Problem Set 1: Demand Estimation

Ryan Mansley Nate Miller

Spring 2023

The objective is to estimate a model with random coefficients logit demand and differentiated-price Bertrand/Nash competition, along the lines of BLP (1995) and Nevo (2001). Please submit your problem set via email. Solutions should include (1) your raw code and (2) a PDF file with written answers to each question, formatted Latex tables as appropriate, and an appendix with documented code.

Collaboration among groups of 2-3 students is encouraged, though I will leave it up to you to determine the optimal group size. Each group should submit a single set of solutions. Groups also should feel free to communicate with each other or with me regarding any difficult parts of the problem set. If you are stuck on a question then move on, as the exercises are designed for you to learn the material, and once you are no longer learning for a question, you should stop spending time on it. That said, you should plan to spend a substantial amount of time on these exercises.

A useful resource is Aviv Nevo's Matlab code from his "Practitioners' Guide" article (2000 JEMS), which is posted with this problem set. Please don't use Matlab for your own code, however. Instead use R, Python, or Julia. Another useful resource is the Berry and Haile Handbook of IO chapter ("Foundations of Demand Estimation").

Data Description

In the data, 4 firms sell a total of 20 products in 20 markets. Only a subset of products are sold in each market, and the number of active firms varies across markets. The dataset contains product-market level observations for the following variables: a market identifier, a product identifier, a firm identifier, price, market share, an observable characteristic of the product (xvar), and a cost shifter (wvar). The sum of the market shares is less than one, reflecting the presence of an unobserved outside good. Thus, the data already reflect some "market size" that converts quantities into shares.

1. Prepare a table with summary statistics for market share, characteristics, price, cost shifters, the number of products per market, and the number of firms per market.

Logit Model Warmup (Use Stata/R/Python/Julia)

Suppose consumers have the following indirect utility function, where i denotes the consumer, j denotes the product, and t denotes the market:

$$u_{ijt} = x'_{jt}\beta + \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

where ξ_{jt} is unobserved quality, ϵ_{ij} is an IID draw from a type-1 extreme value (Gumbel) distribution, and the vector x_j includes both the observable characteristic (xvar) and a constant. Assume that firms observe ξ_j and then set prices in accordance with the Nash-Bertrand equilibrium. Also assume that the x_j variables are exogenous, in the sense of being orthogonal to unobserved quality. Marginal costs are specific to the product-market combination, and constant in market shares.

2. Derive an expression in which log shares are a linear function of data and parameters.

3. Estimate the parameters of this model (α, β) using the cost shifter provided in the data as an instrument. Report your point estimates and standard errors.

4. Why do you need an instrument? What are other examples of IVs you could use, and what conditions must they satisfy to be valid?

5. Compute a matrix of own and cross-price elasticities for market 1.

6. Construct a "Hausman instrument" based on the average price of a product in other markets, and use it to estimate the model. Under which conditions would the Hausman instrument be valid? Explain.

Random Coefficients Logit Model (Use R/Python/Julia)

Suppose now that there is heterogeneity among consumers in the utility they derive from x_{jt}, i.e., from xvar and the constant. The model is:

$$u_{ijt} = \underbrace{x'_{jt}\beta + \alpha p_{jt} + \xi_{jt}}_{\delta_{jt}(x_{jt},p_{jt},\xi_{jt};\theta_1)} + \underbrace{\sum_{k\in\{1,2\}} \pi_k d_{ikt} x_{jkt}}_{\mu_{ijt}(x_{jt},p_{jt},D_{it};\theta_2)} + \epsilon_{ijt}$$

$d_{ikt} \sim N(0,1)\ \forall k$. Create simulated consumers by generating 500 random draws. Because each market has the same distribution of demographics (by assumption), in this application the same draws can be used for all markets. Hold the draws fixed throughout estimation and simulation. Estimate using one-step GMM. For instruments Z, use the weighting matrix $(Z'Z)^{-1}$ in the objective function. Most applications use the efficient two-step GMM estimator, but we won't layer that in here.

7. Write the parameter vector of interest as $\theta = (\theta_1, \theta_2)$ where θ_1 are the linear parameters and θ_2 are the nonlinear parameters. Which parameters are in θ_1 and which are in θ_2? Explain why computation time increases in the length of θ_2 but not θ_1.

8. Explain what variation in the data will allow you to identify each of the parameters. What is the minimum number of instruments you will need to for estimation? Use all the instruments for both the GMM problem and the nested 2SLS estimation.

9. Describe the steps in the estimation procedure. Write down the equation for the contraction map that you will use in your inner loop.

10. Estimate the model using as instruments: cost shifters, BLP IVs, and differentiation IVs. Report your point estimates and standard errors.

11. Explore robustness to different starting values of θ_2 and different sets of instruments.

12. Compute a matrix of own and cross-price elasticities for market 1. Are they different than the elasticities you found in the pure logit model? Discuss.

Supply-Side Inferences

In industrial organization, we most typically use demand estimation to say something about the supply-side. That is where the problem set goes next.

13. Assume the observed prices respresent a Nash-Bertrand equilibrium in each market. Use firms' first order conditions to impute marginal costs for each product in all markets. This typically is done with a coding loop that considers each market in turn. Make two scatter plots: price against marginal cost and price against $\xi(\hat{\theta})$.

14. Now suppose that a tax of \$0.50 per unit is imposed in these markets. Model this as a cost shock, and compute the new equilibrium prices. What fraction of the tax is passed on to consumers? Specifically, calculate the average price increase in each market. Make a scatter plot of average price increase vs. the pre-tax Herfindahl-Hirschmann Index (HHI) in each market. What does this suggest about market concentration and pass-through?

15. Let the marginal cost function be

$$mc_{jt} = \gamma_j + \gamma_0 w_{jt} + \eta_{jt}$$

where w_{jt} is the cost shifter in your dataset and η_{jt} is a supply-side structural error term. Estimate γ_j and γ_0 via OLS. What assumptions must be true for your estimates to be consistent?