Problem Set 2
Empirical IO

Pavel Pronin

October 25, 2022

Contents

1 Problem 1

1.1 Static Model: Theory

- $i \in \{0, 1\}$ is the decision rule, where $i = 1$ if engine is replaced and $i = 0$ if engine is not replaced.
- $x \in \{0, 89\}$ – possible (already discretized) values of mileage.
- Let $c(x, \theta_1)$ denote the maintenance cost function associated with given x.
- Let RC denote the replacement costs.
- Let utility be defined as

$$u(x, i, \theta_1) = \begin{cases} -c(x, \theta_1) + \varepsilon_0 & i = 0 \\ -RC + \varepsilon_1 & i = 1 \end{cases}$$

  where ε is some unobserved error corresponding to choice $i \in \{0, 1\}$.
- Denote by $V(i)$ the observed part of utility function.
- Assume that ε_i follow extreme value distribution type I.

Decision Problem Agent Maximizes

$$\max_{i \in \{0,1\}} u(x, i, \theta_1)$$
$$\implies i = 0 \iff \varepsilon_1 - \varepsilon_0 \leq RC - c(x, \theta_1)$$
$$P_0 = \frac{\exp(RC - c(x, \theta_1))}{1 + \exp(RC - c(x, \theta_1))}$$
$$P_1 = 1 - P_0$$

Estimation Estimate using MLE

$$L = \prod_{n=1}^{N} (P_1 \cdot i_n + (1 - i_n) \cdot P_0)$$
$$LL = \sum_{n=1}^{N} [i_n \cdot \ln P_1 + (1 - i_n) \ln P_0]$$
$$\widehat{RC}, \hat{\theta}_1 \in \arg_{RC,\theta_1} \max LL$$
$$\text{s.t. } \forall x : c(x, \theta_1) \geq 0, RC \geq 0$$

For calculate of boostrap see procedure in Dynamic Model

1.2 Static Model: Implementation

1. Table 1 shows results for Linear Cost Model

$$c(x, \theta_1) = 0.001 \cdot \theta_{10} \cdot x$$

2. Table 2 shows results for Quadratic Cost Model

$$c(x, \theta_1) = 0.001 \cdot \left[\theta_{10} x + \theta_{11} x^2\right]$$

Table 1: Static Linear Cost Model

	Parameter	Estimate	SE (bootsrap, $B = 500$)
0	θ_{10}	26.270304	2.939348
1	RC	5.590550	0.164675

Table 2: Static Quadratic Cost Model

	Parameter	Estimate	SE (bootstrap, $B = 250$)
0	θ_{10}	9.641866	202.534683
1	θ_{11}	0.099207	2.316628
2	RC	5.108464	3.874612

1.3 Why Static Model is not good?

- Suppose that we know there are 2 time periods.

$$\max_{i_0, i_1} u(x_0, i_0, \theta_1) + \beta u(x_1, i_1, \theta_1)$$

Then clearly if x_0, x_1 are independent, then in optimal solution i_0 does not depend on x_1

- Suppose that we know that instead x_0 and x_1 are related and depend on i_t

$$x_1(i_0) = \begin{cases} \alpha x_0 & i_0 = 0 \\ 0 & i_0 = 1 \end{cases}$$
$$\alpha \geq 0$$

- Then decision makers chooses between

1. $(i_0, i_1) = (1, 1)$

$$-RC + \varepsilon_{10} + \beta(-RC + \varepsilon_{11})$$

2. $(i_0, i_1) = (0, 1)$

$$-c(x_0, \theta_1) + \varepsilon_{00} + \beta(-RC + \varepsilon_{11})$$

3. $(i_0, i_1) = (1, 0)$

$$-RC + \varepsilon_{10} + \beta(-c(0, \theta_1) + \varepsilon_{01})$$

4. $(i_0, i_1) = (0, 0)$

$$-c(x_0, \theta_1) + \varepsilon_{00} + \beta(-c(\alpha x_0, \theta_1) + \varepsilon_{01})$$

- Clearly, note that choice in period 1 does depend on the choice in period 0. If decision maker follows dynamic rule, he chooses $(i_0, i_1) = (0, 0)$ if

$$\begin{cases} (\varepsilon_{10} - \varepsilon_{00}) + \beta(\varepsilon_{11} - \varepsilon_{01}) \leq RC - c(x_0, \theta_1) + \beta(RC - c(\alpha x_0, \theta_1)) \\ \varepsilon_{11} - \varepsilon_{01} \leq RC - c(\alpha x_0, \theta_1) \\ \varepsilon_{10} - \varepsilon_{00} \leq RC - c(x_0, \theta_1) + \beta(c(0, \theta_1) - c(\alpha x_0, \theta_1)) \end{cases}$$

- If Decision Maker follows static role instead. He chooses $(i_0, i_1) = (0, 0)$ if

$$\begin{cases} \varepsilon_{10} - \varepsilon_{00} \leq RC - c(x_0, \theta_1) \\ \varepsilon_{11} - \varepsilon_{01} \leq RC - c(\alpha x_0, \theta_1) \end{cases}$$
$$\implies (\varepsilon_{10} - \varepsilon_{00}) + \beta(\varepsilon_{11} - \varepsilon_{01}) \leq RC - c(x_0, \theta_1) + \beta(RC - c(\alpha x_0, \theta_1))$$

- Therefore, optimal static decision coincides with 2 of the 3 conditions of the optimal dynamic decision. Note further that $c(0, \theta_1) < c(\alpha x_0, \theta_1)$ Therefore, clearly

$$R - c(x_0, \theta_1) \geq R - c(x_0, \theta_1) + \beta(c(0, \theta_1) - c(\alpha x_0, \theta_1))$$

  Thus,

  1. If necessary conditions for optimal dynamic decision are met $\implies$ conditions for static static decision are met
  2. If necessary conditions for optimal static decision are met $\not\Longrightarrow$ conditions for optimal dynamic decision are met

- Therefore, dynamic decision is better if

  1. i_t can influence x_{t+1}

1.4 Dynamic Model: Theory

Variables

- $C(x_t)$ – choice set, available values of i_t when state variable is x_t
- $\varepsilon_t \in \{\varepsilon_t(i) | i \in C(x_t)\}$ – A $\#C(x_t)$-dimensional vector of state variables observed by agent but not by researcher
- $x_t = \{x_t(1), ... x_t(K)\}$ – K-dimensional vector of state variables observed by both research and agent
- $u(x_t, i, \theta_1) + \varepsilon_t(i)$ – realized single-period utility of decision i when state variables are (x_t, ε_t), θ_1 is to be estimated.
- $p(x_{t+1}, \varepsilon_{t+1} | x_t, \varepsilon_t, i_t, \theta_2, \theta_3)$ – Markov transition density for state variables (x_t, ε_t) when alternative i_t is selected. θ_2, θ_3 to be estimated.
- $\theta = (\beta, \theta_1, \theta_2, \theta_3)$ – complete vector of $(1 + K_1 + K_2 + K_3)$ parameters to be estimated

Decision Problem Agent maximizes

$$V_\theta(x_t, \varepsilon_t) = \sup_{\Pi} \mathbb{E}_t \left\{ \sum_{j=t}^{\infty} \beta^{(j-t)} [u(x_j, f_j, \theta_1) + \varepsilon_j(f_j)] \middle| x_t, \varepsilon_t, \theta_2, \theta_3 \right\} \tag{1}$$

$$\Pi = \{f_t, f_{t+1}, ...\}, \forall t : f_t \in C(x_t)$$

where x_t is the state variable, ε_t is unobserved component (to the researcher), $f_t \in C(x_t)$ is the replacement decision, $(\beta, \theta_1, \theta_2, \theta_3)$ are exogenous parameters. $\mathbb{E}_t$ denote expectation w.r.t stochastic process (x_t, ε_t). Probability density of stochastic process (x_t, ε_t) is given by:

$$dp(x_{t+1}, \varepsilon_{t+1}, \ldots, x_{t+N}, \varepsilon_{t+N} | x_t, \varepsilon_t) = \prod_{i=t}^{N-1} p(x_{i+1}, \varepsilon_{i+1} | x_i, \varepsilon_i, f_i(x_i, \varepsilon_i), \theta_2, \theta_3)$$

Rust (1987): Policy rule The solution to Problem 1 is given by stationary policy rule:

$$i_t = f(x_t, \varepsilon_t, \theta)$$

Optimal value function is the unique solution to the Bellman's equation

$$V_\theta(x_t, \varepsilon_t) = \max_{i \in C(x_t)} \{u(x_t, i, \theta_1) + \varepsilon_t(i) + \beta \mathbb{E} V_\theta(x_t, \varepsilon_t, i)\}$$
$$\mathbb{E} V_\theta(x_t, \varepsilon_t, i) = \int_y \int_\eta V_\theta(y, \eta) p(dy, d\eta | x_t, \varepsilon_t, i, \theta_2, \theta_3)$$

optimal control policy function f is defined by

$$f(x_t, \varepsilon_t, \theta) \equiv \arg\max_{i \in C(x_t)} \{u(x_t, i, \theta) + \varepsilon_t(i) + \beta \mathbb{E} V_\theta(x_t, \varepsilon_t, i)\}$$

Conditional Independence (CI) Assumption: Assume that transition density of the controlled process $\{x_t, \varepsilon_t\}$ is given by

$$p(x_{t+1}, \varepsilon_{t+1} | x_t, \varepsilon_t, i, \theta_2, \theta_3) = q(\varepsilon_{t+1} | x_{t+1}, \theta_2) p(x_{t+1} | x_t, i, \theta_3)$$

This involves 2 restrictions:

1. x_{t+1} is sufficient statistic for ε_{t+1}, therefore any statistical relationship between ε_{t+1} and ε_t is transmitted entirely through x_{t+1}
2. Probability density of x_{t+1} depends only on x_t and not ε_t

Rust (1987): Theorem 1. Assume CI holds. Let $P(i|x, \theta)$ denote the conditional probability of choosing action $i \in C(x)$ given state variable x.

1. Let

   $$G([u(x, \theta_1) + \beta \mathbb{E} V_\theta(x)] | x, \theta_2)$$

   denote the *social surplus function* (conditional expectation of the maximum) corresponding to the density $q(\varepsilon | x, \theta_2)$, defined by

   $$G([u(x, \theta_1) + \beta \mathbb{E} V_\theta(x)] | x, \theta_2) \equiv \int_\varepsilon \max_{j \in C(x)} [u(x, j, \theta_1) + \beta \mathbb{E} V_\theta(x, j)] q(d\varepsilon | x, \theta_2)$$

2. Then $P(i|x, \theta)$ (conditional probability that the value under i is the largest) is given by

   $$P(i|x, \theta) = G_i([u(x, \theta_1) + \beta \mathbb{E} V_\theta(x)] | x, \theta_2)$$

   where G_i is the partial derivative of G w.r.t $u(x, i, \theta_1)$ and $\mathbb{E} V_\theta$ is the unique fixed point to a contraction mapping $\mathbb{E} V_\theta = T(\mathbb{E} V_\theta)$, defined by

   $$\mathbb{E} V_\theta(x, i) = \int_y G([u(y, \theta_1) + \beta \mathbb{E} V_\theta(y)] | y, \theta_2) p(dy | x, i, \theta_3)$$

Rust (1987). Properties of $G(\cdot)$ Define

$$G(r(x) | x, \theta_2) = \int_\varepsilon \left\{ \max_{i \in C(x)} [r(x, i) + \varepsilon(i)] \right\} q(\varepsilon | x, \theta_2) \lambda(d\varepsilon)$$

Suppose $q(\varepsilon | x, \theta_2)$ has finite first moments.

1. Then G is positively linear homogenous, convex function of $r(x)$
2. G has additive property

   $$\forall \alpha \in \mathbb{R} : G(r(x) + \alpha | x, \theta_2) = \alpha + G(r(x) | x, \theta_2)$$

Rust(1987): Theorem 2. Assume CI holds. The likelihood function l^f is given by

$$l^f(x_1, \ldots, x_T, i_1, \ldots, i_T | x_0, i_0, \theta) = \prod_{t=1}^{T} P(i_t | x_t, \theta) p(x_t | x_{t-1}, i_{t-1}, \theta_3),$$
$$P(i | x, \theta) = G_i([u(x, \theta_1) + \beta \mathbb{E} V_\theta(x)] | x, \theta_2)$$

Assumption: $q(\varepsilon|y,\theta_2)$ given by multivariate extreme value distribution Then

1. $$q(\varepsilon,x,\theta_2)=\prod_{j\in C(x)}\exp(-\varepsilon(j)+\theta_2)\exp(-\exp(-\varepsilon(j)+\theta_2))$$

   with $\theta_2, y = 0.577216$

2. Social surplus is

$$\begin{aligned}G([u(x,\theta_1)+\beta\mathbb{E}V_\theta(x)]|x,\theta_2) &\equiv \int_\varepsilon \max_{j\in C(x)}[u(x,j,\theta_1)+\beta\mathbb{E}V_\theta(x,j)]q(d\varepsilon|x,\theta_2)\\ &= \ln\left\{\sum_{j\in C(x)}\exp(u(x,j,\theta_1)+\beta V_\theta(x,j))\right\}\end{aligned}$$

3. Thus, probability is

$$\begin{aligned}P(i|x,\theta) &= G_i([u(x,\theta_1)+\beta\mathbb{E}V_\theta(x)]|x,\theta_2)\\ &= \frac{\exp(u(x,i,\theta)+\beta\mathbb{E}V_\theta(x,i))}{\sum_{j\in C(x)}\exp(u(x,j,\theta_1)+\beta V_\theta(x,j))}\end{aligned}$$

4. $\mathbb{E}V_\theta(x,i)$ is given by unique solution to the following contraction mapping

$$\mathbb{E}V_\theta(x,i)=\int_y \ln\left\{\sum_{j\in C(x)}\exp(u(y,j,\theta_1)+\beta\mathbb{E}V_\theta(y,j))\right\}p(dy|x,i,\theta_3)$$

Nested Fixed Point Algorithm

1. "Inner" (fixed point algorithm): for each value of θ, compute EV_θ from

$$\mathbb{E}V_\theta(x,i)=\int_y \ln\left\{\sum_{j\in C(x)}\exp(u(y,j,\theta_1)+\beta\mathbb{E}V_\theta(y,j))\right\}p(dy|x,i,\theta_3)$$

   This contraction mapping $\mathbb{E}V_\theta(x,i)=T(\mathbb{E}V_\theta(x,i))$ is **Frechet Differentiable**, therefore, we can use the **Newton-Kantorovich algorithm** to compute $\mathbb{E}V_\theta$.

2. "Outer" (hill climbing algorithm): search for value of θ that maximizes l^f

$$l^f(i_1,x_1,\ldots,i_T,x_T|i_0,x_0,\theta)=\prod_{t=1}^{T}P(i_t|x_t,\theta)p(x_t|x_{t-1},i_{t-1},\theta_3)$$
$$P(i|x,\theta)=\frac{\exp(u(x,i,\theta)+\beta\mathbb{E}V_\theta(x,i))}{\sum_{j\in C(x)}\exp(u(x,j,\theta_1)+\beta V_\theta(x,j))}$$

Preparing to Estimate

1. Choice variables

$$\forall t: i_t\in C(x_t)\equiv\{0,1\}$$

2. Utility functions

$$u(x_t,i,\theta_1)=\begin{cases}-RC-c(0,\theta_1)+\varepsilon_t(1) & i_t=1\\ -c(x_t,\theta_1)+\varepsilon_t(0)\end{cases}$$

3. Transition probabilities

$$p(x_{t+1}|x_t,i_t,\theta_3)=\begin{cases}g(x_{t+1},\theta_3) & i_t=1\\ g(x_{t+1}-x_t,\theta_3) & i_t=0\end{cases}$$

4. Discretization of x_t: 90 intervals of length 5000. Therefore, let $z \in \{0, \dots, 89\}$ denote the discretized value of x_t. From data we know that

$$\forall t : (z_{t+1} - z_t) \in \{0, 1, 2\}$$

Therefore, we can specify

$$g(z_{t+1} - z_t, \theta_3) = \begin{cases} \theta_{30} & z_{t+1} - z_t = 0 \\ \theta_{31} & z_{t+1} - z_t = 1 \\ 1 - \theta_{30} - \theta_{31} & z_{t+1} - z_t = 2 \end{cases}$$

5. Functional form for $c(\cdot)$ is one of the following:

   (a) Polynomial

$$c(z, \theta_1) = \theta_{11} z + \theta_{12} z^2 + \theta_{13} z^3$$

   (b) Exponential

$$c(z, \theta_1) = \theta_{11} \exp(\theta_{12} z)$$

   (c) Hyperbolic

$$c(z, \theta_1) = \frac{\theta_{11}}{91 - z}$$

   (d) Square root

$$c(z, \theta_1) = \theta_{11} \sqrt{z}$$

6. Normalization

   (a) Set $c(0, \theta_1) = 0$ regardless of functional form

   (b) Assume $\{\varepsilon(0), \varepsilon(1)\}$ follow Gumble type 1 extreme value process.

   (c) Normalize

$$\mathbb{E}[\{\varepsilon(0), \varepsilon(1)\}] = (0, 0)$$
$$Var\{\varepsilon(0), \varepsilon(1)\}] = \left(\frac{\pi^2}{6}, \frac{\pi^2}{6}\right)$$

Estimation

1. Take time-series for a single bus $(i_t, x_t)_{t=0}^T$ and form a likelihood

$$L^f(i_1, x_1, \dots, i_T, x_T | i_0, x_0, \theta) = \prod_{t=1}^T P(i_t | x_t, \theta) p(x_t | x_{t-1}, i_{t-1}, \theta_3)$$
$$P(i|x, \theta) = \frac{\exp(u(x, i, \theta) + \beta \mathbb{E} V_\theta(x, i))}{\sum_{j \in C(x)} \exp(u(x, j, \theta_1) + \beta V_\theta(x, j))}$$

2. **Stage 1.** Maximize L^1 (or l^1) to obtain $\hat{\theta}_3 = (\hat{\theta}_{30}, \hat{\theta}_{31})$

$$L^1(i_1, x_1, \dots, i_T, x_T | i_0, x_0, \theta) = \prod_{t=1}^T p(x_t | x_{t-1}, i_{t-1}, \theta_3)$$

   or equivalently

$$l^1(i_1, x_1, \dots, i_T, x_T | i_0, x_0, \theta) = \sum_{t=1}^T \ln p(x_t | x_{t-1}, i_{t-1}, \theta_3)$$

where

$$p(x_{t+1}|x_t, i_t\theta_3) = \begin{cases} g(x_{t+1}, \theta_3) & i_t = 1 \\ g(x_{t+1} - x_t, \theta_3) & i_t = 0 \end{cases}$$

$$g(x_{t+1} - x_t, \theta_3) = \begin{cases} \theta_{30} & x_{t+1} - x_t = 0 \\ \theta_{31} & x_{t+1} - x_t = 1 \\ 1 - \theta_{30} - \theta_{31} & x_{t+1} - x_t = 2 \end{cases}$$

Note that since $\hat{\theta}$ is obtained through MLE we have

$$\sqrt{n}(\hat{\theta}_3) - \theta \sim N(0, V_\theta)$$
$$V_\theta = I(\theta)^{-1}$$
$$\frac{\partial l_1}{\partial \theta_{30}} = \sum_{t=1}^{T} \frac{1}{p(x_t|x_{t-1}, i_{t-1}, \theta_3)} \frac{\partial p(x_t|x_{t-1}, i_{t-1}, \theta_3)}{\partial \theta_{3j}}$$
$$\frac{\partial p(x_t|x_{t-1}, i_{t-1}, \theta_3)}{\partial \theta_{30}} = \begin{cases} \frac{\partial g(x_t - x_{t-1}, \theta_3)}{\partial \theta_{30}} & i_{t-1} = 1 \\ \frac{\partial g(x_t, \theta_3)}{\partial \theta_{30}} & i_{t-1} = 0 \end{cases} = \mathbb{I}(x_t - x_{t-1} = 0) - \mathbb{I}(x_t - x_{t-1} = 2)$$
$$\gamma_t \equiv x_t - x_{t-1}$$
$$\frac{\partial l^1}{\partial \theta_{30}} = \sum_{t=1}^{T} \frac{\mathbb{I}(\gamma_t = 0) - \mathbb{I}(\gamma_t = 2)}{p(x_t|x_{t-1}, i_{t-1}, \theta_3)}$$
$$\frac{\partial^2 l^1}{\partial \theta_{30}^2} = -\sum_{t=1}^{T} \frac{[\mathbb{I}(\gamma_t = 0) - \mathbb{I}(\gamma_t = 2)]^2}{p(x_t|x_{t-1}, i_{t-1}, \theta_3)^2}$$
$$\frac{\partial l^1}{\partial \theta_{31}} = \sum_{t=1}^{T} \frac{\mathbb{I}(\gamma_t = 1) - \mathbb{I}(\gamma_t = 2)}{p(x_t|x_{t-1}, i_{t-1}, \theta_3)}$$
$$\frac{\partial^2 l^1}{\partial \theta_{31}^2} = -\sum_{t=1}^{T} \frac{[\mathbb{I}(\gamma_t = 1) - \mathbb{I}(\gamma_t = 2)]^2}{p(x_t|x_{t-1}, i_{t-1}, \theta_3)^2}$$
$$p(x_t|x_{t-1}, i_{t-1}, \theta_3) = \theta_{30}\mathbb{I}(\gamma_t = 0) + \theta_{31}\mathbb{I}(\gamma_t = 1) + (1 - \theta_{30} - \theta_{31})\mathbb{I}(\gamma_t = 2)$$
$$\implies \frac{\partial^2 l^1}{\partial \theta_{30}^2} = -\sum_{t=1}^{T} \begin{cases} \frac{1}{\theta_{30}^2} & \gamma_t = 0 \\ 0 & \gamma_t = 1 \\ \frac{1}{(1-\theta_{30}-\theta_{31})^2} & \gamma_t = 2 \end{cases} = -\sum_{t=1}^{T} \left[\frac{\mathbb{I}(\gamma_t = 0)}{\theta_{30}^2} + \frac{\mathbb{I}(\gamma_t = 2)}{(1 - \theta_{30} - \theta_{31})^2} \right]$$
$$\implies \frac{\partial^2 l^2}{\partial \theta_{31}^2} = -\sum_{t=1}^{T} \begin{cases} 0 & \gamma_t = 0 \\ \frac{1}{\theta_{31}^2} & \gamma_t = 1 \\ \frac{1}{(1-\theta_{30}-\theta_{31})^2} & \gamma_t = 2 \end{cases} = -\sum_{t=1}^{T} \left[\frac{\mathbb{I}(\gamma_t = 1)}{\theta_{31}^2} + \frac{\mathbb{I}(\gamma_t = 2)}{(1 - \theta_{30} - \theta_{31})^2} \right]$$
$$\frac{\partial^2 l^1}{\partial \theta_{30} \partial \theta_{31}} = -\sum_{t=1}^{T} \frac{(\mathbb{I}(\gamma_t = 0) - \mathbb{I}(\gamma_t = 2))(\mathbb{I}(\gamma_t = 1) - \mathbb{I}(\gamma_t = 2))}{p(x_t|x_{t-1}, i_{t-1}, \theta_3)^2}$$
$$\implies \frac{\partial^2 l^1}{\partial \theta_{30} \partial \theta_{31}} = -\sum_{t=1}^{T} \begin{cases} 0 & \gamma_t = 0 \\ 0 & \gamma_t = 1 \\ \frac{1}{(1-\theta_{30}-\theta_{31})^2} & \gamma_t = 2 \end{cases} = -\sum_{t=1}^{T} \frac{\mathbb{I}(\gamma_t = 2)}{(1 - \theta_{30} - \theta_{31})^2}$$

Taking mathematical expectation we have

$$\mathbb{E}\left[\frac{\partial^2 l^1}{\partial \theta_{30}^2}\right] = -T\left(\frac{1 - \theta_{31}}{\theta_{30}(1 - \theta_{30} - \theta_{31})}\right)$$
$$\mathbb{E}\left[\frac{\partial^2 l^1}{\partial \theta_{31}^2}\right] = -T\left(\frac{1 - \theta_{30}}{\theta_{31}(1 - \theta_{30} - \theta_{31})}\right)$$
$$\mathbb{E}\left[\frac{\partial^2 l^1}{\partial \theta_{30} \partial \theta_{31}}\right] = -T\left(\frac{1}{1 - \theta_{30} - \theta_{31}}\right)$$

Therefore, Fisher Information is

$$I_T(\theta_3) = \mathbb{E}\left[-\frac{\partial^2 l(\theta)}{\partial\theta_3^2}\right] = \frac{T}{1-\theta_{30}-\theta_{31}}\begin{bmatrix}\frac{1-\theta_{31}}{\theta_{30}} & 1\\ 1 & \frac{1-\theta_{30}}{\theta_{31}}\end{bmatrix}$$

$$I_T(\theta_3)^{-1} = \frac{1-\theta_{30}-\theta_{31}}{T}\begin{bmatrix}\frac{1-\theta_{31}}{\theta_{30}} & 1\\ 1 & \frac{1-\theta_{30}}{\theta_{31}}\end{bmatrix}^{-1} = \frac{1-\theta_{30}-\theta_{31}}{T}\left(\frac{(1-\theta_{31})(1-\theta_{30})-\theta_{30}\theta_{31}}{\theta_{30}\theta_{31}}\right)^{-1}\begin{bmatrix}\frac{1-\theta_{30}}{\theta_{31}} & -1\\ -1 & \frac{1-\theta_{31}}{\theta_{30}}\end{bmatrix}$$

$$I_T(\theta_3)^{-1} = \frac{\theta_{30}\theta_{31}}{T}\begin{bmatrix}\frac{1-\theta_{30}}{\theta_{31}} & -1\\ -1 & \frac{1-\theta_{31}}{\theta_{30}}\end{bmatrix}$$

Therefore we have

$$\hat{\theta}_3 \sim N\left(\begin{bmatrix}\theta_{30}\\ \theta_{31}\end{bmatrix}, \frac{\theta_{30}\theta_{31}}{T}\begin{bmatrix}\frac{1-\theta_{30}}{\theta_{31}} & -1\\ -1 & \frac{1-\theta_{31}}{\theta_{30}}\end{bmatrix}\right)$$

And we can use estimated values of $\hat{\theta}_3$ to approximate for $\hat{V}_{\hat{\theta}}$. Thus, standard errors is given by

$$se(\hat{\theta}_{30}) = \sqrt{\frac{\theta_{30}(1-\theta_{30})}{T}}, se(\hat{\theta}_{31}) = \sqrt{\frac{\theta_{31}(1-\theta_{31})}{T}}$$

and we can reconstruct

$$\hat{\theta}_{32} = 1-\hat{\theta}_{30}-\hat{\theta}_{31}$$

$$Var(1-\hat{\theta}_{30}-\hat{\theta}_{31}) = Var(\hat{\theta}_{30}+\hat{\theta}_{31}) = Var(\hat{\theta}_{30}) + Var(\hat{\theta}_{31}) + 2Cov(\hat{\theta}_{30},\hat{\theta}_{31})$$

$$se(\hat{\theta}_{32}) = \sqrt{\frac{\hat{\theta}_{30}(1-\hat{\theta}_{30}) + \hat{\theta}_{31}(1-\hat{\theta}_{31}) - 2\hat{\theta}_{30}\hat{\theta}_{31}}{T}}$$

3. **Stage 2.** Maximize L^2 (or l^2)

$$L^2(i_1, x_1, \dots, i_T, x_T|\theta) = \prod_{t=1}^{T} P(i_t|x_t, \theta_1)$$

or equivalently

$$l^2(i_1, x_1, \dots, i_T, x_T|\theta) = \sum_{t=1}^{T} \ln P(i_t|x_t, \theta_1)$$

where

$$P(i_t|x_t,\theta_1) = \frac{\exp(u(x_t, i_t, \theta_1) + \beta\mathbb{E}V_\theta(x_t, i_t))}{\sum_{j_t\in\{0,1\}}\exp(u(x_t, j_t, \theta_1) + \beta\mathbb{E}V_\theta(x_t, j_t))}$$

utilities are

$$u(x_t, i_t, \theta_1) = \begin{cases}-RC - \varepsilon_t(1) & i_t = 1\\ -c(x_t,\theta_1) + \varepsilon_t(0) & i_t = 0\end{cases}$$

$c(x_t, \theta_1)$ is given by one of the above functional forms and $\mathbb{E}V_\theta(x, i)$ is the solution to

$$\mathbb{E}V_\theta(x_t, i_t) = \sum_{x_{t+1}\in X_{t+1}} \ln\left\{\sum_{j\in\{0,1\}}\exp(u(x_{t+1}, j, \theta_1) + \beta V_\theta(x_{t+1}, j))\right\} p(x_{t+1}|x_t, i_t, \hat{\theta}_3)$$

where

$$p(x_{t+1}|x_t, i_t, \theta_3) = \begin{cases}g(x_{t+1}-x_t, \hat{\theta}_3) & i_t = 0\\ g(x_{t+1}, \hat{\theta}_3) & i_t = 1\end{cases}$$

As $x_{t+1} - x_t \in \{0,1,2\} \implies X_{t+1} = \{x_t, x_t+1, x_t+2\}$ with corresponding probabilities

$$x_{t+1}(i_t = 0) \in \left(x_t, \hat{\theta}_{30}; x_t+1, \hat{\theta}_{31}; x_t+2, \hat{\theta}_{32}\right)$$
$$x_{t+1}(i_t = 1) \in \left(0, \hat{\theta}_{30}; 1, \hat{\theta}_{31}; 2, \hat{\theta}_{32}\right)$$

Therefore, we can rewrite $\mathbb{E}V_\theta$ as

$$\begin{aligned}\mathbb{E}V_\theta(x_t, i_t = 0) = \hat{\theta}_{30} \ln &\left\{\sum_{j\in\{0,1\}} \exp(u(x_t, j, \theta_1) + \beta\mathbb{E}V_\theta(x_t, j))\right\} \\ &+ \hat{\theta}_{31} \ln \left\{\sum_{j\in\{0,1\}} \exp(u(x_t+1, j, \theta_1) + \beta\mathbb{E}V_\theta(x_t+1, j))\right\} \\ &+ \hat{\theta}_{32} \ln \left\{\sum_{j\in\{0,1\}} \exp(u(x_t+2, j, \theta_1) + \beta\mathbb{E}V_\theta(x_t+2, j))\right\}\end{aligned}$$

We can note that

$$\forall x_t : \mathbb{E}V_\theta(x_t, i_t = 1) = \mathbb{E}V_\theta(0, i_t = 0)$$

Therefore, we only need to compute $\mathbb{E}V_\theta(x_t, i_t = 0)$. Define

$$\mathbb{E}V_\theta(x) \equiv \mathbb{E}V_\theta(x, 0)$$
$$p(x_{t+1}|x_t) \equiv p(x_{t+1}|x_t, i_t = 0, \hat{\theta}_3)$$
$$\mathbb{E}V_\theta(x) = \sum_{y\in\{x,x+1,x+2\}} p(y|x_t) \ln\left\{e^{u(y,0,\theta_1)+\beta\mathbb{E}V_\theta(y,0)} + e^{u(y,1,\theta_1)+\beta\mathbb{E}V_\theta(y,1)}\right\}$$

Note that

$$u(y, 0, \theta_1) = -c(y, \theta_1)$$
$$u(y, 1, \theta_1) = -RC$$
$$\mathbb{E}V_\theta(y, 1) = \mathbb{E}V_\theta(0, 0) = \mathbb{E}V_\theta(0)$$
$$\mathbb{E}V_\theta(y, 0) = \mathbb{E}V_\theta(y)$$

Therefore, expression simplifies to

$$\mathbb{E}V_\theta(x) = \sum_{y\in\{x,x+1,x+2\}} p(y|x_t) \ln\left\{e^{-c(y,\theta_1)+\beta\mathbb{E}V_\theta(y)} + e^{-RC+\beta\mathbb{E}V_\theta(0)}\right\}$$

We can rewrite it in matrix form. Let P be (90×90) transition matrix, $\mathbb{E}V_\theta$ be the (90×1) column of $\mathbb{E}V_\theta$ values, $c(\theta_1)$ be (90×1) column of cost values,

$$\mathbb{E}V_\theta = \begin{bmatrix}\mathbb{E}V_\theta(0)\\ \vdots \\ \mathbb{E}V_\theta(89)\end{bmatrix}_{90\times1}, c(\theta_1) = \begin{bmatrix}-RC\\ -c(1,\theta_1)\\ \vdots \\ -c(89,\theta_1)\end{bmatrix}_{90\times1}, P = \begin{bmatrix}\theta_{30} & \theta_{31} & \theta_{32} & \dots & \dots & & & & & \\ 0 & \theta_{30} & \theta_{31} & \theta_{32} & \dots & \dots & & & & \\ 0 & 0 & \theta_{30} & \theta_{31} & \theta_{32} & \dots & \dots & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & \theta_{30} & \theta_{31} & 1-\theta_{32} \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \theta_{30} & 1-\theta_{30} \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1\end{bmatrix}_{90\times90}$$

Then we can express

$$\mathbb{E}V_\theta = P \cdot \ln\{\exp(-c(\theta_1) + \beta \cdot \mathbb{E}V_\theta) + \exp(-[c(\theta_1)]_0 + \beta \cdot [\mathbb{E}V_\theta]_0)\},$$

where β is a scalar and $[c(\theta_1)]_0, [\mathbb{E}V_\theta]_0$ are also scalars (first elements of corresponding columns). Note that we also can simplify expression about $P(i_t|x_t, \theta_1)$ as

$$P(0|x_t, \theta_1) = \frac{\exp(-c(x_t, \theta_1) + \beta\mathbb{E}V_\theta(x_t))}{\exp(-c(x_t, \theta_1) + \beta\mathbb{E}V_\theta(x_t)) + \exp(-RC + \beta\mathbb{E}V_\theta(0))}$$
$$P(0|\theta_1) = \frac{\exp\{-[c(\theta_1)]_{x_t} + \beta \cdot [\mathbb{E}V_\theta]_{x_t}\}}{\exp\{-[c(\theta_1)]_{x_t} + \beta \cdot [\mathbb{E}V_\theta]_{x_t}\} + \exp\{-[c(\theta_1)]_0 + \beta \cdot [\mathbb{E}V_\theta]_0\}}$$
$$P(1|\theta_1) = 1 - P(0|\theta_1),$$

where $P(0|\theta_1)$ is the (90×1) column vector, and $[\cdot]_{x_t}$ denotes x_tth row of corresponding column. Therefore, we can rewrite

$$\begin{aligned}
P(i_t|x_t,\theta_1) &= i_t \cdot [1 - P(0|\theta_1)]_{x_t} + (1-i_t)[P(0|\theta_1)]_{x_t} \\
&= [P(0|\theta_1)]_{x_t} + i_t \cdot (1 - 2 \cdot [P(0|\theta_1)]_{x_t}) \\
\ln P(i_t|x_t,\theta_1) &= \ln \{[P(0|\theta_1)]_{x_t} + i_t \cdot (1 - 2 \cdot [P(0|\theta_1)]_{x_t})\} \\
&= [\ln \{P(0|\theta_1) + i_t \cdot (1 - 2 \cdot P(0|\theta_1))\}]_{x_t} \\
l^2 &= \sum_{t=1}^{T} [\ln \{P(0|\theta_1) + i_t \cdot (1 - 2 \cdot P(0|\theta_1))\}]_{x_t}
\end{aligned}$$

4. **Stage 3.** Use $(\hat{\beta}, \hat{RC}, \hat{\theta}_1, \hat{\theta}_3)$ as initial starting values to maximize

$$L^f = \prod_{t=1}^{T} P(i_t|x_t,\theta) p(x_t|x_{t-1}, i_{t-1}, \theta_3)$$

$$l^f = \sum_{t=1}^{T} \{\ln P(i_t|x_t,\theta) + \ln p(x_t|x_{t-1}, i_{t-1}, \theta_3)\}$$

5. **Bootstrap SE.** Let B be the number of bootstrap replications. $\{(x_t, i_t)_{t=1}^T\}$ – original sample. Bootstrap sample b is $\{(x_t^b, i_t^b)_{t=1}^T\}$ where (x_t^b, i_t^b) are chosen from original sample $\{(x_t, i_t)_{t=1}^T\}$ with replacement. Whole bootstrap sample is

$$\left\{\{(x_t^b, i_t^b)_{t=1}^T\}\right\}_{b=1}^B$$

On each bootstrap sample calculate parameter estimates $\hat{\theta}_b$. Then **symmetric percentile CI** is defined as

$$CI^{|\%|}(\hat{\theta}) = \left[\hat{\theta} \mp q_{1-\alpha}^{|\%|*}\right],$$

where $1-\alpha$ is confidence level and $q_{1-\alpha}^{|\%|*}$ corresponds to $1-\alpha$ quantile of

$$\left\{\left|\hat{\theta} - \hat{\theta}_b\right|\right\}_{i=1}^B$$

Recall that asymptotic confidence interval is defined as

$$CI^{asy}(\theta) = \left[\hat{\theta} \mp se(\hat{\theta}) q_{1-\alpha/2}^{N(0,1)}\right],$$

where $q_{1-\alpha/2}^{N(0,1)}$ is $1-\alpha/2$ quantile from standard normal distribution.
Therefore, we can solve for $se(\hat{\theta})$

$$\hat{\theta} + q_{1-\alpha}^{|\%|*} = \hat{\theta} + se(\hat{\theta}) q_{1-\alpha/2}^{N(0,1)}$$

$$se(\hat{\theta}) = \frac{q_{1-\alpha}^{|\%|*}}{q_{1-\alpha/2}^{N(0,1)}}$$

Estimation In Short

1. Obtain $\hat{\theta}_3$ from:

$$(\hat{\theta}_{30}, \hat{\theta}_{31}) = \arg \max_{(\theta_{30},\theta_{31})} \sum_{t=1}^{T} \ln p(x_t|x_{t-1}, i_{t-1} = 0, \theta_3)$$

$$\text{s.t. } \textbf{NB!}\ \theta_{30} \geq 0, \theta_{31} \geq 0, 1 - \theta_{30} - \theta_{31} \geq 0$$

$$p(x_t|x_{t-1}, i_t = 0, \theta_3) = g(x_t - x_{t-1}, \theta_3)$$

$$g(x_{t+1} - x_t, \theta_3) = \begin{cases} \theta_{30} & x_{t+1} - x_t = 0 \\ \theta_{31} & x_{t+1} - x_t = 1 \\ 1 - \theta_{30} - \theta_{31} & x_{t+1} - x_t = 2 \end{cases}$$

$$\hat{\theta}_{32} = 1 - \hat{\theta}_{30} - \hat{\theta}_{31}$$

2. Calculate standard errors as

$$se(\hat{\theta}_{30}) = \sqrt{\frac{\theta_{30}(1-\theta_{30})}{T}}$$
$$se(\hat{\theta}_{31}) = \sqrt{\frac{\theta_{31}(1-\theta_{31})}{T}}$$
$$se(\hat{\theta}_{32}) = \sqrt{\frac{\hat{\theta}_{30}(1-\hat{\theta}_{30}) + \hat{\theta}_{31}(1-\hat{\theta}_{31}) - 2\hat{\theta}_{30}\hat{\theta}_{31}}{T}}$$

3. Construct $\hat{P}$ form $\hat{\theta}_3$, construct column-vector $c(\theta_1)$ from chosen functional form

4. Use $\hat{P}, c(\theta_1)$ to estimate using nested fixed point algorithm.

$$(\hat{RC}, \hat{\theta}_1) = \arg\max_{(RC,\theta_1)} \sum_{t=1}^{T} [\ln\{P(0|\theta_1) + i_t \cdot (1 - 2\cdot P(0|\theta_1))\}]_{x_t}$$
$$\textbf{NB!}\ \text{s.t. } \theta_1 \in \mathbb{R}, RC \geq 0$$
$$P(0|\theta_1) = \frac{\exp\{-[c(\theta_1)]_{x_t} + \beta\cdot[\mathbb{E}V_\theta]_{x_t}\}}{\exp\{-[c(\theta_1)]_{x_t} + \beta\cdot[\mathbb{E}V_\theta]_{x_t}\} + \exp\{-RC + \beta\cdot[\mathbb{E}V_\theta]_0\}}$$
$$\mathbb{E}V_\theta = \hat{P}\cdot\ln\{\exp(-c(\theta_1) + \beta\cdot\mathbb{E}V_\theta) + \exp(-RC + \beta\cdot[\mathbb{E}V_\theta]_0)\}$$

5. Use estimate values as starting values for maximization of full log-likelihood

$$\max\sum_{t=1}^{T}\{\ln P(i_t|x_t,\theta) + \ln p(x_t|x_{t-1}, i_{t-1}, \theta_3)\}$$

6. Create B bootstrap samples and for each boostrap sample $b \in \{1, \dots, B\}$ estimate $\hat{\theta}_b$, calculate $1-\alpha$ quantile of $\left\{|\hat{\theta} - \hat{\theta}_b|\right\}_{i=1}^{B}$ and calculate standard errors as

$$se(\hat{\theta}) = \frac{q_{1-\alpha}^{|\%|*}}{q_{1-\alpha/2}^{N(0,1)}}$$

1.5 Dynamic Model: Implementation

1. Table 3 shows estimates of linear cost model

$$c(x, \theta_1) = 0.001 \cdot \theta_{10} \cdot x$$

2. Table 4 shows estimates of quadratic cost model

$$c(x, \theta_1) = 0.001 \cdot \left[\theta_{10} \cdot x + \theta_{11} \cdot x^2\right]$$

3. Figure 1 shows estimated value functions corresponding to reported parameters

Table 3: Estimate of Linear Cost Model, $\beta = 0.9$

	Parameter	Estimate	SE	SE (Bootstrap, $B = 500$)
0	θ_{30}	0.351109	0.005252	0.029531
1	θ_{31}	0.637367	0.005290	0.177730
2	θ_{32}	0.011524	0.001174	0.184370
3	θ_{10}	3.200837		0.410856
4	RC	5.721047		0.193570

Table 4: Estimate of Quadratic Cost Model $\beta = 0.9$

	Parameter	Estimate	SE	SE (Bootstrap, $B = 250$)
0	θ_{30}	0.351109	0.005252	0.028056
1	θ_{31}	0.637367	0.005290	0.176798
2	θ_{32}	0.011524	0.001174	0.184577
3	θ_{10}	36.807117	NaN	8.656974
4	θ_{11}	-0.355080	NaN	0.078795
5	RC	10.964093	NaN	1.102520

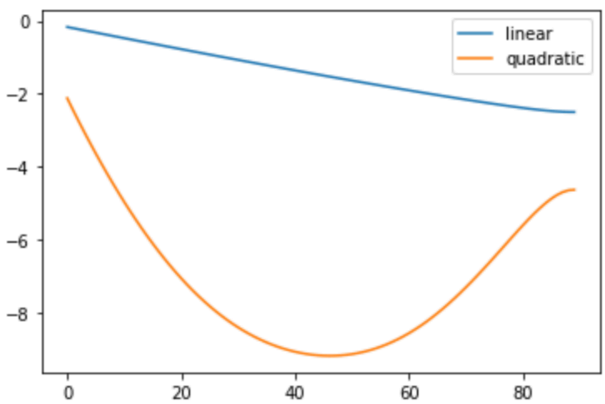

Figure 1: Estimated Value Function for given linear cost form

2 Report

2.1 Summary

- Authors review development in Empirical Industrial Economics (EIO) with focus on five major topics

  1. Estimating Demand in Imperfectly Competitive Markets. Industries are heterogenous and rather distinct to compare. The major trade-off is between the flexibility of substitution patterns and variation in the data. There are two solutions: (1) BLP (1995) and (2) Hausman (1997). One concern regarding the literature is that people do not elaborate enough on the identifying restrictions and identifying assumptions. Some of the best practices relate to modelling the consumer choice explicitly and further structurally estimating parameters.
  2. Market Power and Price Competition. Market power is hard to evaluate, especially due to complex substitution patterns that determine the degree of influence of each particular firm. Common researchers start with a model of general equilibrium, assume that equilibrium is achieved and estimate parameters. One of the most popular models is Bertrand-Nash equilibrium with differentiated products. Typically empirical work this requires data on costs that can either be gathered from accounting data (accounting costs $\neq$ economic costs) to proxy for economic marginal costs, or economic costs can be derived using structural modelling. The later approach requires strong assumption of profit-maximizing behaviour that is not necessarily met in practice. Interesting novel topics include modeling search costs, advertisement campaigns, retail sales, consumer stockpiling, adverse selection and non-price competition strategies.
  3. Competition in Auction Markets. Auction markets behave similar to competitive markets and economic theory has a variety of theoretical models and explanations to offer. Auction markets are ubiquitous too. Some examples include procurement markets and advertisement auctions conducted by Google and many other web-sites.
  4. Determinants of market Structure. There are two main approaches to study market entry: (i) to focus on specific cases, (2) to analyse and estimate strategic models of entry deterrence.
  5. Industry Dynamics. These studies focus on long-run effects, analysing life-cycles of firm dynamics. One of the key finding is that firms' life-cycles depend heavily on firm characteristics rather than on industry differences.

- They highlight that across all studies and fields there was a rapid shift in EIO methodology and approach to identification. Namely,

  1. Research shifted from cross-industry studies to within industry studies
  2. Research heavily relied upon economic theory and careful approaches to identification
  3. Research incorporated granular data

- However, there are also several limitations in the current research. That relate to the somewhat obfuscated reporting of identification assumption and too heavy focus on within-industry dynamics. Authors argue that yet powerful, this narrow intra-industry approach has provided many insights, but overall the field remains disaggregated and IO needs to come up with a more unifying perspectives.

2.2 Assessment

The paper is well-written and provides a broad summary of several decades of research in EIO. Authors also present their perspective on further developments of EIO as a field and argue with the quasi-experimental approach to identification that does not really on economic theory to much extent.

From my perspective, the paper would benefit massively from more in-depth discussion and comparison of various industries. Authors clearly point that EIO and IO in general requires unifying approach. It would be interesting to see how patterns in various industries overlap and what are the major differences. It would also be interesting to ask what characteristics of industries may be responsible for these divergences.