Technical appendix to:

Estimating Demand Heterogeneity Using Aggregated Data: An Application to the Frozen Pizza Category

Paulo Albuquerque and Bart J. Bronnenberg

This appendix serves as a guide to researchers interested in implementing the approach proposed by Albuquerque and Bronnenberg (2008) as described in the paper "Estimating Demand Heterogeneity Using Aggregated Data: An Application to the Frozen Pizza Category." The data and parameters are described in section A. Section B describes the algorithm used to obtain the moment conditions and the optimization function.

A Preliminaries

A.1 Data

There are $j = 1, ..., J$, brands in the market, and we observe $t = 1, ..., T$ weekly observations. Since penetration per brand and purchase set sizes are usually annual measures, then $T \geq 52$. The estimation requires the following data:

1. Market Shares: $\widetilde{s}$, a matrix containing market shares over time, of size $[J \times T]$. These shares add to 1 in each time period and are for inside goods only (i.e., the outside good alternative is not considered here).

2. Marketing mix variables: X, a matrix of size $[(T \cdot J) \times K]$, that stacks marketing mix variables by time and brand. We observe K marketing mix variables.

3. Dummy variables: D, a matrix of size $[(T \cdot J) \times J]$. Each column j of this matrix is a dummy vector, stacked as X, that takes the value of 1 for rows of brand j, and zero otherwise. If one of the brands is not present for some time periods, as in the case of a new product introduction, the number of rows for that brand is limited to the total number of periods when it is present in the market.

4. Empirical distribution of purchase set size: $\mathcal{F}$, vector of $[(S+1) \times 1]$, $S < J$, containing the percentage of households who buy $s = [0, 1, 2, ..., S]$ unique brands from the category in a year. The strict inequality $S < J$ applies because $\Pr(S = J)$ is equal to $1 - \sum_{s=1}^{J-1} \Pr(S = s)$.

5. Penetration by brand: π, vector $[J \times 1]$ of annual penetration for each individual brand.

6. Total sales for the category under study, measured as category expenditure in the store: CE= $[\text{CE}_1 \ldots \text{CE}_t \ldots \text{CE}_T]$, vector of size $[1 \times T]$. We divide CE_t by its mean, so that the vector has mean equal to 1. Although this normalization is not necessary, doing so allows us to interpret the value of parameter λ as the relative weight of the inside goods in the total market.

7. Instrumental variables matrix: Z, of size $[(J \cdot T) \times R]$, where R is the number of instruments. The paper gives details on the instruments used in the simulation and in the empirical study.

8. Random draws from the standard Normal distribution collected in a matrix $[N \times (P + K)]$, where N is the total number of draws (i.e., simulated individuals), H is the number of dimensions in the product perception map (we use $H = 2$) and K is the number of marketing mix variables. We note that in our empirical work, we only compute random effects on price.

A.2 Parameters

The number of parameters to be estimated is $3 \times J + 2 \times K$. The parameters are:

1. J brand intercepts, α.

2. $2 \times J - 1$ locations of the brands, L, for the two-dimensional factor model covariance structure. Several identification restrictions need to be imposed on the factor model: the location of one brand in one dimension is zero and the first dimension of one brand and the second dimension of another brand are restricted to be positive.[1] We also note that without any loss in generality, the interested reader can estimate the locations of the brands in polar coordinates as we do. The paper explains that such coordinates can be interpreted directly in terms of variances of the random effects, and the correlation of random effects in the consumer population. However, since there is a one to one mapping between Cartesian and Polar coordinates, either way of parameterizing the attribute addresses leads to the same model implications.

3. K marketing-mix mean parameters, β;

4. K marketing-mix variance parameters, for unobserved heterogeneity, σ;[2]

[1] For a simplified model with a diagonal variance-covariance matrix, the number of parameters for the variance of brand preference heterogeneity will instead be J.

[2] In the paper, we only compute random effects for the price variable. Thus, in our implementation, we only added one extra parameter, not K.

5. 1 parameter for the scale of the inside goods, λ. This parameter measures the average relative size of the inside goods, when compared to the total market that includes both inside and outside alternatives.

B Estimation steps

In estimation, we minimize a set of moment conditions. In each iteration of this minimization, we need to compute the sample moment values. The discussion below is at points specific to our paper and model specifications. A more general formulation of the standard BLP algorithm is contained in Aviv Nevo (2000, *Journal of Economics & Management Strategy*). Given a set of parameters $\theta = [\alpha, \beta, L, \sigma, \lambda]$, and a weight matrix W which is initialized using a set of preliminary estimates, we obtain the sample value of the moment conditions using the following steps:

1. *Compute market shares that consider the outside good as an alternative*

   To do this, we use the observed data about inside good shares $\widetilde{s}$, and the parameter λ. Start by computing the row-vector of the estimated outside good share by subtracting a multiple of the category volume from 1, $\widehat{s}_0 = 1 - \lambda \times \text{CE}$. Then obtain the market shares with the outside good in market $s_j = \widetilde{s}_j \odot (1 - \widehat{s}_0)$, where $\odot$ stands for element-by-element multiplication. The matrix s, containing shares for all inside brands, will be of dimension $[J \times T]$. Note that, for each time period, these shares add to $1 - \widehat{s}_0$, since a row containing the share of the outside good is not included in the matrix.

2. *Given the outside good, recover the mean utilities.*

   This inversion from shares to mean utilities is done through the familiar contraction mapping procedure. Collect all utility terms that are individual specific in the vector $\mu_{ijt} = L_j \omega_i + x_{jt} \beta_i$, and collect the utility terms that are common to individuals in the variable $\delta_{jt} = \alpha_j + \xi_{jt}$. To compute δ_{jt}, we need to use an iterative procedure. Take any initial value δ_{jt}^0 and compute utilities $u_{ijt}^0 = \mu_{ijt} + \delta_{jt}^0 + \epsilon_{ijt}$ for each brand and for all time periods. Compute choice probabilities using the logit formulation and obtain the predicted share matrix $\widehat{s}\left(\delta^0, \theta\right)$, by averaging the choice probabilities over all "simulated" individuals.[3] Compute $\delta_{jt}^1 = \delta_{jt}^0 + \ln(s_{jt}) - \ln\left(\widehat{s}_{jt}\left(\delta_{jt}^0, \theta\right)\right)$ and the max $\left(\delta_{jt}^1 - \delta_{jt}^0\right)$, over $\forall j, t$. Repeat these steps while max $\left(\delta_{jt}^{n+1} - \delta_{jt}^n\right) > \varsigma$, where ς is a small quantity. In our case, we used $10E - 10$, although we have also used $10E - 11$ on occasion.

[3] $\widehat{s}\left(\delta^0, \theta\right)$ is of dimension $[J \times T]$. Each element $\widehat{s}_{jt}\left(\delta_{jt}^0, \theta\right)$ is the estimated share of brand j at time t among all choices, including the outside good.

3. *Obtaining linear parameters α_j and unobservables ξ_{jt}*

   We regress the values of δ_{jt} obtained in the last iteration of the contraction mapping on the data and instruments to obtain the parameters of the mean utilities (in our case that is only α_j but if the marketing mix variables are specified linearly, this would also involve the mean marketing mix effects) and the ξ_{jt}, using the instrumental variable approach: the mean parameters α_j are obtained from $\hat{\alpha} = \left([D\ X]'Z \cdot (Z'Z)^{-1} \cdot Z'[D\ X]\right)^{-1} \cdot ([D\ X]'Z \cdot (Z'Z)^{-1} \cdot Z'\delta)$, where δ is a vector $[(J \cdot T) \times 1]$, with all the values of $\hat{\delta}_{jt}$ stacked by brand. Finally, we obtain $\hat{\xi} = \hat{\delta} - [D\ X] \cdot \hat{\alpha}$.

4. *Evaluating the BLP moments*

   Once $\hat{\xi}$ are obtained, the BLP moments to be optimized are given by the expected value of the product of ξ with each column of the instrument matrix Z:

$$G_1(\theta) : \begin{bmatrix} \left(\frac{1}{J \cdot T} \sum_{jt} \hat{\xi}_{jt} \cdot Z_{jt,1}\right) \\ \ldots \\ \left(\frac{1}{J \cdot T} \sum_{jt} \hat{\xi}_{jt} \cdot Z_{jt,R}\right) \end{bmatrix} \tag{1}$$

   where $G_1(\theta)$ is a vector of size $[R \times 1]$.

5. *Evaluating the outside good moments*

   The penetration by brand is obtained by making use of the choice probabilities $\Pr_{ijt}(X_t, \xi_t, \theta)$, in equation (4) of the paper, given a set of parameters θ, the data X_t, and the estimated unobservables $\hat{\xi}$ obtained in the previous step. Since the penetration data is an annual quantity, we use the last 52 weeks, from $T - 51$ to T, to compute the penetration that is forecast, or implied, by the model. As explained in the paper, the probability of a consumer only buying the outside good in all those weeks is given by:

$$\Pr_i(\{\varnothing\}) = \prod_{t=T-51}^{T} \Pr_{i0t}(X_t, \xi_t, \theta) \tag{2}$$

   The probability of buying brand j at least once is obtained from:

$$\Pr_i(j \in \{C_i\}) = 1 - \prod_{t=T-51}^{T} (1 - \Pr_{ijt}(X_t, \xi_t, \theta)) \tag{3}$$

   where $\{C_i\}$ is the purchase set of individual j, i.e., the set of different brands bought by that individual during the 52 weeks. The same expression can be use for all other alternatives.

The sample value of the outside good moment conditions is approximated to an arbitrary degree of precision by taking the mean of these probabilities across the simulated households.

$$G_2(\theta): \begin{bmatrix} \frac{1}{N}\sum_i \Pr_i(\{\varnothing\}) \\ \frac{1}{N}\sum_i \Pr_i(1 \in \{\cdot\}) \\ \vdots \\ \frac{1}{N}\sum_i \Pr_i(J \in \{\cdot\}) \end{bmatrix} - \begin{bmatrix} 1-\pi_c \\ \pi_1 \\ \vdots \\ \pi_J \end{bmatrix}$$

We have experimented with different values for N. We have eventually settled on using $N = 500$ in the data experiment. Results seem to suggest that this value works well enough to recover the data generating parameters. To be conservative, we set $N = 1000$ in the empirical analysis.

In the empirical analysis we have 5 brands. For those data, $G_2(\theta)$ is a vector of $[6 \times 1]$, and the first value of this vector applies to the category penetration, which is equal to the fraction of individuals who buy 0 unique brands of pizza (see next point).

6. *Obtaining the heterogeneity moments*

The purchase set size distribution is obtained recursively as explained in Section 3 of the paper, also aggregating choice probabilities of each individual over time. First, we need the estimated percentage of individuals that did not buy from the category for one year, the purchase set size of zero, which is equal to $\Pr_i(\{\varnothing\})$, obtained in the previous step (since we have already included it above, we drop it here).

Next, we compute the estimated percentage of individuals that bought only one brand (or the outside good) during the entire year. For each brand j, add the choice probabilities of brand j, $\Pr_{ijt}(X_t, \xi_t, \theta)$, and the outside good $\Pr_{i0t}(X_t, \xi_t, \theta)$, and for time periods from $T-51$ to T, compute the following product $H_{ij} = \prod_{t=T-51}^{T} [\Pr_{ijt}(X_t, \xi_t, \theta) + \Pr_{i0t}(X_t, \xi_t, \theta)]$ .H_{ij} includes the probabilities that consumers buy only brand j, only the outside good, or both. However, to obtain the purchase set size of 1, we do not want to include the case where consumers only buy the outside good. Thus:

$$\Pr_i(\{j\}) = H_{ij} - \Pr_i(\{\varnothing\}) \quad (4)$$

The probability that individual i has a purchase set size of zero is the accumulation of the previous quantity across $\forall j$:

$$\Pr(S_i = 1) = \sum_{j=1}^{J} \Pr_i(\{j\})$$

Similarly, we get the estimated percentage of consumers with purchase set size of $2, 3, ..., S$ as described in the paper. The sample moment is evaluated as:

$$G_3(\theta) : \frac{1}{N} \Pr(S_i = s) - \mathcal{F}_s$$

for $s = \{1, ..., 4\}$. Thus, in the empirical example $G_3(\theta)$ has size $[4 \times 1]$.

7. *Weight Matrix*

Before we can compute the objective function that combines all moment conditions, we need two weight matrices, W_1, for the BLP moments, and W_2, for the penetration and heterogeneity moments. Using a set of consistent estimates for $\tilde{\theta}$, we compute the weight matrices using the following steps.

(a) Weight matrix for BLP moments: Denote the following matrix by Ψ:

$$\Psi_{JT \times R} = \begin{bmatrix} \xi_1 \cdot Z_{1,1} & ... & \xi_1 \cdot Z_{jt,R} \\ ... & ... & ... \\ \xi_{jt} \cdot Z_{jt,1} & .. & \xi_{jt} \cdot Z_{jt,R} \end{bmatrix}. \tag{5}$$

and compute the "square" of the weighting matrix used in the optimization as:

$$W_1\left(\widetilde{\theta}\right)' W_1\left(\widetilde{\theta}\right) = \left(\frac{\Psi'\Psi}{(JT)^2}\right)^{-1} \tag{6}$$

(b) Weight matrix for penetration and heterogeneity moments: To obtain the variance of the penetration and heterogeneity moment, we draw different realizations of the demand shocks ξ, say $\tilde{\xi}$ and evaluate $G_2\left(\theta|\tilde{\xi}\right)$ and $G_3\left(\theta|\tilde{\xi}\right)$ using the procedure above. By replicating this procedure a number of times we obtain a sample of sample moment values, from which the variance covariance of $G_2\left(\theta|\tilde{\xi}\right)$ and $G_3\left(\theta|\tilde{\xi}\right)$ can be readily computed. The "square" of the weighting matrix $W_2\left(\widetilde{\theta}\right)' W_2\left(\widetilde{\theta}\right)$ is the inverse of this variance covariance matrix.

8. *Objective Function*

Gather the moment expressions and the two weighting matrices in the following vector $\mathbf{G}(\theta)$, of size $[(R + 10) \times 1]$, and matrix $\mathbf{W}$, of dimension $[(R + 10) \times (R + 10)]$:

$$\mathbf{G}(\theta) = \begin{bmatrix} G_1(\theta) \\ G_2(\theta) \\ G_3(\theta) \end{bmatrix}$$

and

$$\mathbf{W} = \begin{bmatrix} W_1\left(\widetilde{\theta}\right)' W_1\left(\widetilde{\theta}\right) & \mathbf{0} \\ \mathbf{0} & W_2\left(\widetilde{\theta}\right)' W_2\left(\widetilde{\theta}\right) \end{bmatrix} \tag{7}$$

Finally, the parameters are obtained when the following expression is minimized:

$$\mathbf{G}(\theta)' \cdot \mathbf{W} \cdot \mathbf{G}(\theta) \tag{8}$$

Steps 7-9 need to be carried out twice. In the first execution, we find consistent estimates for θ. These estimates and a first computation of $\mathbf{W}$ are initialized using a preliminary estimate for θ that fits the demand model to the share and the purchase set size and penetration data, using 25 different sets of initial parameter values. This preliminary estimate is used to initialize the first stage and serves the purpose of finding reasonable starting values for the heterogeneity parameters. In the second execution, we use the parameters that result from the first stage to recompute $\mathbf{W}$, as is done standard in the literature.

9. *Standard Deviation of Parameters*

To compute the standard deviation of the parameters, we use the final estimate of $\hat{\theta}$ to again evaluate the weight matrix $\mathbf{W}$, at $\hat{\theta}$. Call this matrix $\mathbf{A}$. We also numerically compute the Jacobian matrix containing the derivative of the $R+10$ moments with respect to the number of parameters (which was $3 \times J + 2 \times K$ from section A2, but more in general can be written as N_p).

$$\Gamma_{(R+10)\times N_p} = \left[\frac{\partial \mathbf{G}\left(\hat{\theta}\right)}{\partial \hat{\theta}'}\right]. \tag{9}$$

Next, the variance covariance matrix of the inference errors is equal to

$$\text{var}\left(\hat{\theta}\right) = \left(\Gamma'\mathbf{W}\Gamma\right)^{-1} \Gamma'\mathbf{W}\mathbf{A}^{-1}\mathbf{W}\Gamma\left(\Gamma'\mathbf{W}\Gamma\right)^{-1}$$