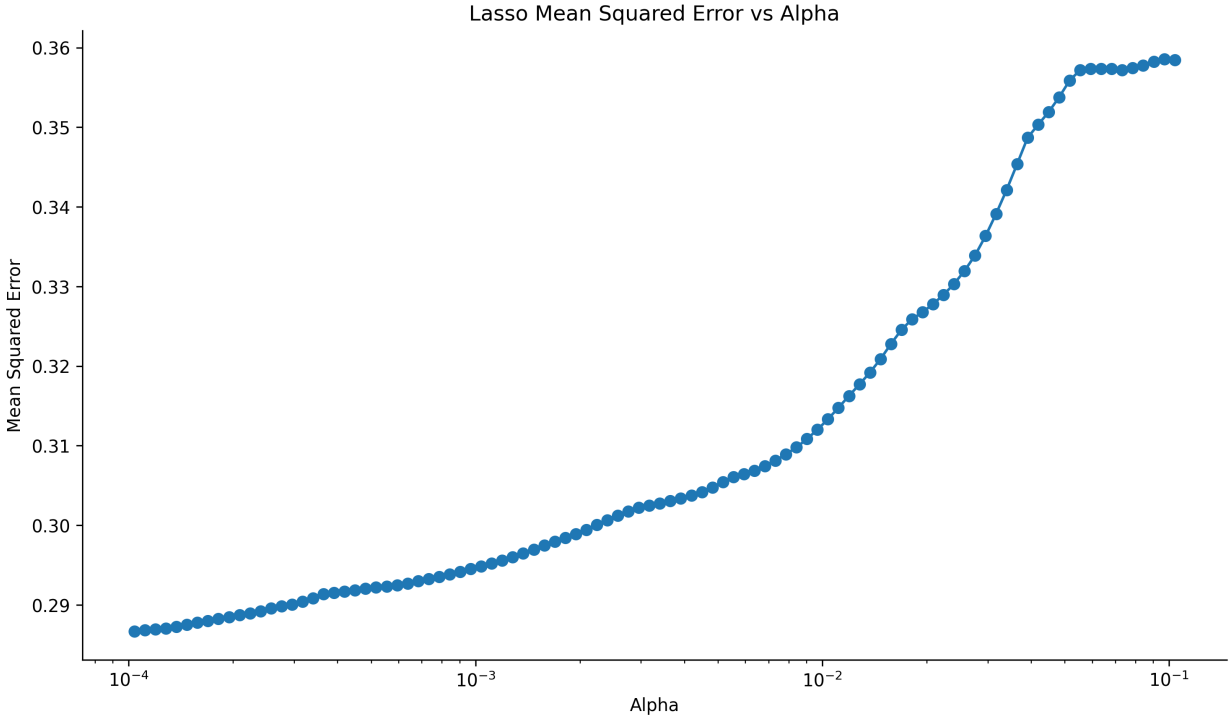


(b) Event Study - Observed and Imputed Fluid Ounces



4.2 Post-Lasso

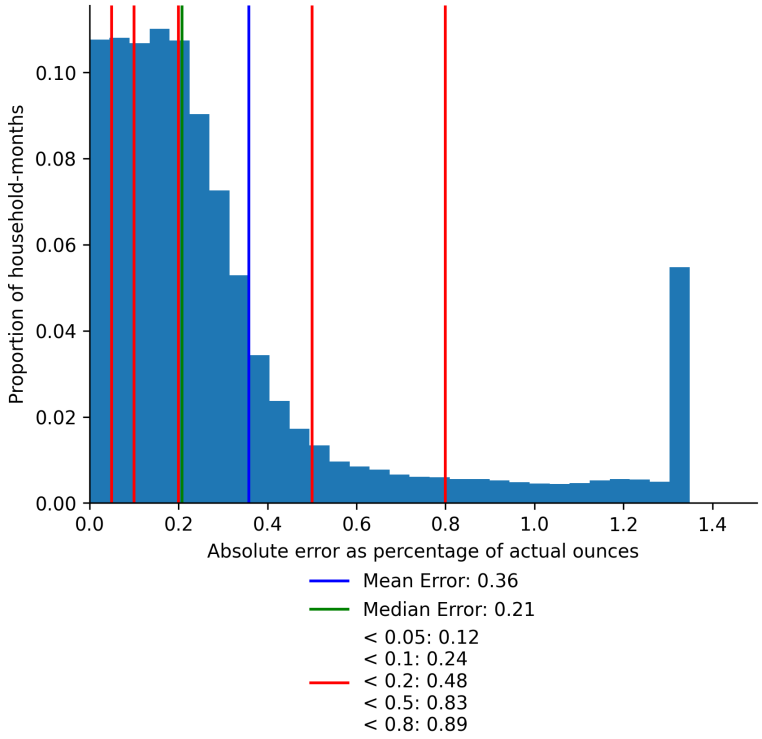
Figure 11: Mean Squared Error against Penalization Term



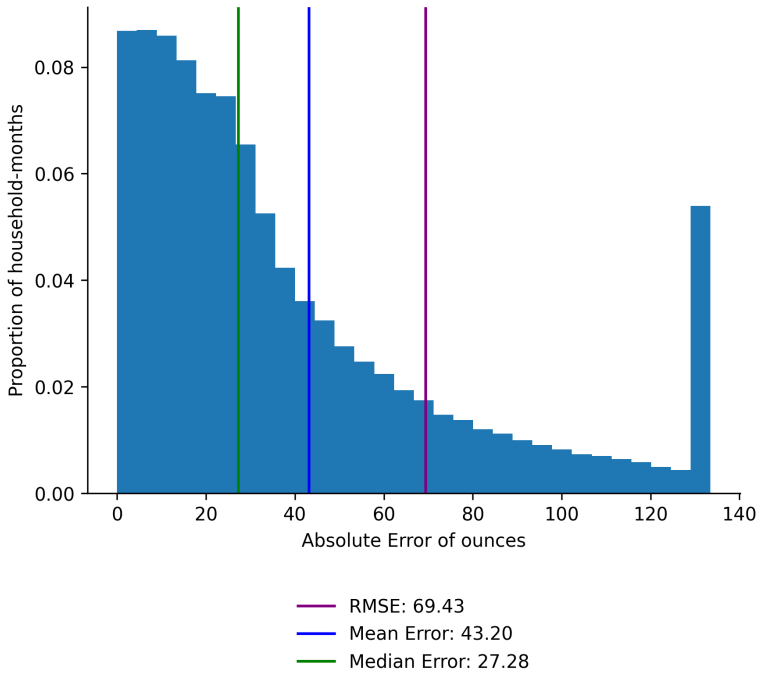
Notes: Best hyperparameter chosen by 10-fold CV.

Figure 12

(a) Absolute error as percentage of observed total fluid ounce per household-month



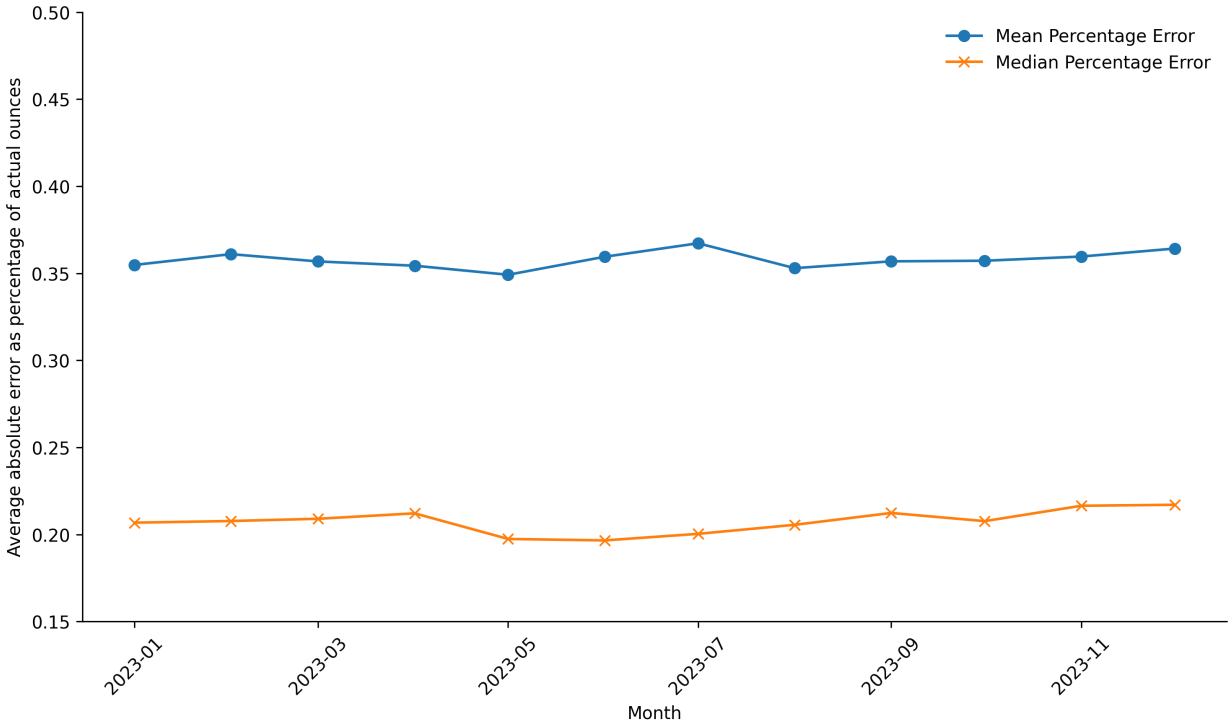
(b) Absolute error for household-month consumption



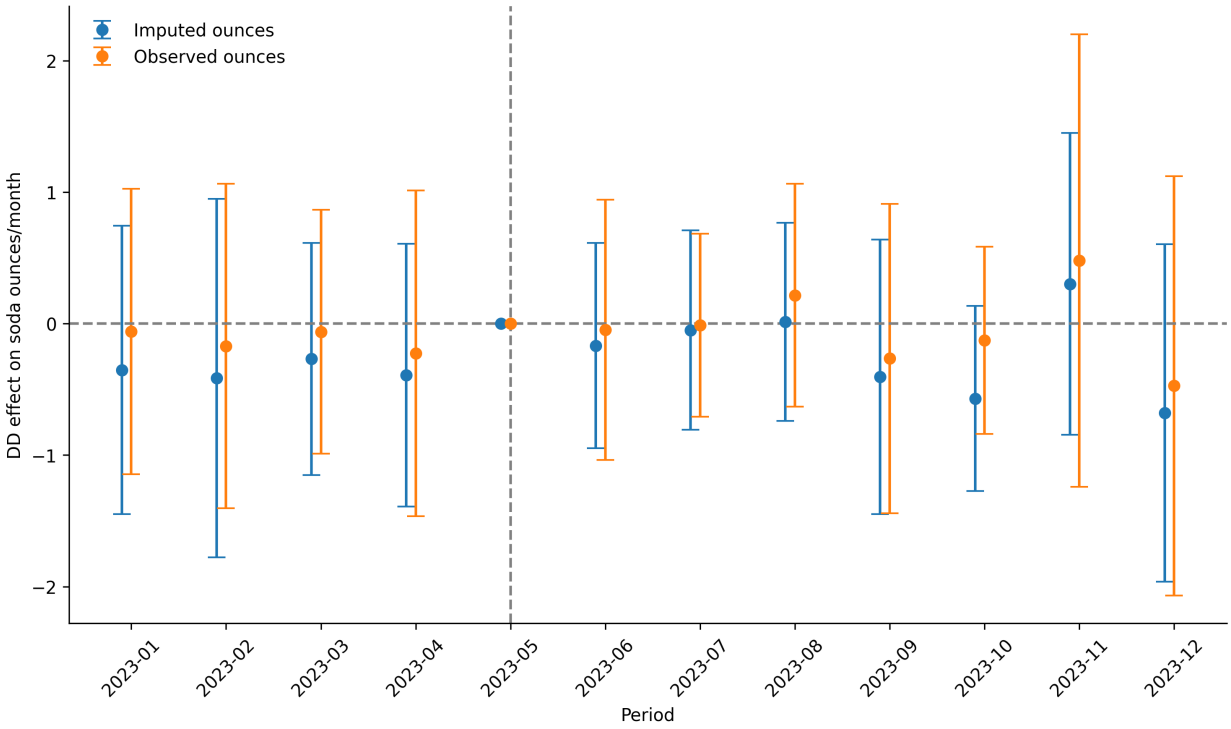
Notes: There are some outliers with large absolute errors. We winsorize at the 95th percentile for visualization purpose.

Figure 13

(a) Monthly average of absolute error as percentage of observed total fluid ounce per household-month



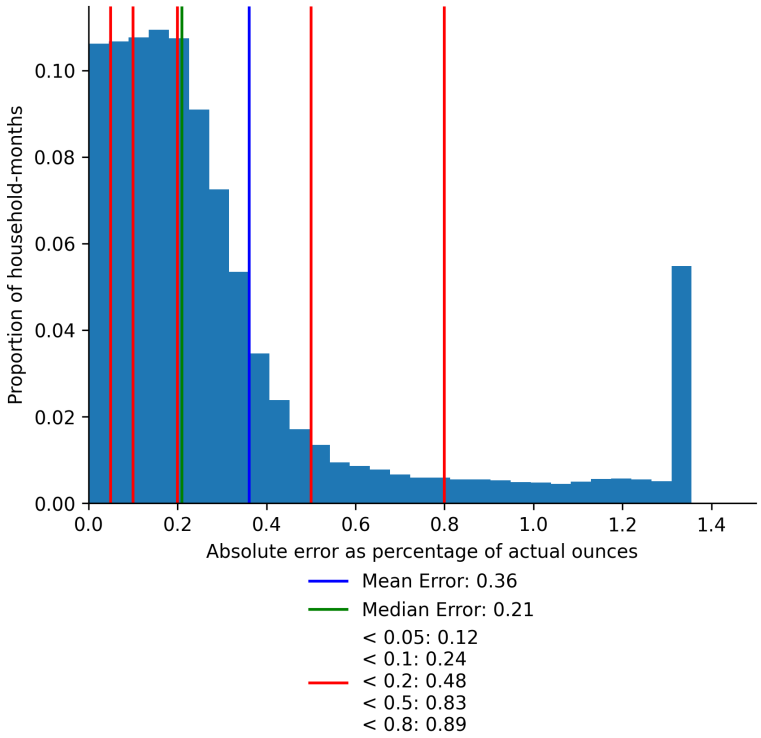
(b) Event Study - Observed and Imputed Fluid Ounces



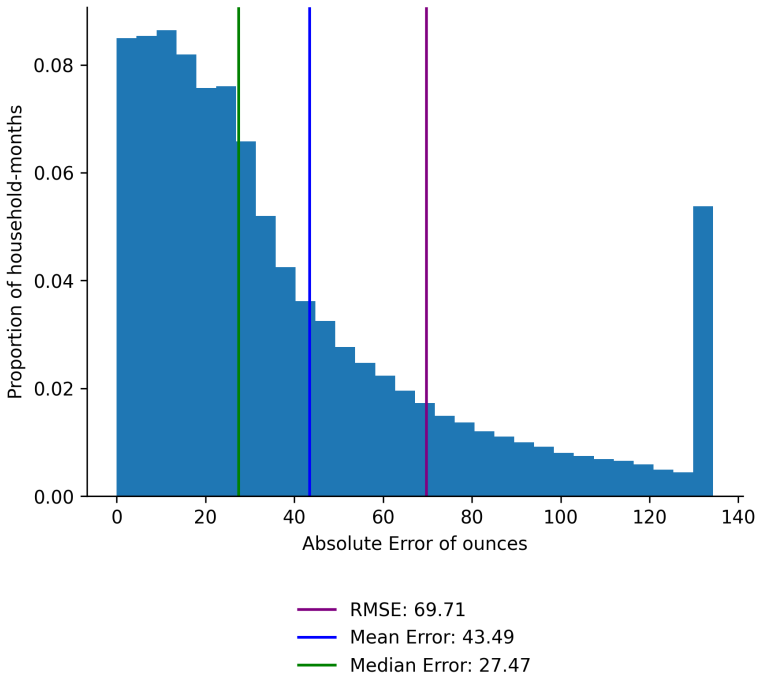
4.3 Lasso

Figure 14

(a) Absolute error as percentage of observed total fluid ounce per household-month



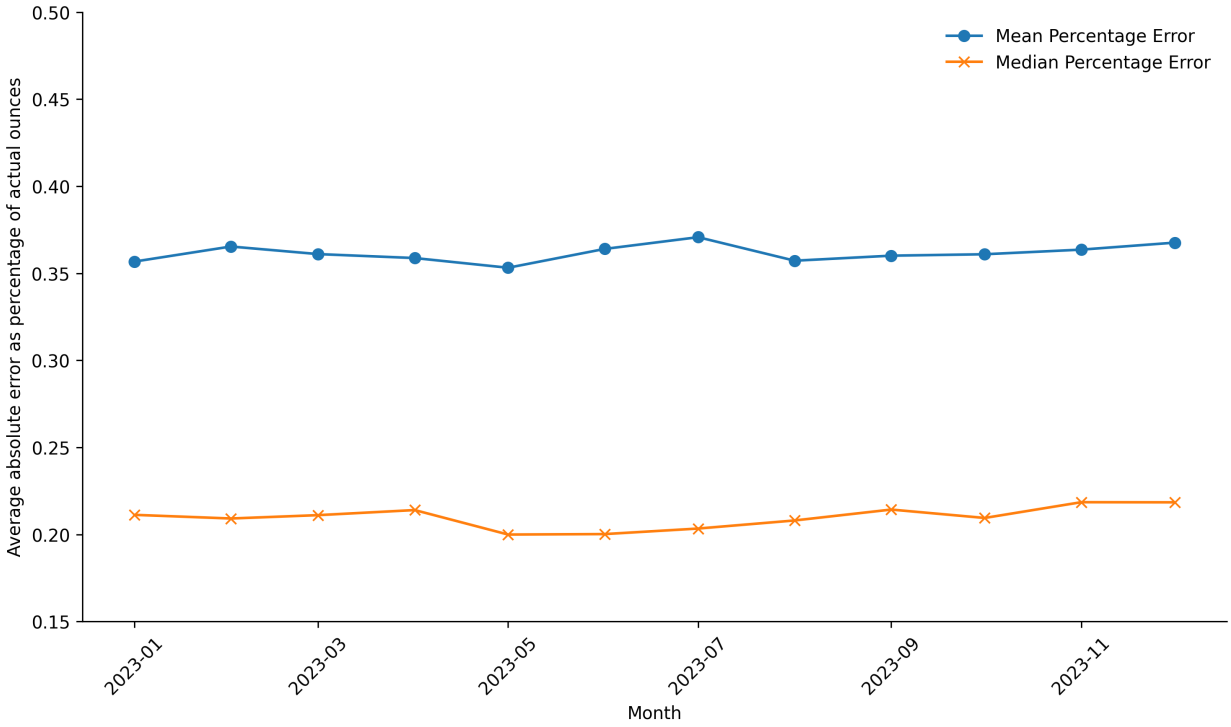
(b) Absolute error for household-month consumption



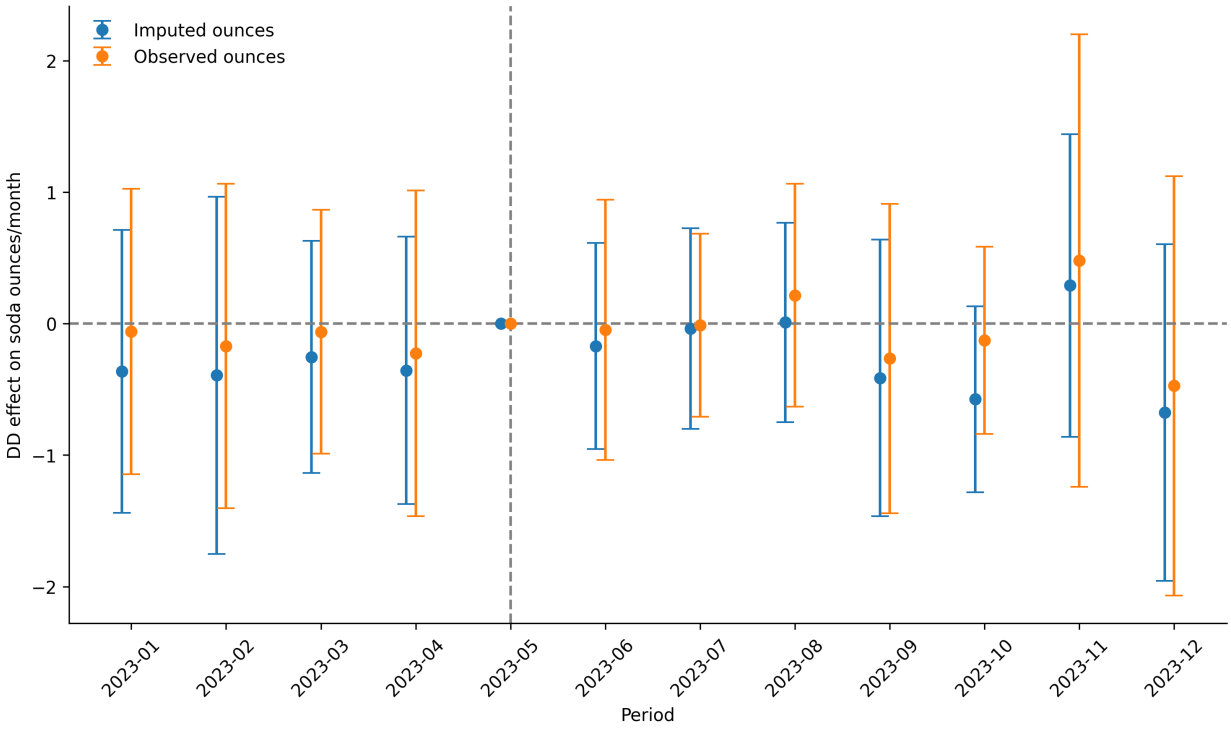
Notes: There are some outliers with large absolute errors. We winsorize at the 95th percentile for visualization purpose.

Figure 15

(a) Monthly average of absolute error as percentage of observed total fluid ounce per household-month



(b) Event Study - Observed and Imputed Fluid Ounces



5 Appendix

Table 1: OLS Results

	OLS Model
Channel type	
Baby & Toy	-0.2840***
	(0.1002)
Beauty	-0.3196***
	(0.0658)
Bodega / Convenience	0.0675***
	(0.0247)
Books / Media	0.2415***
	(0.0538)
Coop / Farm / Feed	-0.1442***
	(0.0302)
Discount / Dollar	-0.6313***
	(0.0245)
Drug Stores	-0.3292***
	(0.0247)
Electronics & Office	0.0482
	(0.0414)
FSR	-0.0074
	(0.0310)
Free Sample / Gift	-0.2937***
	(0.0327)
Grocery	-0.7165***
	(0.0245)
Health / Wellness	-0.7162***
	(0.0525)
Home Furnishings & Delivery	-0.0444
	(0.0349)
Home / Craft / Sporting Goods	-0.2752***
	(0.0256)
LSR / QSR	0.0724**
	(0.0281)
Liquor / Tobacco	-0.3082***
	(0.0268)
Manufacturer Outlet	-0.1092
	(0.3514)
Military	-0.6750***

Continued on next page

Table 1 continued

	OLS Model
	(0.0256)
Online / Remote Retail	-0.7109***
	(0.0247)
Optical Store	-0.1502*
	(0.0840)
Other Retail	-0.3402***
	(0.0251)
Pet Stores	-0.0410
	(0.0542)
Public Markets	-0.2695***
	(0.0952)
Specialty Food Retailer	-0.0851***
	(0.0299)
Transportation Services	0.0460*
	(0.0250)
Vending	0.0322
	(0.0252)
Month	
October	0.0301***
	(0.0024)
November	0.0122***
	(0.0024)
December	0.0153***
	(0.0024)
February	0.0194***
	(0.0024)
March	0.0245***
	(0.0023)
April	0.0196***
	(0.0023)
May	0.0021
	(0.0023)
June	-0.0025
	(0.0024)
July	-0.0057**
	(0.0023)
August	0.0205***
	(0.0024)
September	0.0370***

Continued on next page

Table 1 continued

	OLS Model
	(0.0024)
Census region	
Northeast	0.0287***
	(0.0016)
South	0.0213***
	(0.0012)
West	0.0645***
	(0.0016)
Continuous covariates	
Log household income	-0.0093***
	(0.0005)
Log item unit price	0.0549***
	(0.0005)
Model fit	
R-squared	0.3050
R-squared (Adj.)	0.3048

Table 2: Lasso Results

Feature	Coefficient
Channel type	
Baby & Toy	-0.0000
Beauty	-0.0000
Bodega / Convenience	0.2556
Books / Media	0.0000
Coop / Farm / Feed	0.0000
Discount / Dollar	-0.4359
Drug Stores	-0.1288
Electronics & Office	0.0000
FSR	0.0523
Free Sample / Gift	-0.0000
Grocery	-0.5188
Health / Wellness	-0.0000
Home Furnishings & Delivery	0.0463
Home / Craft / Sporting Goods	-0.0551
LSR / QSR	0.1830
Liquor / Tobacco	-0.0372
Manufacturer Outlet	0.0000
Military	-0.4531
Online / Remote Retail	-0.5085
Optical Store	0.0000
Other Retail	-0.1177
Pet Stores	0.0000
Public Markets	-0.0000
Specialty Food Retailer	0.0118
Transportation Services	0.2293
Vending	0.2123
Month	
October	0.0229
November	0.0044
December	0.0079
February	0.0119
March	0.0170
April	0.0126
May	-0.0029
June	-0.0074
July	-0.0104
August	0.0133

Continued on next page

Table 2 continued

Feature	Coefficient
September	0.0298
Census region	
Northeast	0.0285
South	0.0196
West	0.0651
Continuous covariates	
Log household income	-0.0083
Log item unit price	0.0561

Table 3: Post-Lasso OLS Results

	Post-Lasso OLS Model
Channel type	
Bodega / Convenience	0.2255***
	(0.0106)
Discount / Dollar	-0.4728***
	(0.0101)
Drug Stores	-0.1695***
	(0.0105)
FSR	0.1663***
	(0.0216)
Grocery	-0.5569***
	(0.0101)
Home Furnishings & Delivery	0.1580***
	(0.0267)
Home / Craft / Sporting Goods	-0.1187***
	(0.0125)
LSR / QSR	0.2314***
	(0.0172)
Liquor / Tobacco	-0.1336***
	(0.0149)
Military	-0.5162***
	(0.0126)
Online / Remote Retail	-0.5521***
	(0.0106)
Other Retail	-0.1774***
	(0.0115)
Specialty Food Retailer	0.0911***
	(0.0200)
Transportation Services	0.2054***
	(0.0113)
Vending	0.1932***
	(0.0117)
Month	
October	0.0306***
	(0.0024)
November	0.0122***
	(0.0024)
December	0.0157***
	(0.0024)

Continued on next page

Table 3 continued

	Post-Lasso OLS Model
February	0.0196***
	(0.0024)
March	0.0244***
	(0.0024)
April	0.0201***
	(0.0023)
May	0.0023
	(0.0023)
June	-0.0025
	(0.0024)
July	-0.0052**
	(0.0024)
August	0.0207***
	(0.0024)
September	0.0370***
	(0.0024)
Census region	
Northeast	0.0287***
	(0.0016)
South	0.0210***
	(0.0012)
West	0.0651***
	(0.0016)
Continuous covariates	
Log household income	-0.0090***
	(0.0005)
Log item unit price	0.0565***
	(0.0005)
Model fit	
R-squared	0.2994
R-squared (Adj.)	0.2994

Table 4: Nielsen Channel Mapping

Original Channel	**Mapped Category**
Grocery	Grocery
Bodega	Bodega/Convenience
Swapmeet Flea Market	Public Markets
Fish Market	Specialty Food Retailer
Cheese Stores	Specialty Food Retailer
Pizzeria	LSR/QSR
Home Furnishings	Home Furnishings & Delivery
Apparel Stores	Apparel & Footwear
Craft Stores	Home/Craft/Sporting Goods
Wireless Service Stores	Online/Remote Retail
Garden Stores	Home/Craft/Sporting Goods
Online Shopping	Online/Remote Retail
TV/Home Shopping	Online/Remote Retail
Drug Store	Drug Stores
Coffee Store/Gourmet Coffee Sh	LSR/QSR
Athletic Footwear	Apparel & Footwear
Sporting Goods	Home/Craft/Sporting Goods
Shoe Store	Apparel & Footwear
Pro Shop	Home/Craft/Sporting Goods
Butcher	Specialty Food Retailer
Restaurant	FSR
Fruit Stand	Specialty Food Retailer
Bakery	LSR/QSR
Dollar Store	Discount/Dollar
Convenience Store	Bodega/Convenience
Gas Mini Mart	Bodega/Convenience
Close Out Store	Discount/Dollar
Catalog Showroom	Catalog Showroom
Music/CD Store	Books/Media
Office Supplies Store	Electronics & Office

Continued on next page

Original Channel	Mapped Category
Video Store	Books/Media
Free Sample/Gift	Free Sample/Gift
Health Food Store	Health/Wellness
Party Supply Store	Other Retail
Candy Store	Specialty Food Retailer
Discount Store	Discount/Dollar
Hardware/Home Improvement	Home/Craft/Sporting Goods
Automotive Store	Transportation Services
Quick Serve Restaurants	LSR/QSR
Kennel/Vet	Pet Stores
Dairy Store	Specialty Food Retailer
Tobacco Store	Liquor/Tobacco
Manufacturer Outlet	Manufacturer Outlet
Military Store	Military
Electronics Store	Electronics & Office
Optical Store	Optical Store
Computer Store	Electronics & Office
Toy Store	Baby & Toy
Service Station	Transportation Services
Beverage Store	Specialty Food Retailer
Coop/Farm/Feed	Coop/Farm/Feed
Hypermarket	Grocery
Beauty Supply Store	Beauty
Barber/Salon	Beauty
Vending Machine	Vending
Home Delivery	Home Furnishings & Delivery
Home Inventory	Home Furnishings & Delivery
Department Store	Other Retail
Warehouse Club	Grocery
Mail Order	Online/Remote Retail
Pet Store	Pet Stores

Continued on next page

Original Channel	**Mapped Category**
News/Book Store	Books/Media
Delicatessen	Specialty Food Retailer
Liquor Store	Liquor/Tobacco
Camera Shop	Camera Shop
Stationery Store	Books/Media
All Other Stores	Other Retail

Table 5: Numerator Channel Mapping

Original Channel	**Mapped Category**
Food	Grocery
Pet	Pet Stores
Drug	Drug Stores
Bodega	Bodega/Convenience
Gas & Convenience	Bodega/Convenience
Liquor	Liquor/Tobacco
Other Specialty Store	Other Retail
Other	Other Retail
Sporting Goods Stores	Home/Craft/Sporting Goods
Craft	Home/Craft/Sporting Goods
Dispensaries	Liquor/Tobacco
LSR - Sandwich/Deli	LSR/QSR
LSR - Ethnic/Regional	LSR/QSR
LSR - Bakery/Cafe	LSR/QSR
LSR - Mexican	LSR/QSR
LSR - Chicken	LSR/QSR
LSR - Pizza	LSR/QSR
LSR - Burger	LSR/QSR
LSR - Salad/Healthful	LSR/QSR
LSR - Dessert Snack	LSR/QSR
LSR - Coffee/Bakery	LSR/QSR
LSR - Miscellaneous	LSR/QSR
FSR - Regional/Ethnic	FSR
FSR - Midscale	FSR
FSR - Sports Bar	FSR
FSR - Seafood/Steak	FSR
FSR - Miscellaneous	FSR
FSR - American	FSR
FSR - Italian/Pizza	FSR
Apparel	Apparel & Footwear

Continued on next page

Original Channel	Mapped Category
Shoe	Apparel & Footwear
Online	Online/Remote Retail
Book	Books/Media
Dollar	Discount/Dollar
Postal Services	Other Retail
Beauty	Beauty
Shopping Centers and Malls	Other Retail
Discount Store	Discount/Dollar
Farm	Other Retail
Mass	Grocery
Home Furnishings	Home Furnishings & Delivery
Retail Services	Business/Professional Services
Health	Health/Wellness
Other Retail Store	Other Retail
CloseOut	Discount/Dollar
Baby & Toy	Baby & Toy
Sports and Recreation	Home/Craft/Sporting Goods
Yard Services	Business/Professional Services
Auto Services	Transportation Services
Movie Theatre	Travel & Entertainment
Taxi or Limousine	Transportation Services
Telecom	Online/Remote Retail
Other Entertainment	Travel & Entertainment
Electronics	Electronics & Office
Transport Hub	Transportation Services
Healthcare	Medical/Healthcare
Wireless	Online/Remote Retail
Tobacco Shops	Liquor/Tobacco
Vapor Stores	Liquor/Tobacco
Office	Electronics & Office
Auto	Transportation Services

Continued on next page

Original Channel	Mapped Category
Department Store	Other Retail
Hotels & Resorts	Travel & Entertainment
Specialty Food Retailer	Specialty Food Retailer
School Tuition & Fees	Education/Childcare
Outlet Store	Other Retail
Wholesale	Grocery
Music Stores	Books/Media
Florists	Other Retail
Other Services	Business/Professional Services
Charities	Other Retail
Professional Services	Business/Professional Services
Veterinarians	Pet Stores
Banks	Financial Services
Child Care	Education/Childcare
Laundromat	Business/Professional Services
Video	Books/Media
Spas	Health/Wellness
Amusement Parks	Travel & Entertainment
Parking Lot or Garage	Transportation Services
Security Services	Business/Professional Services
Concert Hall or Theater	Travel & Entertainment
Public Markets	Public Markets
Mortgage Payment	Financial Services
Ticket Outlet	Other Retail
Other Government Payments	Government/Utilities
Gambling	Travel & Entertainment
Other Association Fees	Other Retail
Airline	Travel & Entertainment
Cruise Line	Travel & Entertainment
Doctors Office	Medical/Healthcare
Other Travel	Travel & Entertainment

Continued on next page

Original Channel	**Mapped Category**
Car Rental	Transportation Services
Public Storage	Transportation Services
Sports Entertainment	Travel & Entertainment
Military	Military
Copy Centers	Business/Professional Services
Travel Agency	Travel & Entertainment
Real Estate Services	Business/Professional Services
Dance & Comedy Clubs	Travel & Entertainment
Club	Travel & Entertainment
Utility Company	Government/Utilities
Media	Books/Media
Church Offerings	Other Retail
unknown	Unknown
Movers	Business/Professional Services