IO 2 Problem Set 2

Nadia Lucas and Fern Ramoutar

February 27, 2021

1 Entry Game

There are two firms, $i = 1, 2$ considering whether to enter markets $t = 1, \ldots, T$. The firm would like to enter if and only if net utility of entry is non-negative:

$$y_{it} = 1\left\{-\delta y_{-i,t} + \alpha x_t + \epsilon_{it} \geq 0\right\}$$

1.1 Complete Information

The firms observe $(\epsilon_{1t}, \epsilon_{2t})$, but the econometrician does not observe these. $(\epsilon_{1t}, \epsilon_{2t})$ is distributed iid across t as f_ϵ. In order to find the lower bound and upper bound of $\Pr\left(y_{1t} = 1, y_{2t} = 0 \mid x_t\right)$ in the spirit of Ciliberto and Tamer (2008), we can first summarize the realizations of the Nash equilibrium and the corresponding regions of $\epsilon_{1t}, \epsilon_{2t}$:

$$\begin{aligned}
\mathcal{R}_3 &= \{(y_{1t} = 0, y_{2t} = 0) \text{ is observed }\} = \{(\epsilon_{1t}, \epsilon_{2t}) : \mathbf{x}'_{1t}\alpha_1 + \epsilon_{1t} < 0 \text{ and } \mathbf{x}'_{2t}\alpha_2 + \epsilon_{2t} < 0\} \\
\mathcal{R}_1 &= \{(y_{1t} = 1, y_{2t} = 1) \text{ is observed }\} = \{(\epsilon_{1t}, \epsilon_{2t}) : \mathbf{x}'_{1t}\alpha_1 - \delta_1 + \epsilon_{1t} > 0 \text{ and } \mathbf{x}'_{2t}\alpha_2 - \delta_2 + \epsilon_{2t} > 0\} \\
\mathcal{R}_2 + \mathcal{R}_5 &= \{(y_{1t} = 1, y_{2t} = 0) \text{ is observed }\} = \{(\epsilon_{1t}, \epsilon_{2t}) : \mathbf{x}'_{1t}\alpha_1 + \epsilon_{1t} > 0 \text{ and } \mathbf{x}'_{2t}\alpha_2 - \delta_2 + \epsilon_{2t} < 0\} \\
\mathcal{R}_4 + \mathcal{R}_5 &= \{(y_{1t} = 0, y_{2t} = 1) \text{ is observed }\} = \{(\epsilon_{1t}, \epsilon_{2t}) : \mathbf{x}'_{1t}\alpha_1 - \delta_1 + \epsilon_{1t} < 0 \text{ and } \mathbf{x}'_{2t}\alpha_2 + \epsilon_{2t} > 0\}
\end{aligned}$$

We recognize that $\mathcal{R}_5$ is an overlapping region, so that if the realized value of $(\epsilon_{1t}, \epsilon_{2t})$ is in $\mathcal{R}_5$, then either $(y_{1t} = 1, y_{2t} = 0)$ or $(y_{1t} = 0, y_{2t} = 1)$ can be rationalized. The choice probability predicted by the model is:

$$\begin{aligned}
\Pr(y_{1t} = 1, y_{2t} = 0 \mid x_t) = {} & \Pr\left((\epsilon_{1t}, \epsilon_{2t}) \in \mathcal{R}_1\right) \\
& + \int \Pr\left((y_{1t} = 1, y_{2t} = 0) \mid \epsilon_{1t}, \epsilon_{2t}, x_t\right) 1\left[(\epsilon_{1t}, \epsilon_{2t}) \in \mathcal{R}_2\right] dF_{\epsilon_{1t}, \epsilon_{2t}}
\end{aligned}$$

Since the model provides us with inequality constraints in some regions of unobservables:

$$\begin{aligned}
\Pr\left((\epsilon_{1t}, \epsilon_{2t}) \in \mathcal{R}_1\right) \leq \Pr(y_{1t} = 1, y_{2t} = 0) \leq {} & \Pr\left((\epsilon_{1t}, \epsilon_{2t}) \in \mathcal{R}_1\right) \\
& + \Pr\left((\epsilon_{1t}, \epsilon_{2t}) \in \mathcal{R}_2\right)
\end{aligned}$$

We can bound estimates for the conditional probability of observing $(y_{1t}=1, y_{2t}=0)$ given (x_{1t}, x_{2t}):

$$\int_{\mathcal{R}_2} dF \leq \Pr\left(\{y_{1t}=1, y_{2t}=0\} \mid \mathbf{x}_{1t}, \mathbf{x}_{2t}\right) \leq \int_{\mathcal{R}_2} dF + \int_{\mathcal{R}_5} dF$$

1.2 Demand Estimation with Endogenous Entry

Suppose once firm i enters market t, the firm's market share S_{it} will be

$$S_{it} = z_{it}\beta + \gamma p_{it} + \xi_{it}$$

The structural errors $(\epsilon_{1t}, \epsilon_{2t}, \xi_{1t}, \xi_{2t})$ are distributed iid across t as $f_{\epsilon\xi}$. p_{it}

1. Even though the structural errors are distributed iid, if we run the above as an OLS regression we cannot consistently estimate γ and β if prices are endogenous, such that p_{it} is correlated with ξ_{it}.

2. We will get consistent estimates of γ if we have an instrument d_{it} that satisfies the following conditions: (1) $d_{it} \perp \xi_{it} \forall i, t$; (2) exclusion restriction holds; (3) existence of first stage; (4) monotonicity holds. We will also get consistent estimates of β if z_{it} is exogenous.

3. We can bound the distribution of firm 1's unobservables using the law of total probability, which implies:

$$\begin{aligned}
&\Pr\left(\xi_1 \leq t_1; (\nu_1, \nu_2) \in A^U_{(1,0)}, \mathbf{X}, \mathbf{Z}\right)\\
&\leq \Pr\left(\xi_1 \leq t_1; y_1 = 1; y_2 = 0, \mathbf{X}, \mathbf{Z}\right)\\
&= \Pr\left(S_1 - \alpha_1 P_1 - X_1\beta \leq t_1; y_1 = 1, y_2 = 0, \mathbf{X}, \mathbf{Z}\right)\\
&\leq \Pr\left(\xi_1 \leq t_1; (\nu_1, \nu_2) \in A^U_{(1,0)}, \mathbf{X}, \mathbf{Z}\right) + \Pr\left(\xi_1 \leq t_1; (\nu_1, \nu_2) \in A^M_{(1,0)}, \mathbf{X}, \mathbf{Z}\right)
\end{aligned}$$

1.3 Incomplete information

Now let ϵ_{it} be the private information of the firm i, and all the other variables are common knowledge to both firms. In this setting they play a static Bayesian Nash equilibrium. The econometrician observes $(y_{1t}, y_{2t}, x_t)_t$. Assume $\epsilon \sim \text{Logistic}(0,1)$ is iid across firms and markets. $x_t \in \{1, 2\}$ with $\Pr(x_t = 1) = .5$ is also independent.

1. Player i's strategy includes their posterior belief about the other player's type and the other player's strategy. If we normalize the expected payoff of the option to exit the market to be 0, then we can write down player i's strategy as

$$\begin{aligned}
y_{it} &= \mathbf{1}\left(\mathbf{x}'_{it}\alpha_i - \delta_i E\left[y_{-it} \mid \mathbf{x}_{-it}\right] + \epsilon_{it} \geq 0\right)\\
&= \mathbf{1}\left(\mathbf{x}'_{it}\alpha_i - \delta_i p_{-it} + \epsilon_{it} \geq 0\right)
\end{aligned}$$

where $p_{-it} := E\left[y_{-it} \mid \mathbf{x}_{-it}\right] = \Pr\left(y_{-it} = 1 \mid \mathbf{x}_{-it}\right)$ is the probability of entry. From the other firm's point of view, given the i.i.d nature of the ϵ_{it}'s, we can stack the above result and take expectations to get

$$p_{it} = 1 - F_\epsilon\left(-\mathbf{x}_{it}'\alpha_i + \delta_i p_{-it}\right)$$

2. Invoking the logit error assumption, we get the following set of equilibrium fixed-point equations:

$$\begin{aligned} p_{1t} &= \tfrac{\exp(\mathbf{x}_t'\alpha_1 - \delta_1 p_{2t})}{1+\exp(\mathbf{x}_t'\alpha_1 - \delta_1 p_{2t})} \\ p_{2t} &= \tfrac{\exp(\mathbf{x}_t'\alpha_2 - \delta_2 p_{1t})}{1+\exp(\mathbf{x}_t'\alpha_2 - \delta_2 p_{1t})} \end{aligned}$$

We calculate the Jacobian of this vector of probabilities. We use Newton's method to calculate the fixed points and search along a grid to identify possible multiple equilibria. When $(\alpha, \delta) = (1, 1)$, there is a unique symmetric BNE:

$$(p_{1t}, p_{1t}) \approx \begin{cases} (0.598942, 0.598942) & \text{if } x_t = 1 \\ (0.773249, 0.773249) & \text{if } x_t = 2 \end{cases}$$

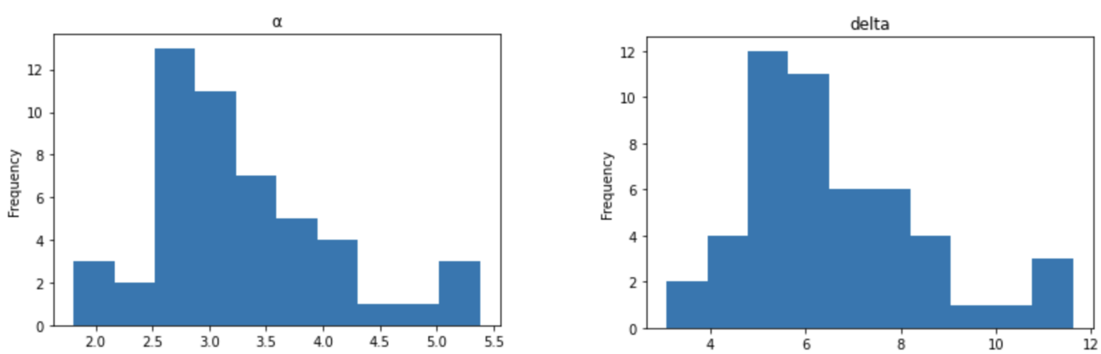
When $(\alpha, \delta) = (3, 6)$ there are multiple BNE:

$$(p_{1t}, p_{1t}) \approx \begin{cases} (0.07072, 0.92928) & \text{if } x_t = 1 \\ (0.5, 0.5) & \\ (0.92928, 0.07072) & \\ (0.71179, 0.849318) & \text{if } x_t = 2 \\ (0.784577, 0.784577) & \\ (0.849318, 0.71179) & \end{cases}$$

3. We use $(\alpha, \delta) = (3, 6)$ and its symmetric BNE to simulate the data and construct the log-likelihood:

$$l(\omega) = \sum_i \sum_t \log\left(p_{it}(\omega)\right)^{y_{it}}\left(1 - p_{it}(\omega)\right)^{1-y_{it}}$$

We estimate the parameters of the game via maximum likelihood. Running a Monte Carlo simulation with T=1000 and S=50, which gives us the following distributions:

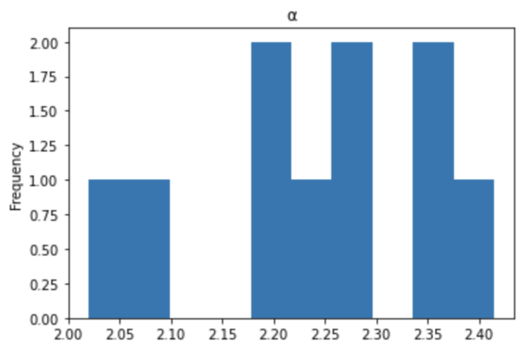


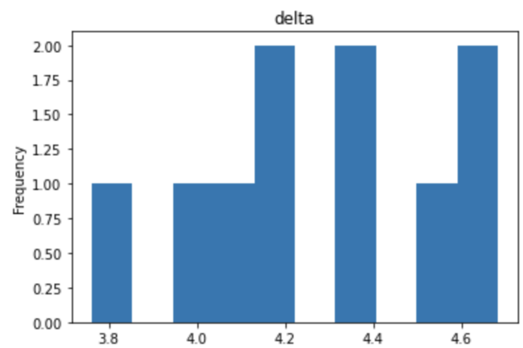
The mean of the α's is 3.25. The mean of the δ's is 6.57.

4. Now we suppose that firms do not necessarily use the symmetric strategy. If the model exhibits multiple equilibria, we allow the probability that a given equilibrium $k = 1, \ldots, K$ is selected to be $\lambda_{it}^{k}(x_t) \propto \exp(k/2)$. We can interpret these equilibrium selection probabilities as mixture weights. We simulate the data using the specified ordering rule - firm 1's choice probability in equilibrium k, $p_{1t}^{k}(x_t)$, satisfies $p_{1t}^{k}(x) \geq p_{1t}^{k+1}(x)$ for all $k = 1, \ldots, K-1$ and given x - and construct the multiple equilibria log-likelihood:

$$l(\omega) = \sum_i \sum_t \log \sum_k \lambda_{it}^{k} \left(p_{it}^{k}(\omega)\right)^{y_{it}} \left(1 - p_{it}^{k}(\omega)\right)^{1-y_{it}}$$

We again estimate the parameters of the game via maximum likelihood. Running a Monte Carlo simulation (we set S=10 in this simulation due to time constraints) gives us the following distributions:





The mean of the α's is 2.24. The mean of the δ's is 4.27.

5. Even though we could not show it in our simulation, we know that the estimator we proposed above is consistent with weights $\lambda_{it}^{k}(x_t) \propto \exp(k/2)$. Therefore, our estimator will be consistent for the new weights if $\lambda_{it}^{k}(x_t, u_t) \propto \exp(k/2)$. Since, $\lambda_{it}^{k}(x_t, u_t) \propto \exp\left((k + u_t)/2\right) \propto \exp(k/2)$, our estimator will be consistent.

2 Dynamics

2.1 The Data

2.2 The Model

2.3 Questions

1. See pset2_nfp.jl for all code related to this question

2. Conditional independence: Given (x_2, a_t), ϵ_t does not influence the realization of x_{t+1}. Next period always draw an entirely new ϵ_t. In other words we can write

$$p(x_{t+1}, \epsilon_{t+1} \mid x_t, \epsilon_t, a_t) = \rho(x_{t+1} \mid x_t, a_t) g(\epsilon_{t+1})$$

When ϵ_t terms are type-1 extreme value, this gives a lot of ease with the math.

3. Again, see code for more details. We interpreted the state space evolution written in equation (4) as dependent only on intervals, not on the previous state space at all. So all jumps in mileage intervals were informative for all others.

$$p(x_{t+1}) = \begin{cases} g(x_{t+1} - 0) & \text{if } d_t = 1 \\ g(x_{t+1} - x_t) & \text{if } d_t = 0 \end{cases}$$

I also augment the maximum state-space to include all miles driven as observed in the data plus the maximum possible interval. Then the probabilities are binned. I've settled on K=10 for computational time purposes.

4. If we define

$$EV(x, d) = \int V_\theta(y, \epsilon) p(d\epsilon) p(dy \mid x, d)$$

We can use the properties of the Bellman to first expand out the inner part of the integral, and then plug in our definition of EV

$$\begin{aligned} V(x, \epsilon) &= \max_{j \in A} \{u(x, j) + \epsilon_j + \beta E[V(y, \epsilon') \mid x, a = j]\} \\ &= \max_{j \in A} \left\{ u(x, j) + \epsilon_j + \beta \int_y \int_{\epsilon'} V(y, \epsilon') p(d\epsilon') p(dy \mid x, a = j) \right\} \end{aligned}$$

where the last equality stems from the conditional independence assumption. For simplification, set:

$$w(x) = \int V(x, \epsilon') p(d\epsilon')$$

Nowe we have a one-dimensional integral with

$$V(x, \epsilon) = \max_{j \in A} \left\{ u(x, j) + \epsilon_j + \beta \int_y w(y) p(dy \mid x, a = j) \right\}$$

Integrating both sides with respect to ϵ and using our definition of $w(x)$ from above, we get

$$w(x) = \int_{j \in A} \max_{j \in A} \left\{ u(x, j) + \epsilon_j + \beta \int_y w(y) p(dy \mid x, a = j) \right\}$$

Using the T1EV property of integral of maximum over utility plus T1EV error, we have

$$w(x) = \log \sum_{j \in A} \exp \left\{ u_j(x) + \beta \int_y w(y) p(dy \mid x, a = j) \right\}$$

Notice that our definition of $EV(x, d)$ can be rewritten using the definition of $w(x)$

$$EV(x, d) = \int_y w(y) p(dy \mid x, a = j)$$

Finally, plugging that into our definition of $w(x)$ we have

$$w(x) = \log \sum_{j \in A} \exp\left\{u_j(x) + \beta EV(y, j)\right\}$$

And now plugging this $w(x)$ into our definition of $EV(x, d)$

$$EV(x, d) = \int_y \log\left(\sum_{j \in A} \exp\left(u_j(x) + \beta EV(y, j)\right)\right) p(dy \mid x, d)$$

which is our objective.

5. The conditional choice probabilities stem from the value functions. In this case, I will call them

$$\begin{aligned} V1 &= u_1(x) + \beta EV(x, 1) \\ V0 &= u_0(x) + \beta EV(x, 0) \end{aligned}$$

For actions replace (1) and don't replace (0). Thus we can define

$$\begin{aligned} Pr(a = j \mid x) &= Pr\{v_j(x) + \epsilon_j \geq v_K(x) + \epsilon_K\} \\ &= \frac{\exp(v_j(x))}{\sum_k \exp(v_k(x))} \end{aligned}$$

Which using our great T1EV error terms, we can find a closed form solution for them. Specifically for this problem, we have

$$\begin{aligned} p_0 &= \frac{\exp(V0)}{\exp(V0) + \exp(V1)} \\ p_1 &= \frac{\exp(V1)}{\exp(V0) + \exp(V1)} \end{aligned}$$

Which can be rewritten to match notation on the slides as

$$p_0 = \frac{1}{1 + \exp(V1 - V0)}$$

For the purposes of our problem, p_k is defined as p_0 or the probability of not replacing.

2.3.1 Nested Fixed Point

1. The regenerative property allows us to reduce the state space of EV. Notice that $EV(x, 1)$ is the same for any state, x so it doesn't really matter what the state is, and notice, that it's also the same as starting back in the 0 mileage position and not replacing the bus engine. Thus we use $EV(0, 0)$ as the place that also holds all the information for $EV(x, 1)$ for all states x.

2. Starting on line 143 in the codefile, we start writing the fixed-point equation as a matrix equation.

3. The poly-algorithm is on line 238 and it employs the Newtoon-Kantarovich method with King-Werner steps.

4. The log-likelihood is given by

$$
\begin{aligned}
l(\theta) &= \sum_t d_t \log(1 - p_{x_t}(\theta)) + (1 - d_t)\log(p_{x_t}(\theta)) \\
p_k &= \frac{1}{1 + \exp(u(k,1 \mid \theta) + \beta EV(0 \mid \theta) - u(k,0 \mid \theta) - \beta EV(k \mid \theta)}
\end{aligned}
$$

5. Gradient of the likelihood: Using the chain rule we have

$$
\frac{\partial l(\theta)}{\partial \theta} = \sum_t \frac{d_t}{1 - p_{x_t}(\theta)} \frac{-\partial p_{x_t}(\theta)}{\partial \theta} + \frac{(1 - d_t)}{p_{x_t}(\theta)} \frac{\partial p_{x_t}(\theta)}{\partial \theta}
$$

Using the shorthand

$$
\begin{aligned}
V1 &\equiv u(k,1 \mid \theta) + \beta EV(0 \mid \theta) \\
V0 &\equiv u(k,0 \mid \theta) + \beta EV(k \mid \theta)
\end{aligned}
$$

Now solving for just

$$
\frac{\partial p_k(\theta)}{\partial \theta} = \frac{-\left[\frac{\partial u(k,1)}{\partial \theta} + \beta \frac{\partial EV(0)}{\partial \theta} - \frac{\partial u(k,0)}{\partial \theta} - \beta \frac{\partial EV}{\partial \theta}\right] * \exp(V1 - V0)}{(1 + \exp(V1 - V0)^2}
$$

Simplifying, notice that

$$
\begin{aligned}
p_k &= \frac{\exp(V0)}{\exp(V0) + \exp(V_1)} \\
&= \frac{1}{1 + \exp(V1 - V0)} \\
\implies 1 - p_k &= \frac{\exp(V1)}{\exp(V1) + \exp(V0)} \\
&= \frac{\exp(V1 - V0)}{1 + \exp(V1 - V0)} \\
\implies p_k(1 - p_k) &= \frac{\exp(V1 - V0)}{(1 + \exp(V1 - V0))^2}
\end{aligned}
$$

This implies

$$
\frac{\partial p_k(\theta)}{\partial \theta} = -p_k(1 - p_k)\left[\frac{\partial u(k,1)}{\partial \theta} + \beta \frac{\partial EV(0)}{\partial \theta} - \frac{\partial u(k,0)}{\partial \theta} - \beta \frac{\partial EV}{\partial \theta}\right]
$$

Finally we just have to solve out for

$$\frac{\partial u(k,1)}{\partial\theta} = \begin{bmatrix} -x \\ -\left(\frac{x}{100}\right)^2 \\ 0 \end{bmatrix}$$

$$\frac{\partial u(k,0)}{\partial\theta} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$$

$$\frac{\partial EV}{\partial\theta} = (I - \Gamma')^{-1}\frac{\partial\Gamma}{\partial\theta}$$

where

$$\frac{\partial\Gamma}{\partial\theta} = F_0 \frac{\exp(V1)\frac{\partial u(k,1)}{\partial\theta} + \exp(V0)\frac{\partial u(k,0)}{\partial\theta}}{\exp(V1) + \exp(V0)}$$

and plug everything back in to the original likelihood formulation.

This begins in line 299 in the code. Commented out are sanity checks that use a random θ to do sanity checks and compare against ForwardDiff at every step. The analytical estimates I was getting were very close at each step which helped with debugging.

At every point we can perform a derivative check against Automatic Differentiation, there is one available in the code, commented out.

6. To estimate the model parameters, we feed in a starting θ into an built-in Julia optimizer along with our maximum-likelihood and corresponding gradient functions. Unfortunately, the results are very sensitive to the starting value of θ. We end up performing a grid-search over various starting values of θ until honing in on a starting value that gave us the best results. This estimate for θ gave a likelihood of -.909

$$\hat{\theta} = \begin{bmatrix} 0.0294 \\ -1.1103 \\ 9.071 \end{bmatrix}$$

Economically speaking, this doesn't make a lot of intuitive sense for me because it results in an EV vector that trends slightly upward, or in other words, that more miles is slightly better for utility. We think this all stems from the fact that our transition probability matrix was derived from intervals only instead of being dependent on currents state, while more computationally tractable, we had a hard time wrapping our heads around this intuitively, and it did lead to some odd estimates for θ.