Empirical IO: Problem Set - 2

Not all questions need to be answered. There is a choice in Question 3.
The total marks for the problem set are 165. This will then be normalized to 20 using:

$$\text{Final Grade} = \frac{\text{your group's marks}}{165} \times 20.$$

What to submit: The problem set should be submitted in PDF and the R script. Your R code should be capable of (re-)producing the numbers that you present in the LaTeX tables.
Submission Deadline: The problem set must be submitted by **8-December-2025 23:59:59**. A one-hour delay leads to a deduction of 5 points (i.e., the maximum grade will be 15), a two-hour delay results in a deduction of 10 points, ..., and a delay of at least four hours results in a deduction of 20 points. The solutions will be released on 12-October-2025 as soon as the deadline passes. When you submit the problem set, please email both me and Tushar.

---

1 Problem 1 [45 marks]

You are given market shares of four products $s_1 = 0.1134, s_2 = 0.1638, s_3 = 0.3114, s_4 = 0.3114$. The utility of a consumer is given by $u_{ij} = \alpha * p_j + \delta_j + e_{ij}$, where e_{ij} is T1EV shock. Suppose the prices are $p_1 = 4, p_2 = 5, p_3 = 6, p_4 = 2$. No R-code should be used for this, you are supposed to do this with a calculator and math.

1) When for firm 1 markup i.e. $(p - MC_1)/MC_1 = 0.05$. Find α and δ_j's. Derive the equations you use for this part. [15 marks]

2) When for firm 1 own-price elasticity is 7.0928. Find α and δ_j's. Derive the equations you use for this part. [15 marks]

3) Suppose firm do not operate in a competitive environment and you suspect some coordination. Assume that $CV_{i \to j} = CV$ for all $i \neq j$. Also assume $(p - MC_1)/MC_1 = 0.05$ and $(p - MC_2)/MC_2 = 0.2$. Then what is CV, α, and δ_j's. Derive the equations you use for this part. [15 marks]

2 Problem 2 [60 marks]

Estimate the following three Random Coefficients models. Use the same dataset as used in Lecture-8 for this part and the same package. Below I am also writing the intercept term, however please note and recall from the lecture that you do not have to do anything additional as the intercept term is included by default, only focus on the named variables apart from the intercept term while creating the data objects.

1) Utility specification: [15 marks]

$$U_{ijm} = \beta_{\text{intercept},im} \cdot 1\{j \neq 0\} - \alpha_{im} \cdot p_{jm} + mushy_{jm} \cdot \beta_{mushy,im} + sugar_{jm} \cdot \beta_{sugar,im} + \xi_{jm} + \varepsilon_{ijm}$$

where $\beta_{\text{intercept},im} \cdot 1\{j \neq 0\}$ denotes the intercept term common to all products except for the outside good. Random coefficient specifcation:

$$\begin{bmatrix} \alpha_{im} \\ \beta_{im} \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix} + \nu_{im}$$

The parameteric restrictions needed are:

$$\Pi = \begin{bmatrix} \Pi_{(Intercept),\text{unobs\_sd}} \\ \Pi_{(\text{price}),\text{unobs\_sd}} \\ \Pi_{(\text{sugar}),\text{unobs\_sd}} \\ \Pi_{(\text{mushy}),\text{unobs\_sd}} \end{bmatrix}$$

2) Utility specification: [15 marks]

$$U_{ijm} = \beta_{\text{intercept},im} \cdot 1\{j \neq 0\} - \alpha_{im} \cdot p_{jm} + mushy_{jm} \cdot \beta_{im} + \xi_{jm} + \varepsilon_{ijm}$$

where $\beta_{\text{intercept},im} \cdot 1\{j \neq 0\}$ denotes the intercept term common to all products except for the outside good. Random coefficient specifcation:

$$\begin{bmatrix} \alpha_{im} \\ \beta_{im} \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix} + \Pi^{Income}_{(K+1)} Income_{rim} + \Pi^{age}_{(K+1)} age_{rim} + \nu_{im}$$

The parameteric restrictions needed are:

$$\Pi = \begin{bmatrix} \Pi_{(Intercept),\text{unobs\_sd}} & \Pi_{(Intercept),\text{income}} & NA \\ \Pi_{(\text{price}),\text{unobs\_sd}} & \Pi_{(\text{price}),\text{income}} & NA \\ \Pi_{(\text{mushy}),\text{unobs\_sd}} & NA & \Pi_{(\text{mushy}),\text{age}} \end{bmatrix}$$

3) Utility specification: [15 marks]

$$U_{ijm} = \beta_{\text{intercept},im} \cdot 1\{j \neq 0\} - \alpha_{im} \cdot p_{jm} + mushy_{jm} \cdot \beta_{mushy,im} + sugar_{jm} \cdot \beta_{sugar,im} + \xi_{jm} + \varepsilon_{ijm}$$

where $\beta_{\text{intercept},im} \cdot 1\{j \neq 0\}$ denotes the intercept term common to all products except for the outside good. Random coefficient specifcation:

$$\begin{bmatrix} \alpha_{im} \\ \beta_{im} \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix} + \Pi^{Income}_{(K+1)} Income_{rim} + \Pi^{age}_{(K+1)} age_{rim} + \nu_{im}$$

The parameteric restrictions needed are:

$$\Pi = \begin{bmatrix} \Pi_{(Intercept),\text{unobs\_sd}} & \Pi_{(Intercept),\text{income}} & NA \\ \Pi_{(\text{price}),\text{unobs\_sd}} & \Pi_{(\text{price}),\text{income}} & NA \\ \Pi_{(\text{mushy}),\text{unobs\_sd}} & NA & \Pi_{(\text{mushy}),\text{age}} \\ \Pi_{(\text{sugar}),\text{unobs\_sd}} & NA & \Pi_{(\text{sugar}),\text{age}} \end{bmatrix}$$

4) For each of the above calculate the elasticities and average price taste parameter for a voter with income 0.5 and age 0.0. (These are standardized, 0 means the mean age, 0.5 means 0.5 standard deviation from the mean). [15 marks]

3 Problem 3

Option 1: Demand and Supply Estimation (60 marks) (estimating_industry.csv)

You are provided with simulated market-level data for a differentiated-products industry. Each market m contains J products indexed by $j = 1, \ldots, J$, with observed characteristics RAM_j, Screen_j, price p_{jm}, market-level income income_m, market shares s_{jm}, and outside share s_{0m}. A simple logit demand system governs consumer choice, and firms set prices in static Bertrand–Nash equilibrium under single-product ownership. Each firm faces marginal costs that depend on product characteristics and on equilibrium output.

You will estimate (i) the demand parameters using an instrumental-variables logit regression, and (ii) the supply-side parameters using the implied marginal costs and an estimated supply equation.

Answer all parts. Show all algebraic steps clearly.

Model

Consumers have indirect utility

$$u_{ijm} = \beta_0 + \beta_1 \, \text{RAM}_j + \beta_2 \, \text{Screen}_j + \alpha_m \, p_{jm} + \xi_j + \varepsilon_{ijm},$$

where

$$\alpha_m = \alpha + \gamma \cdot \text{income}_m,$$

and ξ_j is an unobserved product characteristic common across markets.

Firms set prices by choosing p_{jm} to maximize profits $p_{jm} \cdot q_{jm} - \int_0^{q_{jm}} mc_{jm}(t)dt$, taking demand as given, where marginal cost is

$$mc_{jm}(q_{jm}) = c_0 + c_1 \, \text{RAM}_j + c_2 \, \text{Screen}_j + c_3 \, \text{input-prices}_j + \kappa \, q_{jm} + \omega_{jm}, \qquad q_{jm} = Q_m \, s_{jm}.$$

A. Theory (Derivation) — 20 marks

1) **(10 marks)** Starting from the utility specification of consumers, derive the linear estimating equation for the demand side. Explain each step.

2) **(10 marks)** Starting from the firm's profit-maximization problem derive the supply-side estimating equation.

B. Empirical Tasks in R — 40 marks

All steps below must be implemented in `R`. Include your code and output.

1) **(4 marks)** Construct the Berry transformed dependent variable:

$$\delta_{jm} = \log(s_{jm}) - \log(s_{0m}).$$

Show the exact `R` code used to create this variable.

2) **(8 marks)** Construct the following sets of instruments:

   a) **BLP rival-characteristics IVs**

$$Z_{jm}^{\text{ram}} = \sqrt{\frac{1}{J-1} \sum_{k \neq j} (\text{RAM}_k - \text{RAM}_j)^2}, \qquad Z_{jm}^{\text{scr}} = \sqrt{\frac{1}{J-1} \sum_{k \neq j} (\text{Screen}_k - \text{Screen}_j)^2}.$$

b) **Hausman–Nevo IVs:** mean price of product j in all *other* markets.

c) **Waldfogel IVs:** the number of competing products in the market.

d) The interaction of each instrument with income (including the interaction of input prices with income):

$$Z \times \text{income}_m.$$

Show the R code defining all instruments.

3) **(10 marks)** Estimate the demand model using the following three specifications (For each of the following you will also use its interactions with $income_m$):

a) BLP instruments only.

b) BLP + Hausman–Nevo instruments.

c) BLP + Hausman–Nevo + Input-prices

Do not use fixed effects for this part.

4) **(10 marks)** Using (c)'s estimates, estimate the supply-side implied by the model. Do not use fixed effects for this part. Suggest (if needed construct) the instruments you need and use them. Provide with three specifications that differ by the choice of instruments. The estimate of the coefficient with respect to the quantity should be close to 5×10^{-5}. Make sure all three specifications provide that. Assume that all markets have size, $Q_m = 10^5$.

5) **(8 marks)** Consider a counterfactual where a new product is introduced into every market with:

$$\text{Screen} = 16, \qquad \text{RAM} = 500, \qquad p = 2.$$

All competing products' prices remain fixed.

a) Compute market shares of all products and report their summary statistics. Essentially the mean shares, mean standard errors and how do these values change from the situation when this product was not present in the market.

OR

Option 2: Vote and Platform Choice Estimation (60 marks) (estimating_elections.csv)

You are provided with simulated constituency-level data for a national election. Each constituencyt m contains J candidates indexed by $j = 1, \ldots, J$, with observed characteristics educ_{jm}, crime_{jm}, ideology i_{jm}, constituency-level income income_m, vote shares s_{jm}, and candidate family income fam income_{jm}. A simple logit demand system governs consumer choice, and politicians choose ideology in static Nash equilibrium. Each candidate faces marginal disutility for deviating from their ideal ideological position, which we will denote by MC (you can think of this as marginal cost for ideological compromise).

You will estimate (i) the demand parameters using an instrumental-variable logit regression, and (ii) the supply-side parameters using the implied marginal costs and an estimated supply equation.

Answer all parts. Show all algebraic steps clearly.

Model

Consumers have indirect utility

$$u_{ijm} = \beta_0 + \beta_1 \,\text{educ}_{jm} + \beta_2 \,\text{crime}_{jm} + \alpha_m \, ideo_{jm} + w \cdot ideo_{jm}^2 + \xi_j + \varepsilon_{ijm},$$

where

$$\alpha_m = \alpha + \gamma \cdot \text{income}_m,$$

and ξ_j is an unobserved product characteristic common across markets. Note here a voter's ideal point for a politician's ideology is given by $\frac{\alpha_m}{2w}$, which varies from constituency to another.

Candidates choose ideology to maximize profits

$$s_{jm} - \int_0^{\text{ideo}_{jm}} MC(t)dt \tag{1}$$

$$MC_{jm}(q_{jm}) = c_0 + c_1 \,\text{educ}_{jm} + c_2 \,\text{crime}_{jm} + c_3 \,\text{fam income}_{jm} + \kappa \cdot \ \text{ideo}_{jm} + e_{jm}$$

Candidate's ideal point here is given by $\text{ideo}^*_{jm} = -\frac{1}{\kappa}\left(c_0 + c_1 \,\text{educ}_{jm} + c_2 \,\text{crime}_{jm} + c_3 \,\text{fam income}_{jm} + e_{jm}\right)$

A. Theory (Derivation) — 20 marks

1) **(10 marks)** Starting from the utility specification of consumers, derive the linear estimating equation for the vote-choice, the steps should be the same as in for the industry case. Explain each step.

2) **(10 marks)** Starting from the candidate's problem derive the supply-side estimating equation. Note this is not profit - cost, but vote-shares (s_{jm}) - dis-utility from ideological compromise. The integral essentially computes the dis-utility candidates feel from deviating from their ideal ideological position to increase their chances of winning.

B. Empirical Tasks in R — 40 marks

All steps below must be implemented in `R`. Include your code and output.

1) **(4 marks)** Construct the Berry transformed dependent variable:

$$\delta_{jm} = \log(s_{jm}) - \log(s_{0m}).$$

Show the exact `R` code used to create this variable.

2) **(8 marks)** Construct the following sets of instruments:

a) **BLP rival-characteristics IVs**

$$Z^{\text{educ}}_{jm} = \sqrt{\frac{1}{J-1}\sum_{k \neq j}(\text{educ}_{km} - \text{educ}_{jm})^2}, \qquad Z^{\text{crime}}_{jm} = \sqrt{\frac{1}{J-1}\sum_{k \neq j}(\text{crime}_{km} - \text{crime}_{jm})^2}.$$

b) **Hausman–Nevo IVs:** mean ideology of candidate from party j in all *other* constituencies.

c) **Waldfogel IVs:** the income of voters in other markets.

d) The interaction of each instrument with voter-income (including the interaction of input prices with income):

$$Z \times \text{income}_m.$$

Show the `R` code defining all instruments.

3) **(10 marks)** Estimate the vote-choice model using the following three specifications (For each of the following you will also use its interactions with $income_m$):

   a) BLP instruments only.
   b) BLP + Hausman–Nevo instruments.
   c) BLP + Hausman–Nevo + Input-prices

   Do not use fixed effects for this part.

4) **(10 marks)** Using (c)'s estimates, estimate the supply-side implied by the model. Do not use fixed effects for this part. Suggest (if needed construct) the instruments you need and use them. Provide with three specifications that differ by the choice of instruments.

   Do not use fixed effects for this part.

   [Hint: Right instruments will ensure that the ratio of your coefficient estimates and the standard deviation of residuals agree across specifications. i.e. `coefficients(m_supply_A)/sd(m_supply_A$residuals)` and `coefficients(m_supply_B)/sd(m_supply_B$residuals)` would be roughly the same).]

5) **(8 marks)** Consider a counterfactual where a new party enters:

$$\text{educ} = 100, \qquad \text{crime} = 0, \qquad \text{ideo} = 2.$$

   Assume competing candidates' ideology remain fixed.

   a) Compute vote-shares of all candidates and report their summary statistics. Essentially the mean shares, mean standard errors and how do these values change from the situation when this party was was not present in any constituency.