Assignment 2: Nonlinear Demand Estimation

Econ 543

Due: October 27

This homework assignment will give you practice working with the basics of the random coefficients multinomial logit demand model (Berry 1994). Compared to Problem set 1, we will now introduce three new sources of consumer heterogeneity:

- Consumer demographic characteristics (income)
- Unobserved preference for sugar
- Unobserved preference for caffeine

The file **product_data_ps2.csv** contains information on 50 different soft drinks, for 50 different markets. This dataset contains the following variables: product_ID (the product ID), type (a variable indicating if the product is a diet or a regular soda), price (the product's price during period t, in USD), quantity (the number of cans sold in market t), sugar (the grams of sugar contained in product j at time t), caffeine (the milligrams of caffeine contained in product j at time t), caffeine_extract_price (the price of caffeine at time t), corn_syrup_price (the price of high fructose corn syrup at time t), and t (the observation's time period). Throughout this problem set, assume that the market size is $M_t = \bar{M} = 10,000$ for all markets.

Let $\mathcal{J}_t = \{1, \ldots, J_t\}$ denote the set of all products available in market t. As usual, let $j = 0$ denote the outside good (not consuming any product).

Consumer i's indirect utility from consuming soda j in market t is given by the following expression:

$$u_{ijt} = \gamma + \alpha_{it} p_{jt} + \beta_{1i} sug_{jt} + \beta_{2i} caf_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

$$\begin{aligned} \text{with} \quad & \alpha_{it} = \bar{\alpha} + \alpha^y y_{it}, \\ \beta_{1it} = & \bar{\beta}_1 + \nu_{1it}\sigma_1, \quad \nu_{1it} \underset{iid}{\sim} N(0,1) \\ \beta_{2it} = & \bar{\beta}_2 + \nu_{2\text{it}}\, \sigma_2, \quad \nu_{2it} \underset{iid}{\sim} N(0,1) \end{aligned}$$

The utility of the outside option ($j = 0$) is normalized to be mean zero across consumers:

$$u_{\text{i0t}} = \varepsilon_{\text{iot}}$$

p_{jt} denotes the price (in USD) of product j at time t, sug_{jt} denotes j's sugar content (in grams) at time t, caf_{jt} denotes j's milligrams of caffeine at time t, and ξ_{jt} represents the unobserved quality of product j during time period t. y_{it} denotes the income of consumer i, and ν_{1it} and ν_{2it} represent i's unobserved preferences for sugar and caffeine, respectively. Finally, γ represents an intercept. Moreover, assume that ε_{ijt} are *iid* draws from a standard type 1 extreme value distribution.

Following BLP, the mean (common) utility derived by all consumers from product j in market t is given by:

$$\delta_{jt} = \bar{\alpha} p_{jt} + \bar{\beta}_1 sug_{jt} + \bar{\beta}_2 caf_{jt} + \gamma + \xi_{jt}$$

The consumer-specific (heterogeneous) component of utility is given by

$$\mu_{ijt} = \alpha^y y_{it} p_{jt} + \sigma_1 \nu_{i1t} sug_{jt} + \sigma_2 \nu_{i2t} caf_{jt}$$

1 Computing market shares with simulation draws

Problem 1(a). The file **consumer_census.csv** contains micro-level data on households' income for each market t in our sample. Let $\bar{y}_t$ denote the average income in market t, let σ_{y_t} denote the variance of y_{it} in market t, and let $\bar{p}_t = \sum_j p_{jt} s_{jt}$ denote the average price paid in market t.

- Compute $\bar{y}_t$, σ_{y_t}, and $\bar{p}_t$ for each market.

- Compute $\text{cor}(\bar{y}_t, \bar{p}_t)$. Do consumers in wealthier markets buy more expensive products on average?

Problem 1(b). Write a function that takes the observed distribution of income of each market, the number of consumers to be simulated (S), and a random seed (for replicability) as inputs, and uses pseudo-Monte Carlo draws to return a simulated sample of S consumers for each market t. For each consumer i and each market t, the output of this function must contain y_{it}, ν_{1it}, and ν_{2it} . Note that you will use the observed distribution of income (from the census) to draw each consumer's income.

1. Use this function to draw 50 samples of size 100 (50 different samples, each sample includes 100 simulated consumers). Report, for each sample, the mean and the variance of y_{it}, ν_{i1}, and ν_{i2}.
2. Use this function to draw a sample of 500 consumers for each market t.

Problem 1(c). Given our distributional assumptions over ε_{ijt}, the probability that consumer i in market t, with heterogeneity given by y_{it}, ν_{1it}, and ν_{2it} purchases product j is given by:

$$s_{ijt}(\delta_t, \theta_2) = \Pr\left(u_{ijt} \geq u_{ikt}, \forall k \in \mathcal{J}_t\right) = \frac{\exp\left(\delta_{jt} + \mu_{ijt}(\theta_2)\right)}{1 + \sum_k \exp\left(\delta_{kt} + \mu_{ikt}(\theta_2)\right)}$$

Let $\mathcal{S}$ denote the set of simulated consumers you obtained in Problem 1(b). Write a function that, given the parameters $\delta_t = \{\delta_{1t}, ..., \delta_{J_t,t}\}$ and $\theta_2 = \{\alpha, \sigma_1, \sigma_2\}$ computes the vector of individual choice probabilities $s_{ijt} = \Pr\left(u_{ijt} \geq u_{ikt} \forall k \in \mathcal{J}_t\right)$ for all $i \in \mathcal{S}$. You will use this function for part 1 (d).

Hint:

A common issue when trying to compute shares s_{ijt} is overflow, which arises due to rounding when trying to compute very large or very small values such as $\exp(800)$ (in which case the computational software can return Inf or $\exp(-800)$ (in which case it can return 0).

One way to deal with this is by using the log-sum-exp function

$$\text{LSE}(x) = \log \sum_k \exp x_k$$

Let $u_{it} = (0, u_{i1t}, \ldots, u_{iJ_t,t})$ denote the vector of utilities of all products, with

$$u_{ijt} = \delta_{jt} + \mu_{ijt}$$

Let $IncVal_{it} = \text{LSE}\left(u_{it}\right)$ We can alternatively compute s_{ijt} as

$$s_{ijt} = \exp\left(\left(\delta_{jt} + \mu_{ijt}\right) - IncVal_{it}\right)$$

Which guarantees overflow safety (see Conlon & Gortmaker (2020) for further details).

Problem 1(d). Suppose the linear parameters are $\bar{\alpha} = -1$, $\beta_1 = 0.5$, $\beta_2 = 0.5$, and $\gamma = -1.5$. Compute δ_{jt} for each product and each market, assuming $\xi_{jt} = 0, \forall\ j, t$. Call the vector of δs you get δ_t^p. Let $\theta_2' = (\alpha^{y'}, \sigma_1', \sigma_2') = (0, 0, 0)$. Compute the conditional shares $s_{ijt}(\delta_t^p, \theta_2')$ for each consumer. Let $\theta_2'' = (\alpha^{y''}, \sigma_1'', \sigma_2'') = (0, 1, 1)$. Compute the conditional shares $s_{ijt}(\delta_t^p, \theta_2'')$.Comment on how the conditional probabilities change if we use θ_2' or θ_2''.

Problem 1(e). Let $\hat{s}_{jt}$ denote the predicted market share of product j at time t, which is given by:

$$\hat{s}_{jt}\left(\delta_t, \theta_2\right) = \int_i \frac{\exp\left(\delta_{jt} + \mu_{ijt}\left(\theta_2\right)\right)}{1 + \sum_k \exp\left(\delta_{kt} + \mu_{ikt}\left(\theta_2\right)\right)} dF_t(i)$$

where θ_2 denotes the non-linear parameters of this model, δ_t is the vector of mean utilities of all products, and $F_t(i)$ denotes the distribution of consumer characteristics in market t.

Given that we have a sample of 500 consumers drawn from $F_t(i)$, we can approximate the integral over $F_t(i)$ by summing over all consumers in market t:

$$\hat{s}_{jt}\left(\delta_t, \theta_2\right) \approx \sum_{i \in \mathcal{S}_t} w_{it} s_{ijt} = \sum_{i \in \mathcal{S}_t} w_{it} \frac{\exp\left(\delta_{jt} + \mu_{ijt}\left(\theta_2\right)\right)}{1 + \sum_{k \in J_t} \exp\left(\delta_{kt} + \mu_{ikt}\left(\theta_2\right)\right)}$$

Where $\mathcal{S}_t$ denotes the set of simulated consumers, and $w_{it} = \frac{1}{|\mathcal{S}_t|} = \frac{1}{500}$.

Write a function that aggregates the individual choice probabilities of all consumers to compute the predicted market shares for each product, as a function of $\delta_t = (\delta_{1t}, ..., \delta_{J_t,t})$ and $\theta_2 = (\alpha, \sigma_1, \sigma_2)$. Given nonlinear parameters $\alpha^{y''} = 0$, $\sigma_1'' = 1$, $\sigma_2'' = 1$, the vector δ_t^p from Problem 1(d), and the sample of simulated consumers you obtained in part 1 (b), compute $\hat{s}_{jt}(\delta_t^p, \theta_2'')$ for all products and all markets. Compute $\bar{p}_t'' = \sum_j p_{jt}\hat{s}_{jt}(\delta_t^p, \theta_2'')$ and $\mathrm{cor}(\bar{p}_t'', \bar{y}_t)$. Compare the correlation between the average prices paid by consumers predicted by the model and those observed in the data (from Problem 1(a)). What does the difference between these results suggest about the true value of α^y?

2 Inverting shares to recover mean product qualities given non-linear parameters

Problem 2(a). Let s_{jt} be the market shares for product j in market t observed in the data, and $\hat{s}_{jt}\left(\delta_t, \theta_2\right)$ be the market shares predicted by the model. Berry (1994) shows that, given θ_2, there exists a unique vector $\hat{\delta}_t\left(\theta_2\right)$ that equates the observed and the predicted market shares:

$$s_{jt} = \hat{s}_{jt}\left(\hat{\delta}_t\left(\theta_2\right), \theta_2\right)$$

In order to find the vector $\hat{\delta}_t\left(\theta_2\right)$ that equates the observed and the predicted market shares, we can use the following contraction mapping:

$$\delta_t^{(\tau+1)} = \delta_t^{(\tau)} + \log\left(s_t\right) - \log\left(\hat{s}_t\left(\delta_t^{(\tau)}, \theta_2\right)\right)$$

where $\delta_t^{(\tau)}$ denotes the current-period guess for $\hat{\delta}_t\left(\theta_2\right)$. Write a function that employs this contraction mapping (or one of the alternatives discussed in Conlon & Gortmaker (2020)) to find $\hat{\delta}_j\left(\theta_2\right)$ for any given θ_2.

Problem 2(b). In this exercise, we will study how the accuracy of $\hat{\delta}_t\left(\theta_2\right)$ changes depending on the size of our sample of simulated consumers. As before, let $\theta_2'' = (0, 1, 1)$. For market $t = 1$ and $n = 1, 2, ..., 100$, draw a sample $\mathcal{S}_n$ of consumers using the function you wrote for Problem 1(b), each of them containing $S_n = 100$ simulated consumers (100 different samples, each sample must include 100 consumers). For each sample $\mathcal{S}_n \in \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_{100}\}$, let

$$\hat{\delta}_n(\theta_2'') = \left(\hat{\delta}_{1t_n}(\theta_2''), \ldots, \hat{\delta}_{J_t t_n}(\theta_2'')\right)$$

1. Plot the distribution of $\hat{\delta}_{1t_n}(\theta_2'')$.
2. Report the sample mean:
$$\bar{\hat{\delta}}_{1t} = \frac{1}{100} \sum_{n=1}^{100} \hat{\delta}_{1t_n}(\theta_2'')$$
3. Report the sample variance
$$\frac{1}{100} \sum_{n=1}^{100} \left(\hat{\delta}_{1t_n}(\theta_2'') - \bar{\hat{\delta}}_{1t}\right)^2$$
4. Repeat 1, 2, and 3 increasing the size of each sample from 100 to 500. How do the results from parts 2(a) and 2(b) compare to each other?

3 The objective function

Problem 3(a). In this problem, we will follow Gandhi and Houde (2023) to construct differentiation instruments $A_j(x_t, w_t)$ using the products' observed characteristics

$$x_{jt} = (1, sug_{jt}, caf_{jt})$$

and price instruments

$$w_{jt} = (\text{corn\_syrup\_price}_t \times sug_{jt}, \text{caffeine\_extract\_price}_t \times caf_{jt})$$

.

Let $\hat{p}_{jt}$ be the predicted values form the following regression

$$p_{jt} = \beta_0 + \phi_1 \text{corn\_syrup\_price}_t \times sug_{jt} + \phi_2 \text{caffeine\_extract\_price}_t \times caf_{jt} + \varphi_{jt}$$

Note that, by construction, $\hat{p}_{jt}$ satisfies $\mathbb{E}[\xi_{jt} \mid x_t, \hat{p}_t] = 0$

Our instrument function is given by

$$\left.\begin{array}{l} x_{jt} \\ \hat{p}_{jt} \\ \hat{p}_{jt}\bar{y}_t \\ \sum_{\substack{k\in J_t \\ k\neq j}} (caf_{kt} - caf_{jt})^2 \\ \sum_{\substack{k\in J_t \\ k\neq j}} (sug_{kt} - sug_{jt})^2 \end{array}\right\} = A_j(x_t, w_t) = z_{jt}$$

Where $\bar{y}_t$ denotes the average income in market t. Construct the instruments $A_j(x_t, w_t)$.

Problem 3(b). For each product j and each market t, let $z_{jt} \in \mathbb{R}^M$ denote the vector of instruments you constructed in part (a). Note that z_{jt} satisfies

$$\mathbb{E}[\xi_{jt} \mid z_{jt}] = 0$$

Therefore, we have that $\mathbb{E}[\xi_{jt} z_{jt}] = 0$.

Recall that

$$\xi_{jt} = \delta_{jt} - \bar{\alpha}_{p_{jt}} - \bar{\beta}_1 sug_{jt} - \bar{\beta}_2 caf_{jt} - \gamma$$

Let

$$\hat{\xi}_{jt}(\theta_2) = \hat{\delta}_{jt}(\theta_2) - \hat{\bar{\alpha}} p_{jt} - \hat{\bar{\beta}}_1 sug_{jt} - \hat{\bar{\beta}}_2 caf_{jt} - \hat{\gamma}$$

where $\hat{\theta}_1 = \left\{\hat{\bar{\alpha}}, \hat{\bar{\beta}}_1, \hat{\bar{\beta}}_2, \hat{\gamma}\right\}$ is the vector of parameters estimated by linear GMM:

$$\hat{\theta}_1 = \left(X^\top ZWZ^\top X\right)^{-1} X^\top ZWZ^\top \hat{\delta}(\theta_2)$$

Z is a $(J_t \times T) \times M$ matrix created by stacking the vectors z_{jt} for all j and all t on top of each other, and X is a $(J_t \times T) \times 4$ matrix created by stacking the vector $(p_{jt}, sug_{jt}, caf_{jt}, 1)$ for all j and all t on top of each other.

Let W be an $M \times M$ matrix given by[1]:

$$W = (Z^\top Z/(J_t \times T))^{-1}$$

Our vector of sample moment conditions g_1 is then given by

$$g_1(\theta_2) = \frac{1}{J_t \times T}\left(Z^\top \hat{\xi}_{jt}(\theta_2)\right)$$

[1]Note: for efficient GMM estimation, we would need to do a two step procedure. In the first stage, we would use a positive-definite weighting matrix (usually the 2SLS weighting matrix) and get some pilot estimates θ^p. In the second stage, we set the weighting matrix equal to the sample Var-Cov matrix of the moment conditions evaluated at θ^p. To make things simpler in this exercise, we will just perform the first-step estimation using the 2SLS weighting matrix, which is the optimal weighting matrix under the assumption of homoskedasticity.

The GMM objective function is defined as

$$q(\theta_2) = g_1(\theta_2)^\top W g_1(\theta_2)$$

The GMM problem is

$$\min_{\theta_2} q(\theta_2) = g_1(\theta_2)^\top W g_1(\theta_2)$$

Code the GMM objective function that you would use to estimate the nonlinear parameters θ_2, using the instruments you derived in part (a). Evaluate the objective function at $\tilde{\theta} = (-0.5, 2, 2)$.

Problem 3(c). The gradient of the GMM objective function is

$$\nabla q(\theta_2) = 2G(\theta_2)^\top W g(\theta_2)$$

with

$$\underbrace{G(\theta_2)}_{7\times 3} = \frac{1}{N} \underbrace{Z^T}_{7\times N} \underbrace{\frac{\partial \xi}{\partial \theta_2}}_{N\times 3}$$

$$\frac{\partial \xi}{\partial \theta_2} = \frac{\partial \delta}{\partial \theta_2}$$

$\frac{\partial \delta}{\partial \theta_2}$ can be separated by market t, after using the implicit function theorem:

$$\underbrace{\frac{\partial \delta_t}{\partial \theta_2}(\theta_2)}_{J_t\times 3} = -\underbrace{\left(\frac{\partial \hat{s}_t}{\partial \delta_t}(\theta_2)\right)^{-1}}_{J_t\times J_t} \underbrace{\left[\frac{\partial \hat{s}_t}{\partial \theta_2}(\theta_2)\right]}_{J_t\times 3}$$

$$\frac{\partial \hat{s}_t}{\partial \delta_t}(\theta_2) = \begin{bmatrix} \frac{\partial \hat{s}_{1t}}{\partial \delta_{1t}} & \cdots & \frac{\partial \hat{s}_{1t}}{\partial \delta_{J_t,t}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{s}_{J_t,t}}{\partial \delta_{1t}} & \cdots & \frac{\partial \hat{s}_{J_t,t}}{\partial \delta_{J_t,t}} \end{bmatrix}$$

$$\frac{\partial \hat{s}_t}{\partial \theta_2}(\theta_2) = \begin{bmatrix} \frac{\partial \hat{s}_{1t}}{\partial \alpha} & \frac{\partial \hat{s}_{1t}}{\partial \sigma_1} & \frac{\partial \hat{s}_{1,t}}{\partial \sigma_2} \\ \vdots & \vdots & \vdots \\ \frac{\partial \hat{s}_{J_t,t}}{\partial \alpha} & \frac{\partial \hat{s}_{J_t,t}}{\partial \sigma_1} & \frac{\partial \hat{s}_{J_t,t}}{\partial \sigma_2} \end{bmatrix},$$

with

$$\begin{aligned} \frac{\partial \hat{s}_{jt}}{\partial \delta_{kt}} &= \int \frac{\partial s_{ijt}}{\partial \delta_{kt}} dF(i) \\ &= \int s_{ijt}(1 - s_{ijt}) dF(i) \quad \text{if } j = k, \\ &= \int -s_{ijt} s_{ikt} dF(i) \quad \text{if } j \neq k. \end{aligned}$$

$$\frac{\partial \hat{s}_t}{\partial \theta_2}(\theta_2) = \begin{bmatrix} \frac{\partial \hat{s}_{1t}}{\partial \alpha} & \frac{\partial \hat{s}_{1t}}{\partial \sigma_1} & \frac{\partial \hat{s}_{1,t}}{\partial \sigma_2} \\ \vdots & \vdots & \vdots \\ \frac{\partial \hat{s}_{J_t,t}}{\partial \alpha} & \frac{\partial \hat{s}_{J_t,t}}{\partial \sigma_1} & \frac{\partial \hat{s}_{J_t,t}}{\partial \sigma_2} \end{bmatrix},$$

with

$$\frac{\partial \hat{s}_{jt}}{\partial \alpha^y} = \int y_{it} p_{jt} s_{ijt}(1 - s_{ijt}) - y_{it} s_{ijt} \left(\sum_{k \neq j} p_{kt} s_{ikt} \right) dF(i)$$

$$\frac{\partial \hat{s}_{jt}}{\partial \sigma_1} = \int \nu_{1it} sug_{jt} s_{ijt}(1 - s_{ijt}) - \nu_{1it} s_{ijt} \left(\sum_{k \neq j} sug_{kt} s_{ikt} \right) dF(i)$$

$$\frac{\partial \hat{s}_{jt}}{\partial \sigma_2} = \int \nu_{2it} caf_{jt} s_{ijt}(1 - s_{ijt}) - \nu_{2it} s_{ijt} \left(\sum_{k \neq j} caf_{kt} s_{ikt} \right) dF(i)$$

Write a function that computes the analytic gradient as a function of the vector of nonlinear parameters θ_2.

Problem 3(d). Compute the gradient of the GMM objective function numerically, evaluated at $\tilde{\theta} = (-0.5, 2, 2)$. Compare your result to the one you obtain using the function you wrote in Problem 3(c)

Problem 3(e). Estimate the vector $\hat{\theta}_2$ such that

$$\hat{\theta}_2 = \arg\min_{\theta_2} q\left(\theta_2\right) = g_1\left(\theta_2\right)^\top W g_1\left(\theta_2\right)$$

Report the estimated coefficients $\{\hat{\bar{\alpha}}, \hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}, \hat{\alpha}^y, \hat{\sigma}_1, \hat{\sigma}_2\}$.

4 References

1. Berry, S. T. (1994). *Estimating Discrete-Choice Models of Product Differentiation.* The RAND Journal of Economics, 25(2), pp. 242-262.

2. Conlon, Christopher & Gortmaker, Jeff. (2020). *Best practices for differentiated products demand estimation with PyBLP*. The RAND Journal of Economics, 51(4), 1108-1161.

3. Gandhi, A., & Houde, J.-F. (2023). *Measuring Substitution Patterns in Differentiated-Products Industries.* Working paper. Available at: https://www.dropbox.com/s/98h94s7rasvnnnr/GH_submission_ecma.pdf?dl=0