

# Gasoline Demand Estimation

Daniel Posthumus

March 7, 2025

## 1 Introduction

In this memo, I explain and present results from **preliminary** demand estimation in the California gasoline markets. Specifically, I address the following:

- Describing the merged dataset of A15 and OPIS Retail Data, along with external datasources.
- Presenting Pure Logit estimation and results.
- Presenting Nested Logit estimation and results.

## 2 Data

### 2.1 Sources of Data

I use the following data series

- Daily, gas station-level retail prices from OPIS.
  - I collapse on the (unweighted) average for each gas station in a given year.
- Annual, gas station-level A15 data. This includes quantity and station characteristics.
- EIA’s annual series, ”U.S. Gulf Coast Conventional Gasoline Regular Spot Price FOB”.
- Census data on the total population of Zip Code Tabulation Areas (ZCTAs).

### 2.2 Defining Market Size

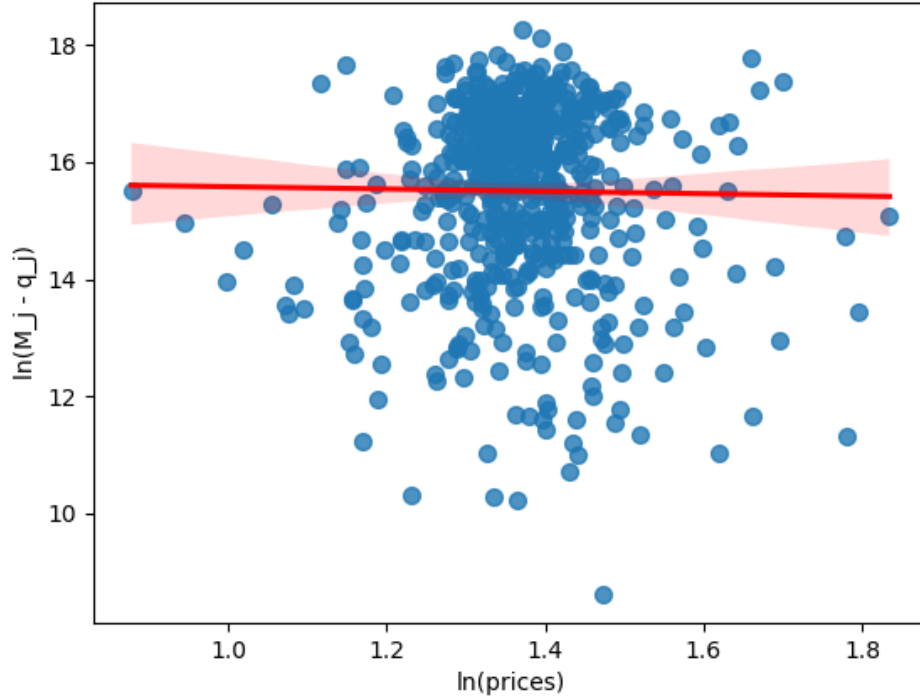
While we have observed gas station-level quantities sold, this doesn’t allow us to directly calculate observed market shares; instead, we need the denominator—market size  $M_t$ . This isn’t as simple as summing the observed quantities sold, as we need to capture ‘consumption’ of the outside option in each market, effectively all consumers that did *not* purchase gasoline for a given zip code-year.

In practice, estimates of market size are very rough; in BLP’s seminal paper (1995), for example, they simply took the number of registered vehicles in all of the United States as an estimate for the market size  $M_t$  of the national automobile market. Clearly, this is only a rough approximation—many people would be happy with their cars and not looking to buy a new one, regardless of price changes.

Our need for the market size is simple: we need a way to exogenously determine the outside good share, so the outside good share is properly scaled across different zip codes with very different market sizes (such as a zip code in downtown San Francisco compared to a zip code in the north Central Valley). To accomplish this, I pull population figures from the census at a Zip Code Tabulation Area (ZCTA) and multiply this population figure by 489 gallons/person. 489 gallons is the estimated average consumption of gasoline for a registered car nationally. This resulted in some non-zero share (2-3%) of negative outside good shares (i.e., instances where the observed quantity of gasoline sold in a given year exceeded the population times 489). To reduce the share of instances where this occurred, I multiplied all market sizes by 2; this would not bias our results as it is constantly scaled for all markets and years.

We can quickly check whether this market share metric makes sense by examining the relationship between the outside good’s share (represented by  $s_{0t}$  for market  $t$ ), such as in the following scatterplot, with a line of best fit superimposed on it:

Figure 1: Log Price and Log Outside Good Quantity



### 2.3 Descriptive Statistics

Here is a table containing descriptive statistics of markets, as defined by Zip-Years. The rest of the paper will proceed with this definition of a market, on the Zip-Year level.

Table 1: Summary Statistics by Market Definition

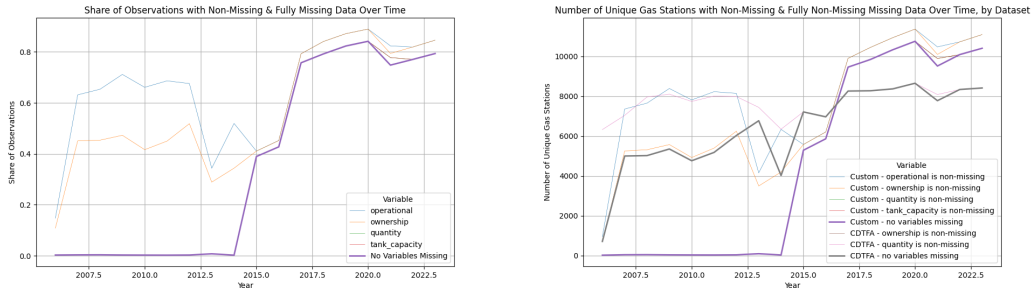
|  | ZIP-Year |
|--|----------|
| Mean Number of Stations                                  | 4.22     |
| $s_{0t}$   | 0.81     |
| $s_{jt}$   | 0.063    |
| Average % of Stations That Are Unbranded                 | 16.2     |
| Average Station-Level Observed Quantity Sold             | 1123167  |
| Average Market Coefficient of Variation on Quantity Sold | 0.71     |
| Average Retail Prices                                    | 3.8      |
| Average Market Coefficient of Variation on Prices        | 0.05     |
| Total Number of Markets                                  | 9994     |

Here is a table showing summary statistics by store brand (I selected the top 10 most common brands in California by the number of unique gas stations belonging to each brand):

Table 2: Summary Statistics by Store Brand

| Store Brand | Retail Prices |          |        |          |      | Total # of Stations | % of Observed Stations |
|-------------|---------------|----------|--------|----------|------|---------------------|------------------------|
|             | Min.          | 25pctile | Median | 75pctile | Max. |                     |                        |
| unbranded   | 2.05          | 3.00     | 3.50   | 4.17     | 6.66 | 1453                | 14.86                  |
| chevron     | 2.29          | 3.24     | 3.73   | 4.45     | 8.70 | 1309                | 13.39                  |
| shell       | 2.22          | 3.23     | 3.73   | 4.48     | 7.82 | 943                 | 9.65                   |
| 76          | 2.16          | 3.21     | 3.66   | 4.35     | 6.67 | 855                 | 8.75                   |
| ampm        | 2.28          | 2.96     | 3.43   | 4.50     | 6.12 | 598                 | 6.12                   |
| valero      | 2.40          | 3.12     | 3.63   | 4.53     | 6.88 | 565                 | 5.78                   |
| arco        | 2.26          | 2.93     | 3.41   | 4.03     | 6.10 | 498                 | 5.09                   |
| 7_eleven    | 2.35          | 3.02     | 3.47   | 4.81     | 6.51 | 452                 | 4.62                   |
| mobil       | 2.38          | 3.30     | 3.74   | 4.54     | 6.67 | 429                 | 4.39                   |
| extramile   | 3.03          | 3.65     | 3.91   | 5.15     | 6.20 | 342                 | 3.50                   |

Then, from PIIRA's A15 Form, we have data on station characteristics. First, there are two possible A15 datasets: 1) the dataset CDTFA prepared and 2) the dataset I compiled from raw A15 data. As the graphics below make clear, the CDTFA-compiled data has improved data coverage pre-2014, while post-2014 the customly-put together A15 data has superior coverage in raw number of stations, as well as more variables.



(a) Share of Non-Missing Observations by Variable (b) Number of Unique Gas Stations by Dataset

Figure 2: Data Coverage of A15 Data Over Time

Next, we can get a sense of what these variables mean. Below, I've compiled summary statistics for each of the A15 variables, broken out by 1) all stations, 2) branded non-hypermarket stations, and 3) hypermarket stations, and 4) unbranded stations. **Note:** one point of confusion with respect to the A15 data is how to treat missing amenity data; for the purpose of this project, I've assumed that where amenity data is missing and all other variables are not missing, that station does not have any amenities.

Table 3: Summary Statistics of Station Characteristics from PIIRA A15 Data

|                                | All Stations | Unbranded | Hypermarket | Branded   |
|--------------------------------|--------------|-----------|-------------|-----------|
| Unique Station Count           | 9775.00      | 1453.00   | 150.00      | 9625.00   |
| Mean Tank Capacity             | 14063.29     | 12768.54  | 24615.31    | 13802.68  |
| Mean Quantity                  | 1157251.95   | 625257.58 | 8274247.80  | 981473.19 |
| Mean Year                      | 2019.37      | 2019.39   | 2019.06     | 2019.38   |
| % Company Owned/Operated       | 33.05        | 27.40     | 95.93       | 31.49     |
| % Franchisee/Lessee Operated   | 22.32        | 3.80      | 3.88        | 22.78     |
| % Independently Owned/Operated | 44.13        | 68.47     | 0.19        | 45.22     |
| % Service Station/Gas Station  | 93.51        | 88.24     | 14.73       | 95.46     |
| % Truck Stop                   | 0.30         | 0.56      | 0.00        | 0.31      |
| % Hypermart                    | 2.75         | 0.52      | 84.30       | 0.73      |
| % Automotive Repair Bay        | 1.08         | 2.05      | 0.00        | 1.11      |
| % Car Wash                     | 0.71         | 0.85      | 0.00        | 0.73      |
| % Convenience Store            | 75.14        | 67.40     | 1.74        | 76.96     |
| % Equipment Rental             | 0.00         | 0.00      | 0.00        | 0.00      |
| % Kiosk                        | 6.71         | 3.71      | 35.27       | 6.00      |
| % Propane                      | 0.00         | 0.00      | 0.00        | 0.00      |
| % Restaurant/Fast Food         | 0.06         | 0.05      | 0.00        | 0.06      |
| % Supermarket/General Store    | 0.51         | 1.21      | 7.75        | 0.33      |
| % Truck Service Shop           | 0.00         | 0.00      | 0.00        | 0.00      |

Regarding the amenity data, we can see the hypermarkets are much likelier to have a kiosk or a supermarket general store, while the vast majority of non-hypermarket gas stations, branded or unbranded, have an attached convenience store – although, interestingly, the rate is significantly lower for unbranded stations than branded non-hypermarket stations (the same is true for kiosks).

### 3 Pure Logit

#### 3.1 Estimation

As discussed further below in the appendix, by assuming there is no systematic consumer heterogeneity in preferences, we are left with the following equation:

$$\ln(s_{jt}) - \ln(s_{0t}) = \ln\left(\frac{s_{jt}}{s_{0t}}\right) = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} \quad (1)$$

This equation is fully identified; market shares, including for the outside good; prices; and product characteristics are all observed. This allows us to estimate the equation using OLS or 2SLS. The motivation for 2SLS is clear; price is endogenous, as it will be correlated with unobserved demand shocks, represented by  $\xi_{jt}$ . There are several possible instruments for price; for now, we are using one type—a cost shifter that changes marginal cost in an exogenous fashion—RBOB spot prices on the gulf coast. RBOB represents one input cost for producing CARBOB, the CA-specific blend of gasoline required by law to be sold in the state. However, the marginal barrel of gasoline sold in California originates from Asia, so we can conclude that demand in California would not substantively shift RBOB spot prices on the Gulf Coast.

For now, the only non-price characteristics included are selected brand fixed effects; I selected the top 5 most common brands in the data as well as hypermarket gas stations and included fixed effects for these 6 categories of stations. Thus, the estimates of the fixed effects should be interpreted as relative to all gas stations in the market outside of the top 5 most common brands (which appear to align with refiner brands) and hypermarkets. I included a constant; all fixed estimates should be interpreted relative to unbranded stations.

Derived in the appendix, the elasticities for this pure logit model take the following forms (own- and cross-price elasticities, respectively):

$$\eta_{jjt} = \alpha \cdot p_{jt} \cdot (1 - s_{jt}) \quad (2)$$

$$\eta_{jkt} = -\alpha \cdot p_{kt} \cdot s_{kt} \quad (3)$$

The elasticities reported in the results table below are the simple average across all products in all markets.

## 3.2 Results

First, we start with the simplest model, where there is no product differentiation. We have just two products: the inside good and the outside good. To formalize this market, there are only  $j \in 0, 1$  products where  $j_{0t}$  is the outside good and  $j_{1t}$  is the inside good. We calculate  $s_{1t}$  as the sum of the observed shares in the data. We take  $p_{1t}$  as the market share-weighted average price of all observations in the data for market  $t$ . Since there is no intra-market variation in prices, we are not able to use ZIP fixed effects, and run two specifications: 1) without and 2) with the RBOB Gulf Coast Spot Price as an instrument for retail price.

The results from this estimation are reported below in Panel A, with the first stage coefficient on the RBOB instrument in Panel B, the results of which indicate the instrument is quite strong.

Table 4: Pure Logit Results Without Product Differentiation

|   | (1)                  | (2)                 |
|---|----------------------|---------------------|
| <b>Panel A: Estimation Results</b>      |                      |                     |
| prices                                  | -0.103***<br>(0.012) | -0.021<br>(0.014)   |
| $\eta_{jkt}^-$ (cross-price elasticity) | nan                  | nan                 |
| $\eta_{jjt}^-$ (own-price elasticity)   | -0.324               | -0.065              |
| N                                       | 9989                 | 9989                |
| RBOB Instrument                         | NO                   | YES                 |
| <b>Panel B: First-Stage Results</b>     |                      |                     |
| RBOB Instrument                         | NA                   | 1.501***<br>(0.004) |

<sup>a</sup> Significance levels: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Next, I estimated the pure logit model *with* product differentiation among the inside goods, introducing brand fixed effects for the top 5 most common brands and unbranded stations. I've included results estimated directly using either OLS or IV in Panel A. I've also included the first stage coefficient on the RBOB instrument in Panel B.

Table 5: Pure Logit Results With Product Differentiation

|   | (1)                  | (2)                  | (3)                  | (4)                  |
|---|----------------------|----------------------|----------------------|----------------------|
| <b>Panel A: Estimation Results</b>      |                      |                      |                      |                      |
| prices                                  | -0.130***<br>(0.006) | -0.154***<br>(0.004) | -0.075***<br>(0.007) | -0.071***<br>(0.005) |
| brand chevron                           | 0.386***<br>(0.023)  | 0.446***<br>(0.018)  | 0.369***<br>(0.023)  | 0.423***<br>(0.018)  |
| brand shell                             | 0.320***<br>(0.025)  | 0.397***<br>(0.019)  | 0.303***<br>(0.025)  | 0.375***<br>(0.019)  |
| brand 76                                | 0.063**<br>(0.025)   | 0.191***<br>(0.021)  | 0.051**<br>(0.025)   | 0.176***<br>(0.021)  |
| brand ampm                              | 0.922***<br>(0.025)  | 1.212***<br>(0.022)  | 0.921***<br>(0.025)  | 1.206***<br>(0.022)  |
| brand 7 eleven                          | -0.023<br>(0.027)    | 0.242***<br>(0.021)  | -0.031<br>(0.026)    | 0.227***<br>(0.021)  |
| hypermarket                             | 2.543***<br>(0.041)  | 2.654***<br>(0.034)  | 2.554***<br>(0.042)  | 2.667***<br>(0.034)  |
| other brand                             | 0.321***<br>(0.020)  | 0.507***<br>(0.016)  | 0.312***<br>(0.020)  | 0.491***<br>(0.016)  |
| amenity car wash                        | -0.949***<br>(0.083) | -0.855***<br>(0.078) | -0.951***<br>(0.083) | -0.859***<br>(0.078) |
| amenity convenience store               | 0.198***<br>(0.018)  | 0.189***<br>(0.014)  | 0.206***<br>(0.018)  | 0.200***<br>(0.014)  |
| amenity restaurant fast food outlet     | 0.855**<br>(0.428)   | 0.120<br>(0.172)     | 0.851**<br>(0.429)   | 0.115<br>(0.172)     |
| amenity supermarket general store       | 0.740***<br>(0.094)  | 0.297***<br>(0.068)  | 0.764***<br>(0.093)  | 0.337***<br>(0.068)  |
| amenity kiosk                           | 0.374***<br>(0.027)  | 0.518***<br>(0.022)  | 0.394***<br>(0.027)  | 0.550***<br>(0.022)  |
| $\eta_{jkt}^-$ (cross-price elasticity) | 0.013                | 0.016                | 0.008                | 0.007                |
| $\eta_{jkt}^-$ (own-price elasticity)   | -0.479               | -0.568               | -0.275               | -0.263               |
| N                                       | 42675                | 42675                | 42675                | 42675                |
| ZIP CODE FE                             | NO                   | YES                  | NO                   | YES                  |
| RBOB Instrument                         | NO                   | NO                   | YES                  | YES                  |
| <b>Panel B: First-Stage Results</b>     |                      |                      |                      |                      |
| RBOB Instrument                         | NA                   | NA                   | 1.502***             | 1.497***             |
|   | NA                   | NA                   | (0.003)              | (0.003)              |

<sup>a</sup> Significance levels: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

### 3.3 Aggregate Elasticities

While *mean* elasticities are constructive, we also want to understand the aggregate elasticity – which will give us an idea about the extensive margin, or substitution to the outside good.

We can start by expressing the total share of inside goods across all markets with respect to mean utility, denoted

as  $\delta_{jt}$  for product  $j$  in market  $t$ .

$$\sum_{l \in T} \sum_{k \in J_l} s_{kl} = \sum_{l \in T} \frac{\sum_{k \in J_l} \exp(\delta_{kl})}{1 + \sum_{k \in J_l} \delta_{kl}} \quad (4)$$

where  $T$  represents the set of all markets and  $J_l$  represents the set of all products in market  $l$ .

We can observe that markets are perfectly segmented. Say a characteristic of product 1 in market 1 changes and all other products/markets are held constant. Consumption patterns in market 1 will change, as  $\delta_{1,1}$  enters into market share equation for all products in market 1. However, consumption patterns in any market  $t \neq 1$  will remain unchanged.

Let's say we want to the aggregate elasticity with respect to average price, constant across all products and markets denoted as  $\bar{p}$ . Then we can re-write the mean utility function to be the following:

$$p_{jt} = \bar{p} + e_{jt} \quad (5)$$

$$\delta_{jt} = \beta X_{jt} - \alpha_1(\bar{p}) + \alpha_2(e_{jt}) + \epsilon_{jt} \quad (6)$$

this is directly estimable and we can easily derive with respect to average price:

$$\frac{\partial \delta_{jt}}{\partial \bar{p}} = -\alpha_1 \quad (7)$$

While this sounds promising in theory, we are caught in something of a bind;  $\bar{p}$  is constant for every observation – thus computationally it will be calculated as part of the intercept (with perfect collinearity between it and the constant term). Yet, if we introduce variation in the mean price – say, at the ZIP level – we can't aggregate our elasticity. **Elasticity can only be aggregated at the level of aggregation at which we average price.** Thus, we want some variation but not too much.

An elegant solution is to observe elasticity at the year level; average price will be subscripted at the year-level (denoted by the subscript  $y$ ) and we can derive different statewide aggregate elasticities by year. Then, let's re-write our equations from above:

$$p_{jt} = \bar{p}_y + e_{jt} \quad (8)$$

$$\delta_{jt} = \beta X_{jt} - \alpha_1(\bar{p}_y) + \alpha_2(e_{jt}) + \epsilon_{jt} \quad (9)$$

this is directly estimable and we can easily derive with respect to average price:

$$\frac{\partial \delta_{jt}}{\partial \bar{p}_y} = -\alpha_1 \quad (10)$$

Then, we can derive the aggregate market share with respect to average price (where  $m_l$  is the market size of each market). I don't subscript all quantities by year for simplicity's sake.

$$Q = \sum_{l \in T} m_l \sum_{k \in J_l} s_{kl} \quad (11)$$

$$M_t = \sum_{k \in K_t} \exp(\delta_{kt}) \quad (12)$$

$$\frac{\partial M_t}{\partial \bar{p}} = -\alpha_1 M_t \quad (13)$$

$$\frac{\partial}{\partial \bar{p}} \left( \sum_{k \in J_t} \delta_{kt} \right) = -\alpha_1 N_t \quad (14)$$

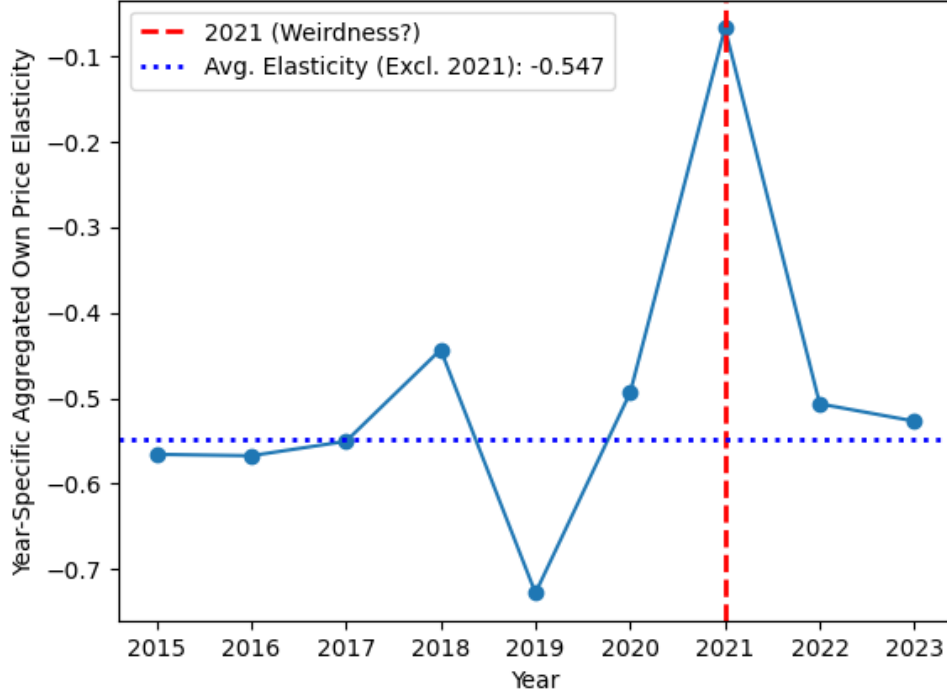
$$\frac{Q}{\partial \bar{p}_y} = -\alpha_1 \sum_{l \in T_y} m_l \left( \sum_{k \in J_t} s_{kl} \right) \left( \frac{1 + (\sum_{k \in J_t} \delta_{kt}) - N_t}{1 + (\sum_{k \in J_t} \delta_{kt})} \right) \quad (15)$$

Where  $N_t$  is the number of products in market  $t$  and  $T_y$  the set of all markets in year  $y$  (recall that markets are defined as zip-years). Then, the own-price elasticity can be written as the following:

$$\eta_{jjy} = \frac{Q}{\partial \bar{p}_y} \times \frac{\bar{p}_y}{Q} \quad (16)$$

this aggregate own-price elasticity, **the aggregate own-price elasticity with respect to the average statewide price of gasoline in California**, is directly estimable. Note that it is subscripted by year,  $y$ . Below I include a simple time-series plot of the aggregate own-price elasticity for California over time, for the years of our best data coverage 2015-2022:

Figure 3: Yearly Elasticity of Demand for Statewide Gasoline Consumption in California



## 4 Nested Logit

As discussed in the appendix, cross-price elasticities estimated by the pure logit model have a serious shortcoming: they only depend on one of the two products' price. We can counter this shortcoming, called the Independence of Irrelevant Alternatives (IIA) property, through a nested logit which has looser restrictions on the elasticity patterns. There are generally three approaches we can take to nested logit based on different nesting schemes:

1. **One nest:** 1) all stations
2. **Two nests:** 1) unbranded and 2) branded stations (including hypermarkets)
3. **Three nests:** 1) unbranded, 2) hypermarket, and 3) non-hypermarket branded stations

### 4.1 Estimation

Based on the derivation in the appendix, we can use the following linear estimating equation for nested logit (for nest  $h$ ):

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = x_{jt}\beta - \alpha p_{jt} + \rho \ln s_{j|h(j)t} + \xi_{jt} \quad (17)$$

where  $s_{j|h(j)}$  is the within-nest market share (i.e., market share conditional on nest-market) for product  $j$  in nest  $h$  in market  $t$ . Note that estimating this equation requires two different instruments, one for price (we'll continue using the gulf coast spot price for RBOB) and the other for the within-group market share, which is also correlated with unobserved idiosyncratic demand shocks.

Some ideas for instruments for within-nest market shares:



- Number of stations in a nest-market
- Sum of characteristics of other firms in market, i.e. a station's competitors

In the results tables, below, I report the coefficients on  $\ln(p_j)$  (referred to as 'log prices') and  $\ln s_{j|h(j)}$  (referred to as 'log conditional shares'); however, when deriving the elasticities, I use  $\alpha$  and  $\rho$  per the above linear equation.

Derived in the appendix, the elasticities for the nested logit take the following forms for 1) own-price elasticity ( $\eta_{jj}$ ), 2) cross-price elasticity between two products in the same nest ( $\eta_{jk}$ ), and 3) cross-price elasticity between three products in different nests ( $\eta_{jf}$ ).

$$\eta_{jjt} = \frac{\partial s_{jt}}{\partial p_{jt}} \frac{p_{jt}}{s_{jt}} = \frac{-\alpha \cdot p_{jt}}{1 - \rho} (1 - \rho s_{j|ht} - (1 - \rho) s_{jt}) \quad (18)$$

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \alpha \cdot p_{kt} s_{kt} (1 + \frac{\rho}{1 - \rho} s_{j|ht}) \quad (19)$$

$$\eta_{jft} = \frac{\partial s_{jt}}{\partial p_{ft}} \frac{p_{ft}}{s_{jt}} = \alpha \cdot p_{ft} s_{ft} \quad (20)$$

What I report in the tables below are the simple average of each of these elasticities across all products.

## 4.2 Results

Here are the results from Nested Logit, manually estimated using IV methods, with number of firms per nest-market as the instrument for conditional share in Panel A, and the first stage results for both RBOB and the conditional share in Panel B. Also reported are the mean elasticities.

Note that  $\eta_{jft}^-$  reports the elasticity of good  $j$  to the price of good  $f$  in market  $t$ , which *is not* contained within the nest of good  $j$ . On the other hand,  $\eta_{jkt}^-$  reports the elasticity of different goods in the same nest.

Table 6: Nested Logit Estimation Results

|   | (1)                  | (2)                  | (3)   |
|---|----------------------|----------------------|---|
| <b>Panel A: IV Estimation Results</b>               |                      |                      |   |
| prices  | -0.045***<br>(0.005) | -0.050***<br>(0.005) | -0.051***<br>(0.005)                        |
| nesting parameter ( $\rho$ )                        | 0.488***<br>(0.006)  | 0.454***<br>(0.006)  | 0.482***<br>(0.006)                         |
| brand chevron                                       | 0.195***<br>(0.017)  | 0.901***<br>(0.021)  | 0.876***<br>(0.021)                         |
| brand shell   | 0.159***<br>(0.019)  | 0.861***<br>(0.023)  | 0.832***<br>(0.022)                         |
| brand 76  | -0.037**<br>(0.019)  | 0.667***<br>(0.023)  | 0.663***<br>(0.022)                         |
| brand ampm  | 0.582***<br>(0.020)  | 1.306***<br>(0.022)  | 1.270***<br>(0.021)                         |
| brand 7 eleven                                      | -0.038*<br>(0.020)   | 0.659***<br>(0.024)  | 0.625***<br>(0.023)                         |
| hypermarket   | 1.750***<br>(0.035)  | 2.520***<br>(0.036)  | 2.289***<br>(0.038)                         |
| other brand   | 0.169***<br>(0.014)  | 0.871***<br>(0.019)  | 0.849***<br>(0.018)                         |
| amenity car wash                                    | -0.534***<br>(0.056) | -0.660***<br>(0.065) | -0.621***<br>(0.063)                        |
| amenity convenience store                           | 0.113***<br>(0.013)  | 0.135***<br>(0.014)  | 0.125***<br>(0.014)                         |
| amenity restaurant fast food outlet                 | 0.723**<br>(0.359)   | 0.800**<br>(0.352)   | 0.788**<br>(0.356)                          |
| amenity supermarket general store                   | 0.491***<br>(0.066)  | 0.669***<br>(0.077)  | 0.663***<br>(0.078)                         |
| amenity kiosk                                       | 0.168***<br>(0.021)  | 0.186***<br>(0.022)  | 0.161***<br>(0.022)                         |
| $\eta_{j t}^-$ (own-price elasticity)               | -0.290               | -0.294               | -0.312                                      |
| $\eta_{j kt}^-$ (intra-nest cross-price elasticity) | 0.004                | 0.005                | 0.004                                       |
| $\eta_{j ft}^-$ (inter-nest cross-price elasticity) | nan                  | 0.004                | 0.007                                       |
| N   | 42675                | 42675                | 42675                                       |
| <b>Panel B: First-Stage Results</b>                 |                      |                      |   |
| Number of Firms In Nest-Market                      | -0.173***<br>(0.001) | -0.216***<br>(0.001) | -0.212***<br>(0.001)                        |
| NESTS   | INSIDE               | BRANDED – UNBRANDED  | NON-HYPER BRANDED – HYPERMARKET – UNBRANDED |

<sup>a</sup> Significance levels: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

### 4.3 Aggregate Elasticities

We want to understand how elasticity and consumption patterns are different across nests. Let there be three nests: 1) branded, 2) unbranded, and 3) hypermarket gasoline stations.

Then we can write that the market share of unbranded stations in market  $t$  is defined by the following:

$$s_{ut} = \frac{D_{ut}^{(1-\rho)}}{D_{ut}^{(1-\rho)} + D_{bt}^{(1-\rho)} + D_{ht}^{(1-\rho)}} \quad (21)$$

$$D_{ut} = \sum_{k \in J_u} \exp\left(\frac{\delta_k}{1-\rho}\right) \quad (22)$$

where  $\rho$  is the previously-estimated nesting parameter  $J_u$  is the set of all unbranded goods in the market, and  $\delta_k$  (as before) is the mean utility of good  $k$ . We can find aggregate elasticities using a similar strategy that we did for the

pure logit estimation. Let:

$$p_{jt} = \bar{p}_{hy} + e_{it} \quad (23)$$

$$\delta_{jt} = \beta X_{jt} - \sum_{o \in H} [(d_o \alpha_{1,o})(\bar{p}_o)] - \alpha_2(e_{jt}) + \epsilon_{jt} \quad (24)$$

Where  $d_h$  is a dummy equal to 1 if good  $j$  is in nest  $b$ . **Note that, just like pure logit, average price varies by year and nest.** This means that we'll have different aggregate elasticities by nest and by year.  $H$  is the set of all nests. This gives us flexibility to estimate a different price coefficient for different segments. We're essentially interacting price and a 'characteristic' dummy. Thus, we can model a counterfactual where the average price of unbranded gas stations falls by \$1 and branded gas stations are the same. Or, perhaps more helpful, how demand would change if branded gas stations charged the same price for a gallon of gasoline that the nearest unbranded gas station does.

Note that each nest has its own average and thus its own  $\alpha_{1,h}$  parameter. First, let's differentiate the share of the unbranded nest with respect to the mean unbranded price ( $\bar{p}_h$ ). The derivative takes the following form:

$$\frac{\partial D_{ut}}{\partial \bar{p}_{uy}} = -\alpha_{1,u} D_u \quad (25)$$

$$\frac{\partial D_{bt}}{\partial \bar{p}_{uy}} = 0 \quad (26)$$

$$\frac{\partial D_{ht}}{\partial \bar{p}_{uy}} = 0 \quad (27)$$

$$\frac{\partial (D_{ut}^{(1-\rho)})}{\partial \bar{p}_{uy}} = (1-\rho)(D_{ut}^{(1-\rho)})(-\alpha_{1,u}) \quad (28)$$

$$\frac{\partial s_{ut}}{\partial \bar{p}_{uy}} = -\alpha_{1,u}(1-\rho)(D_u^{1-\rho}) \frac{D_b^{(1-\rho)} + D_h^{(1-\rho)}}{(D_u^{(1-\rho)} + D_b^{(1-\rho)} + D_h^{(1-\rho)})^2} = -\alpha_{1,u}(1-\rho)(s_{ut}) \frac{D_b^{(1-\rho)} + D_h^{(1-\rho)}}{D_u^{(1-\rho)} + D_b^{(1-\rho)} + D_h^{(1-\rho)}} \quad (29)$$

Then, we can derive the nest market share of unbranded with respect to the mean price of branded gasoline:

$$\frac{\partial D_{bt}}{\partial \bar{p}_{by}} = -\alpha_{1,b} D_{bt} \quad (30)$$

$$\frac{\partial D_{ut}}{\partial \bar{p}_{by}} = 0 \quad (31)$$

$$\frac{\partial (D_{bt}^{(1-\rho)})}{\partial \bar{p}_{by}} = (1-\rho)(D_{bt}^{(1-\rho)})(-\alpha_{1,b}) \quad (32)$$

$$\frac{\partial s_{ut}}{\partial \bar{p}_{by}} = \alpha_{1,b}(1-\rho)(D_b^{1-\rho}) \frac{D_u^{(1-\rho)}}{(D_u^{(1-\rho)} + D_b^{(1-\rho)} + D_h^{(1-\rho)})^2} = (\alpha_{1,b})(1-\rho)(s_{bt}) \frac{D_u^{(1-\rho)}}{D_u^{(1-\rho)} + D_b^{(1-\rho)} + D_h^{(1-\rho)}} \quad (33)$$

From there, we can infer the derivative with respect to the average price of hypermarket stations to be the following:

$$\frac{\partial s_{ut}}{\partial \bar{p}_{hy}} = \alpha_{1,h}(1-\rho)(D_h^{(1-\rho)}) \frac{D_u^{(1-\rho)}}{(D_u^{(1-\rho)} + D_b^{(1-\rho)} + D_h^{(1-\rho)})^2} = \alpha_{1,h}(1-\rho)s_{ht} \frac{D_u^{(1-\rho)}}{(D_u^{(1-\rho)} + D_b^{(1-\rho)} + D_h^{(1-\rho)})} \quad (34)$$

But, we want *aggregates*. Then, the aggregated share of unbranded gasoline sold in the state in year  $y$  is the following:

$$S_{uy} = \frac{1}{Q_y} \sum_{l \in T_y} m_l s_{ul} \quad (35)$$

where  $Q_y$  is the total quantity of gasoline sold in California in year  $y$  and  $T_y$  is the set of all markets associated with

year  $y$  (recall that markets are defined as zip-years). Then we can apply our previously-found derivative:

$$\frac{\partial S_{uy}}{\partial \bar{p}_{uy}} = \frac{1}{Q_y} \sum_{l \in T_y} m_l \frac{\partial s_{ul}}{\partial \bar{p}_{uy}} \quad (36)$$

$$\eta_{uuy} = \frac{\partial S_{uy}}{\partial \bar{p}_{uy}} \frac{\bar{p}_{uy}}{S_{uy}} = -\frac{\bar{p}_{uy} \alpha_{1,b} (1-\rho)}{Q_y} \sum_{l \in T_y} \frac{q_{ul} (D_b^{(1-\rho)} + D_h^{(1-\rho)})}{D_u^{(1-\rho)} + D_b^{(1-\rho)} + D_h^{(1-\rho)}} \quad (37)$$

$$\frac{\partial S_{uy}}{\partial \bar{p}_{by}} = \frac{1}{Q_y} \sum_{l \in T_y} m_l \frac{\partial s_{ul}}{\partial \bar{p}_{by}} \quad (38)$$

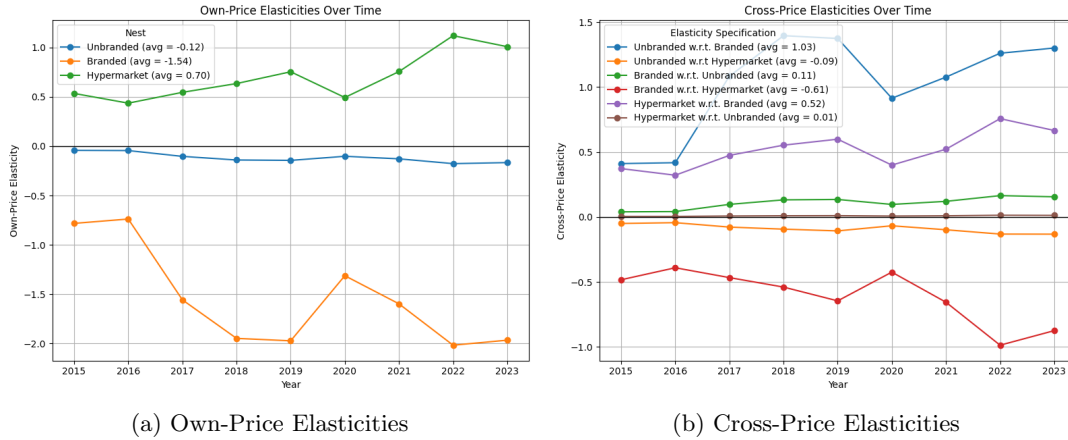
$$\eta_{uby} = \frac{\partial S_{uy}}{\partial \bar{p}_{by}} \frac{\bar{p}_{by}}{S_{uy}} = \frac{\bar{p}_{by} \alpha_{1,b} (1-\rho)}{Q_y} \sum_{l \in T_y} \frac{q_{bl} D_{ul}^{(1-\rho)}}{D_{ul}^{(1-\rho)} + D_{bl}^{(1-\rho)} + D_{hl}^{(1-\rho)}} \quad (39)$$

$$\frac{\partial S_{uy}}{\partial \bar{p}_{hy}} = \frac{1}{Q_y} \sum_{l \in T_y} m_l \frac{\partial s_{ul}}{\partial \bar{p}_{hy}} \quad (40)$$

$$\eta_{uhy} = \frac{\partial S_{uy}}{\partial \bar{p}_{hy}} \frac{\bar{p}_{hy}}{S_{uy}} = \frac{\bar{p}_{hy} \alpha_{1,h} (1-\rho)}{Q_y} \sum_{l \in T_y} \frac{q_{hl} D_{ul}^{(1-\rho)}}{D_{ul}^{(1-\rho)} + D_{bl}^{(1-\rho)} + D_{hl}^{(1-\rho)}} \quad (41)$$

Where  $q_{ul}$  refers to the total quantity of unbranded gasoline sold in market  $l$  (correspondingly for branded and hypermarket gasoline). Now we can compute each of these elasticities and plot them by year. To compare apples-to-apples, below I've compared the estimates of own-price elasticities over time across the three nests, and in another graph I've compared the different cross-price elasticities. **Note that the sign is flipped for hypermarkets; I'm not sure why this is showing up in the data, but essentially the  $\alpha_{1,h}$  coefficient is positive, suggesting higher average hypermarket prices in a given market increase the likelihood of consuming hypermarket gasoline.**

Figure 4: Yearly Aggregated Elasticities of Nest Statewide Market Shares



## 5 Appendix

### 5.1 ‘Cheat Sheet’ of Market Share and Elasticity Equations

#### 5.1.1 Pure Logit Model

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = x_{jt} - \alpha p_{jt} + \xi_{jt} \quad (42)$$

$$\eta_{jzt} = \alpha \cdot p_{jt} \cdot (1 - s_{jt}) \quad (43)$$

$$\eta_{jkt} = -\alpha \cdot p_{kt} \cdot s_{kt} \quad (44)$$

### 5.1.2 Nested Logit Model

For nested logit, we have three elasticities: 1) own-price elasticity ( $\eta_{jkt}$ ), 2) cross-price elasticity between two different products in the same nest ( $\eta_{jkt}$ ), and 3) cross-price elasticity between two products in different nests ( $\eta_{jft}$ ).

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = x_{jt}\beta - \alpha p_{jt} + \rho \ln s_{j|h(j)t} + \xi_{jt} \quad (45)$$

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{jt}} \frac{p_{jt}}{s_{jt}} = \frac{-\alpha \cdot p_{jt}}{1 - \rho} (1 - \rho s_{j|ht} - (1 - \rho)s_{jt}) \quad (46)$$

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \alpha \cdot p_{kt} s_{kt} (1 + \frac{\rho}{1 - \rho} s_{j|ht}) \quad (47)$$

$$\eta_{jft} = \frac{\partial s_{jt}}{\partial p_{ft}} \frac{p_{ft}}{s_{jt}} = \alpha \cdot p_{ft} s_{ft} \quad (48)$$

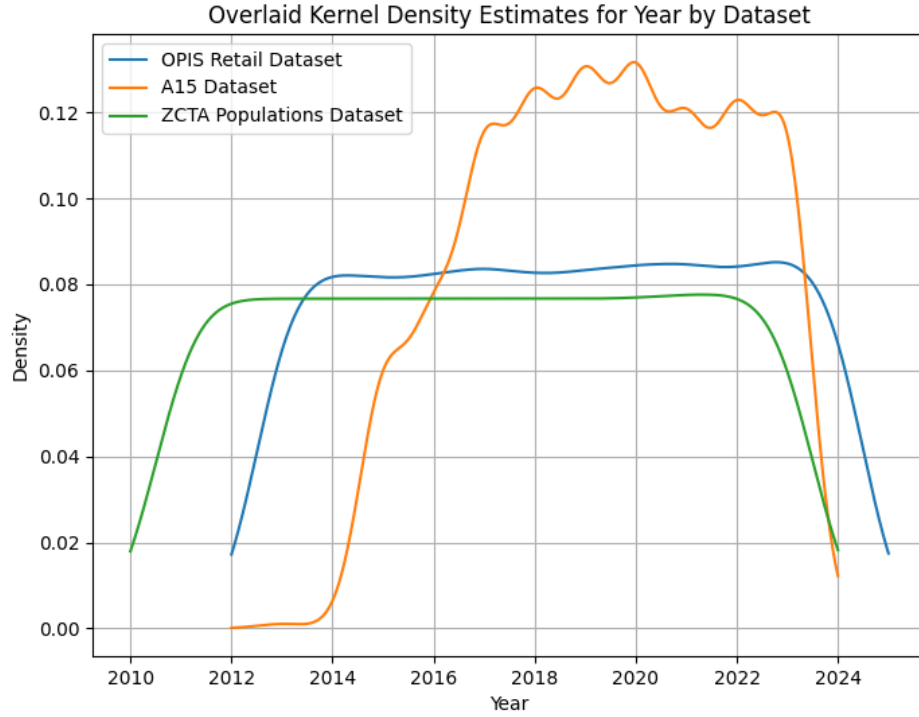
## 5.2 Data Build Notes

Building the data required different sources of data and upon merging these sources of data, some observations were lost. To ensure this didn't introduce bias into the data, I've compiled the following table comparing summary statistics for different stages of the data build:

Table 7: Summary Statistics Across Steps of Data Build

|   | OPIS Retail Dataset | A15 Data | Present in Both OPIS and A15 Data | $s_0 > 0$ | $0 < s_j < 1$ |
|---|---------------------|----------|-----------------------------------|-----------|---------------|
| Number of Observations                    | 117955              | 49041    | 43381                             | 42816     | 42675         |
| Observations as % of OPIS Retail Data     | 100.0               | 41.6     | 36.8                              | 36.3      | 36.2          |
| Number of Unique Stations                 | 15835               | 10813    | 9859                              | 9775      | 9761          |
| Unique Stations as % of OPIS Retail Data  | 100.0               | 68.3     | 62.3                              | 61.7      | 61.6          |
| % of Observations From the Bay Area       | 17.3                | 16.8     | 0.0                               | 0.0       | 0.0           |
| % of Observations From Los Angeles County | 20.7                | 20.5     | 0.0                               | 0.0       | 0.0           |
| % of Observations From Unbranded Stations | NA                  | NA       | NA                                | NA        | NA            |
| Average Retail Price                      | 3.87                | NA       | 3.83                              | 3.83      | 3.83          |
| Average Quantity Sold                     | NA                  | 1144061  | 1159662                           | 1157252   | 1161076       |
| Average Year                              | 2018.5              | 2019.4   | 2019.4                            | 2019.4    | 2019.4        |

Our different sources of data have differing intervals of time covered, as the kernel density plot estimate below illustrates:



Thus, I can recreate the summary statistics table, *first restricting* the OPIS and A15 data by date (to 2015-2023):

Table 8: Summary Statistics Across Steps of Data Build

|   | OPIS Retail Dataset | A15 Data | Present in Both OPIS and A15 Data | $s_0 > 0$ | $0 < s_j < 1$ |
|---|---------------------|----------|-----------------------------------|-----------|---------------|
| Number of Observations                    | 88792               | 49041    | 43323                             | 42760     | 42619         |
| Observations as % of OPIS Retail Data     | 100.0               | 55.2     | 48.8                              | 48.2      | 48.0          |
| Number of Unique Stations                 | 14504               | 10813    | 9841                              | 9756      | 9742          |
| Unique Stations as % of OPIS Retail Data  | 100.0               | 74.6     | 67.9                              | 67.3      | 67.2          |
| % of Observations From the Bay Area       | 17.3                | 16.8     | 0.0                               | 0.0       | 0.0           |
| % of Observations From Los Angeles County | 20.7                | 20.5     | 0.0                               | 0.0       | 0.0           |
| % of Observations From Unbranded Stations | NA                  | NA       | NA                                | NA        | NA            |
| Average Retail Price                      | 3.78                | NA       | 3.83                              | 3.83      | 3.83          |
| Average Quantity Sold                     | NA                  | 1144061  | 1160258                           | 1157865   | 1161696       |
| Average Year                              | 2019.0              | 2019.4   | 2019.4                            | 2019.4    | 2019.4        |

### 5.3 State-Wide Geographic Visualizations

Figure 5: Heatmaps of Volume-Weighted Average Retail Prices on ZIP Level

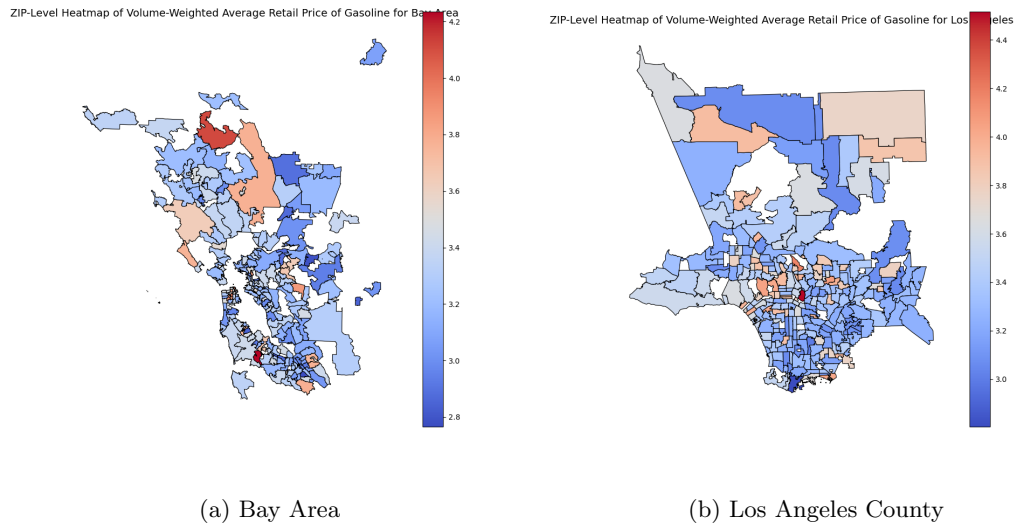


Figure 6: Heatmaps of Uniquely Identified Gasoline Stations on ZIP Level

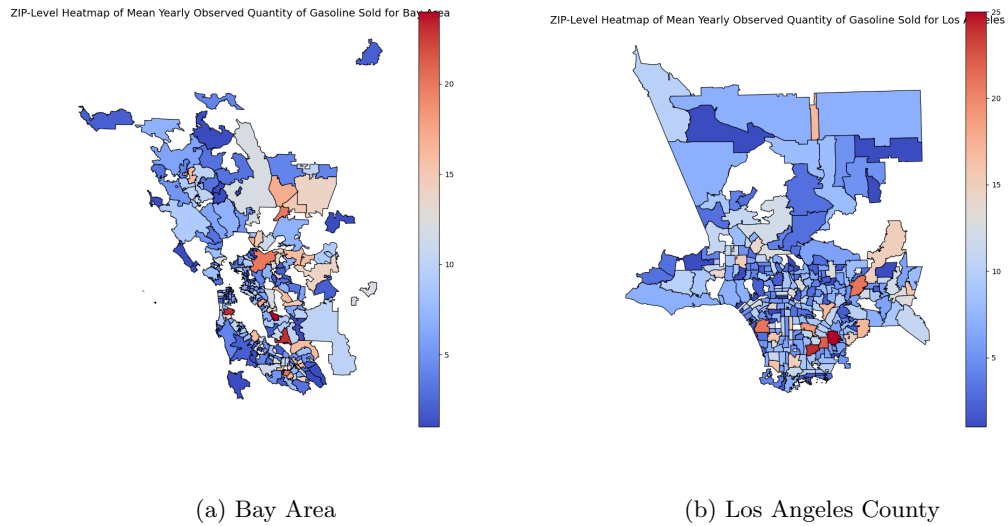
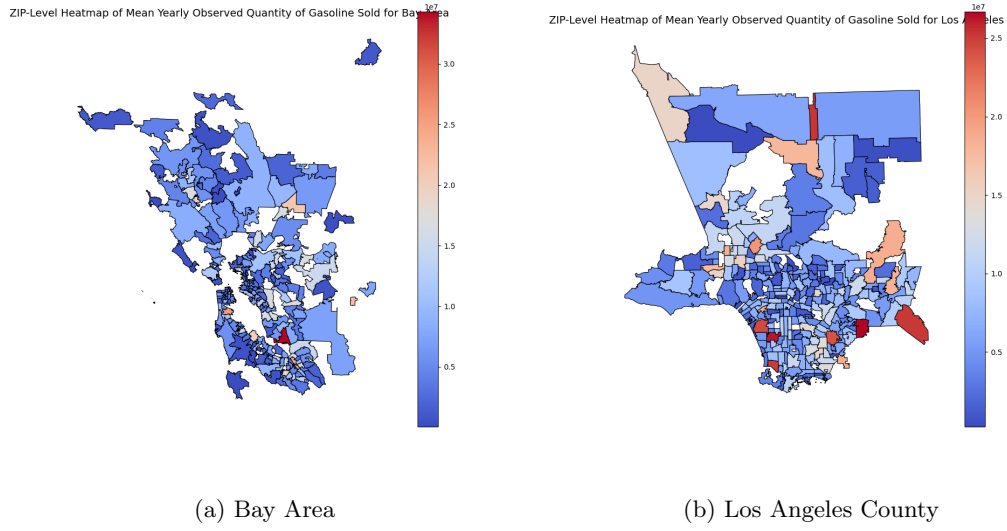


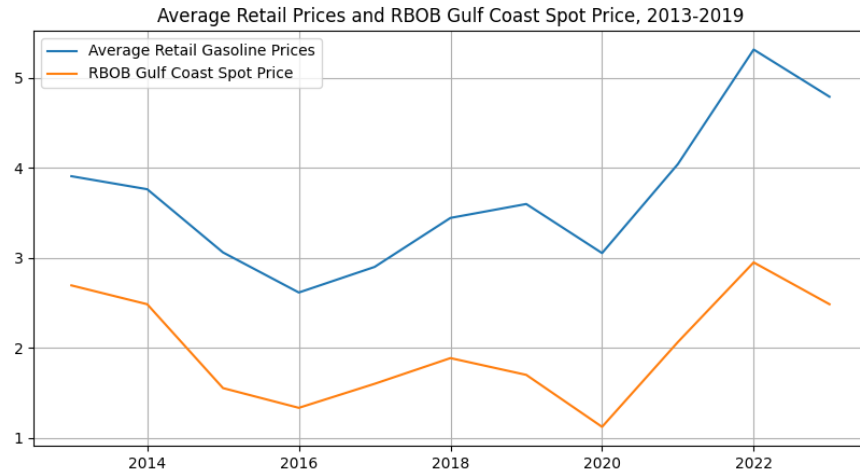
Figure 7: Heatmaps of Average Annual Observed Quantity Sold of Gasoline on ZIP Level



## 5.4 Instruments Descriptives

### 5.4.1 RBOB Gulf Coast

Here is a time-series, with the volume-weighted average gasoline prices in the sample plotted alongside the RBOB Gulf Coast Spot Price (all nominal) over time:

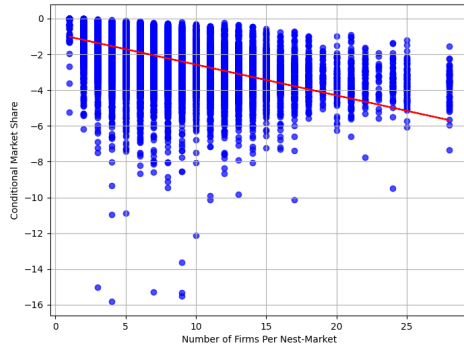


### 5.4.2 Number of Firms in Each Nest-Market

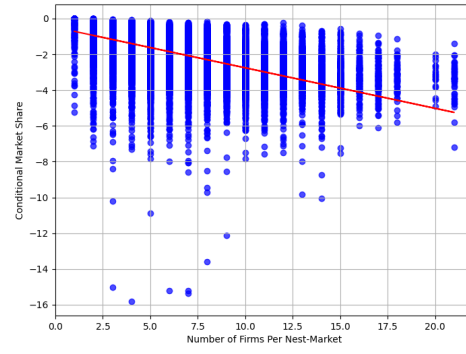
Below, I've included a scatterplot showing the relationship between the number of products in each nest-market (by nest designation) and log conditional market share. This provides suggestive evidence that there is a positive correlation here, suggesting the number of products in nest-market has some strength as an instrument.



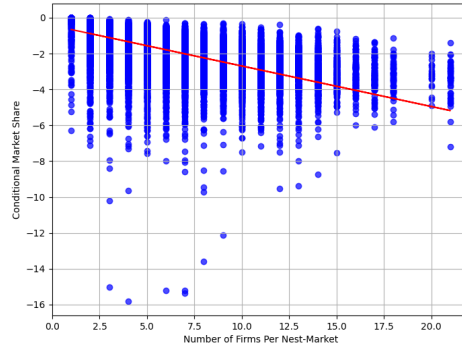
Figure 8: Scatterplots of Number of Products In Nest-Market and Market Share Conditional on Nest-Market



(a) Inside Good Nest



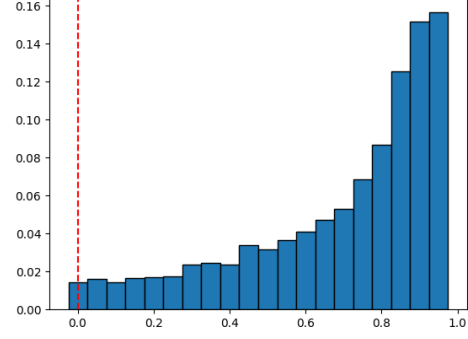
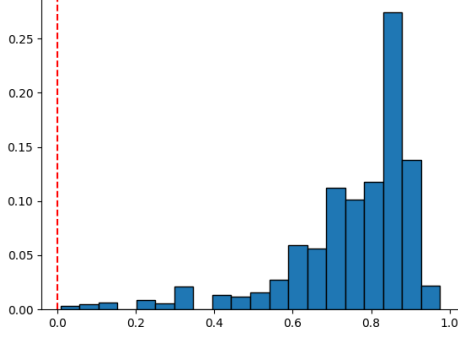
(b) Branded – Unbranded Nests



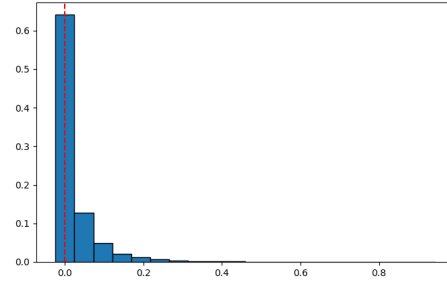
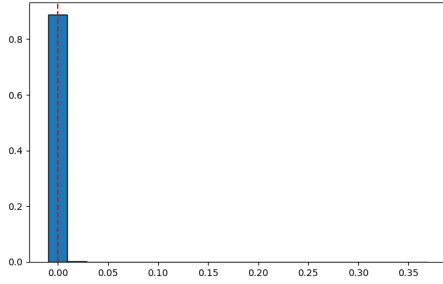
(c) Non-Hypermarket Branded – Hypermarket – Unbranded Nests

## 5.5 Market Shares Descriptives

After dropping instances where the resulting outside good market share ( $s_{0t}$ ) was negative (this amounted to ), we have the following distributions for the outside good and inside good shares:



(a) Zip3-Level Market Outside Good Shares ( $s_{0t}$ ) (b) Zip-Level Markets Outside Good Shares ( $s_{0t}$ )



(c) Zip3-Level Market Inside Good Shares ( $s_{jt}$ ) (d) Zip-Level Markets Inside Good Shares ( $s_{jt}$ )

Although very large, these outside shares appear reasonable and, most importantly, incorporate the significant heterogeneity in California in population density.

## 5.6 Elasticities Check

Here, we try to contextualize some of the elasticity estimates from above by running some preliminary checks on elasticity.

First, I estimate the following equation:

$$\ln(q_t) = \beta_1 \ln(p_t) + \epsilon_t \quad (49)$$

where  $q_t$  is the total observed quantity of gasoline sold in zip-year  $t$  and  $p_t$  is the market share-weighted average price of gasoline sold in zip-year  $t$ . This is very similar to the model we ran for simple logit without product differentiation, with a different dependent variable (here simply  $\ln(q)$  rather than  $\ln(\frac{s_{jt}}{s_{0t}})$ ). In one specification I include the RBOB Gulf Coast Spot Price as the instrument for price, while in the other I don't instrument for price. The results of those estimations can be found here:

Table 9: Market-Level Demand Estimation

|                     | OLS                  | IV                 |
|---------------------|----------------------|--------------------|
| $\ln(\text{price})$ | -0.792***<br>(0.055) | -0.104*<br>(0.063) |
| N                   | 9989                 | 9989               |
| RBOB Instrument     | NO                   | YES                |

<sup>a</sup> Significance levels: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Then, we can visualize the relationship between price and quantity on an aggregated, statewide level (prices volume-weighted average across the whole state and observed quantities sold summed) in the following graphics:

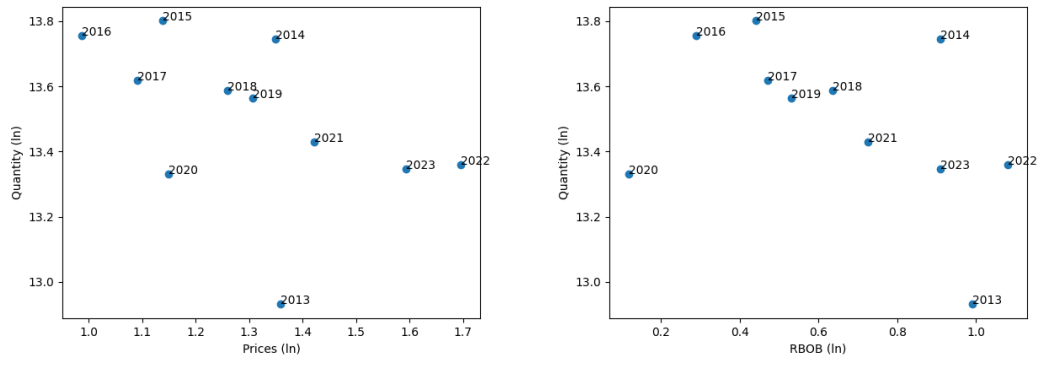


Figure 10: State-Wide Relationships Between Price/RBOB Gulf Coast and Quantity