Assignment # 1

Ian Jaeyeon Kim

1 Results & Discussions

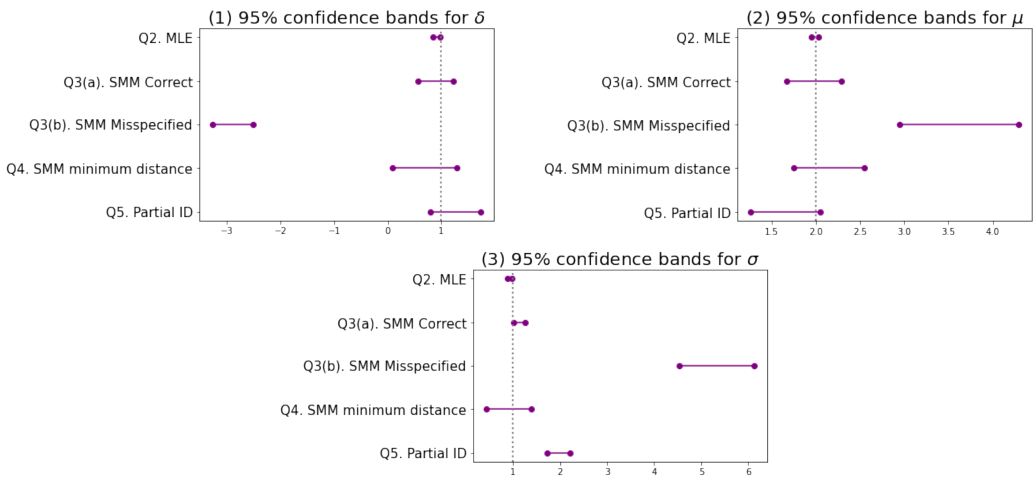

Figure 1. Graph (1), (2) and (3) describe estimated confidence bounds for δ, μ and $, \sigma$ respectively. In each graph, vertical grey dashed line stands for the location of the true parameter.

In the figure 1 above, I depict estimated results from each question. 'MLE' stands for the MLE estimator in question 1. 'SMM Correct' stands for the estimator in question 3 with the correctly specified model, whereas 'SMM Misspecified' means the estimator in the same question with the misspecified model. I named the estimator of question 4 as 'SMM minimum distance', since it only utilizes a single moment condition when it estimates three different parameters. Lastly, 'Partial ID' stands for the result from the question 5. Simple figure depicted above gives me following lessons: (1) If it's tractable, MLE is ideal, (2) Simulation works, (3) Individual information is valuable, but equilibrium selection matters, (4) Why I failed to recover σ in partial identification?

(1) If it's tractable, MLE is ideal In this practice, calculating the probability of the number of entrants observed in each market is tractable because the maximum number of potential entrants is 4. MLE estimator does not require any equilibrium selection and utilizes distributional assumption on cost shocks directly, in the sense that it does not rely on simulations. Hence it could be free from misspecification and fully utilized information observed from the data (except for individual firm's entry decision), which guarantees consistency and efficiency of MLE estimator as we can check from the figure 1. MLE estiamtor's shortest length of confidence band in figure 1 also substantiates the Cramer-Rao lower bound theorem. Therefore, in practice it is recommended to check the tractability of MLE estimator first.

(2) Simulation works As the maximum number of potential entrants goes up, applying MLE estimator becomes complicated. For this reason, Berry(1992) suggested simulation method which searches for the minimizers that minimize the distance between the observed number of entrants from the data and the expected number of entrants from the model with simulated cost shocks. As we can check from the figure 1, estimator named 'SMM minimum distance' performs okay in the sense that it entails true parameters in its confidence band. Although there exists efficiency loss from relying on the simulated errors compared to 'MLE', 'SMM minimum distance' is at least a tractable consistent estimator. In other words, we could achieve tractability by forgiving efficiency to some extent.

(3) Individual information is valuable, but equilibrium selection matters But, still there's a way to recover efficiency of simulation approach. Results of the estimator 'SMM Correct', which incorporates the information of individual firm's entry decision into the estimating procedure with the correct equilibrium selection assumption, shows thinner confidence bands compared to those of 'SMM minimum distance' while entailing true parameters in its bands. In other words, with a correct assumption for equilibrium mechanism we can achieve efficiency by incorporating more information into the estimation procedure. However, the **correct** assumption matters a lot. Results of the estimator 'SMM Misspecified', which assumed an incorrect equilibrium selection, show entirely incorrect estimates. This is because under the incorrect assumption, we let less cost efficient firms enter the market first, which makes our model structure incorrect under which true parameters can't bring the model to the observed data. Hence, knowledge about institutional background or reduced form research about the selection mechanism is essential for a correct model assumption.

(4) Why I failed to recover σ in partial identification? Under the assumption that my bound estimators $\hat{H}_1(\theta, X)$ and $\hat{H}_2(\theta, X)$ are correctly built, what matters for the bound estimation is the non parametric frequency estimator $\hat{P}(y|X)$. A critical step in constructing the frequency estimator is to discretize X space properly. If I discretize the X space into too small bins, then the resulting estimates will be inconsistent since each bin will not have enough data points in it. On the other hand, if I discretize the X space too sparsely, $\hat{P}(y|X)$ will be imprecise and the distance between the lower and upper bound $\hat{H}_1(\theta, X)$ and $\hat{H}_2(\theta, X)$ should be larger. To mitigate between the two extremes, I discretized my observed data space as follows:

1. market with 2 potential entrants: 3 bins for X_m $\times$ 2 bins for Z_{1m} $\times$ 2 bins for $Z_{2m} = 12$ cells
2. market with 3 potential entrants: 2 bins for X_m $\times$ 2 bins for Z_{1m} $\times$ 2 bins for Z_{2m} $\times$ 2 bins for $Z_{3m} = 16$ cells

3. market with 4 potential entrants: 4 bins for $X_m = 4$ cells

Under the above specification, partial identification succeeds to recover δ and μ, but fails to recover σ. I conjecture that under the current specification $\hat{P}(y|X)$ works enough to locate δ and μ, but it is not fully precise so the distance between the lower and upper bound can't be narrowed down. And the associated gap between the two bounds leads to high estimates for error variance σ.

To check my conjecture, I did some experiments by discretizing the X space thinner. The more I made my bins thinner, estimates for δ and μ deviate from the true values whereas estimate for σ gets close to 1. I think that the restricted number of observations ($M = 250$) does not allow me to discretize the X space to the extent that I could recover all the three parameters at the same time. After increasing the number of markets to 1,000 and modifying the data generating process to only generate markets with two potential entrants(in this case the size of y is only 4×1), I could recover all the three parameters.

I conclude that in the partial identification approach, discretizing X space in a proper manner matters. However, it was hard to come up with the 'proper' discretization. As Elena and Milena's working paper 'Data driven nests in discrete choice models' in which the authors search for the nests by applying clustering method, I think one can find a proper way to discretize X space in a more robust way.

2 Question 2

2.1 MLE Estimate

- **STEP 1** For each market m, calculate the probability of the market outcome N_m (the number of entrants).
  - $Pr(N^* = N_m) = Pr(N^* \leq N_m) - Pr(N^* \leq N_m - 1) = \sum_{J=0}^{N} H_{J(N+1)} - \sum_{J=0}^{N-1} H_{JN}$
  - where $H_{JN} = \sum_{s \in S(J)} \int_{A_1(s,N)} f(u_1)du_1 \cdots \int_{A_k(s,N)} f(u_k)du_k$ with $u_f \overset{iid}{\sim} N(\mu, \sigma)$
    - $S(J) := \{s : \sum_k s_k = J\}$
    - $A_k(s, N) = (-\infty, X_m - \delta \ln N_m - Z_{km})$ if $s_k = 1$
    - $A_k(s, N) = (X_m - \delta \ln N_m - Z_{km}, \infty)$ if $s_k = 0$
  - Strength and weakness of MLE come from $S(J)$. **Without assuming any equilibrium selection**, MLE considers all the possible market outcomes that make observed N_m as the equilibrium number of market entrants. Hence, MLE estimator can be free from the issue of model misspecification, which makes it **consistent**. Furthermore, in the procedure, MLE estimator directly utilizes the distributional assumption on cost shock in the sense that it does not rely on the simulation. And this makes MLE estimator the most **efficient** compared to the other estimators in this practice.
  - However, utilization of this consistent and efficient estimator heavily depends on the maximum number of potential entrants. In this practice, the maximum number of

potential entrants is 4, which allows me to work with $S(J)$ easily. As the number of potential entrants goes up, working with $S(J)$ in practice becomes more complicated.

- **STEP 2** Estimate $\hat{\theta}_{MLE} = \arg\min_{\theta} \ln \Pi_{m=1}^{M} Pr(N^* = N_m)$

2.2 Confidence Interval

- To estimate sample standard errors, I used the Hessian matrix calculated in the estimating procedure STEP 2 above.
- I used square root of diagonal elements of inversed Hessian divided by $\sqrt{n = 250}$ as sample standard errors.

3 Question 3

3.1 SMM Estimate

- **STEP1** Construct a vector of moment conditions $g_m(\theta)$.
  - Following Berry(1992), I included market and firm level moment conditions into $g_m(\theta)$. Since the minimum number of potential entrants in a market is two, I used only two firms' information to construct $g_m(\theta)$ to meet the conformity of matrices in the SMM procedure. Unlike Berry(1992), where the author utilized previous quarter's data to choose four firms with highest city share, in this practice I randomly chose two firms in each market since prior information about firms is not available in this practice.
  - With the randomly chosen two firms, I constructed following errors which are functions of the set of parameters (δ, μ, σ) conditional on observed data X.

$$v_{mo}(\theta; X_m) := \underbrace{\hat{N}_m(\theta; X_m)}_{\substack{\text{simulated number of entrants} \\ \text{from the model}}} - \underbrace{N_m(X_m)}_{\substack{\text{observed number of entrants} \\ \text{from the data}}} \quad \forall m = 1, \cdots, 250$$

$$v_{mk}(\theta; X_m) := \underbrace{\hat{E}_{mk}(\theta; X_m)}_{\substack{\text{simulated entry probability of firm } k \\ \text{from the model}}} - \underbrace{E_{mk}(X_m)}_{\substack{\text{observed entry decision of firm } k \\ \text{from the data}}} \quad \forall m = 1, \cdots, 250, k = 1, 2$$

  - Using the errors defined above, I defined $g_m(\theta)' := [v_{m0}(\theta), v_{m1}(\theta), v_{m2}(\theta), v_{m1}(\theta)^2, v_{m2}(\theta)^2]$
  - I also tried multiple set of moment conditions interacting v_{m0}, v_{m1}, v_{m2} with X_m, Z_{m1}, Z_{m2} and the number of potential entrants in each market. However, performance was not better compared to just including second moment of v_{m0}, v_{m1}, v_{m2} into moment conditions.

- **STEP2** Find a consistent estimate $\hat{\theta}_{1step}$
  - Define the sample analogy $\bar{g}_m(\theta) := \frac{1}{M} \sum_{m=1}^{M} g_m(\theta)$

- $\hat{\theta}_{1step} = \arg\min_\theta \overline{g}_m(\theta)'\overline{g}_m(\theta)$

- **STEP3** Estimate weighting matrix $\hat{W}$
  - Estimate robust variance $\hat{\Omega}^* = \frac{1}{M}\sum_{m=1}^{M}(g_m(\hat{\theta}_{1step}) - \bar{g}_m(\hat{\theta}_{1step}))(g_m(\hat{\theta}_{1step}) - \bar{g}_m(\hat{\theta}_{1step}))'$
  - Let $\hat{W} = \hat{\Omega}^{*-1}$

- **STEP4** Efficient point estimate
  - $\hat{\theta}_{2step} = \arg\min_\theta \overline{g}_m(\theta)'\hat{W}\overline{g}_m(\theta)$

3.2 Confidence Interval

- Our simulation approach does not guarantee the differentiability of the objective function. So I followed the method of Pakes and Pollard (1989). Recall the formula:

$$\sqrt{n}(\hat{\theta}_{2step} - \theta_0) \to N(0, (\Gamma'\Gamma)^{-1}\Gamma' V \Gamma(\Gamma'\Gamma)^{-1})$$

  where Γ is the Jacobian of sample analogy $\bar{g}_m(\hat{\theta}_{2step})$ with respect to θ, and V will be replaced with its robust estimator $\hat{\Omega}^*$.

- Square root of each diagonal element of $(\Gamma'\Gamma)^{-1}\Gamma'\hat{\Omega}^*\Gamma(\Gamma'\Gamma)^{-1}$ divided by $\sqrt{n = 250}$ will be our standard errors.

4 Question 4

4.1 SMM Estimate

- Unlike the GMM like approach used in Question 3 where I worked with over-identifying five moment conditions, I adopted minimum distance approach in Question 4 with a single moment condition. In other words, estimation strategy of Question 4 under-identifies the set of parameter and I had to minimize the objective function coming from single moment condition as much as possible.

- **STEP1** Construct an error value $v_{m0}(\theta|X_i)$ as follows:

$$v_{mo}(\theta; X_m) := \underbrace{\hat{N}_m(\theta; X_m)}_{\substack{\text{simulated number of entrants}\\ \text{from the model}}} - \underbrace{N_m(X_m)}_{\substack{\text{observed number of entrants}\\ \text{from the data}}} \quad \forall m = 1, \cdots, 250$$

- **STEP2** Estimate $\hat{\theta}_{md}$

$$\hat{\theta}_{md} = \arg\min_\theta \frac{1}{M}\sum_{m=1}^{M} v_{mo}(\theta; X_m)^2$$

4.2 Confidence Interval

- The number of moment conditions is smaller than the number of parameters we estimate. Therefore, I could not apply the method of Pakes and Pollard (1989) because $\Gamma'\Gamma$ becomes singular.

- Hence, for question 4, I reported empirical confidence interval which I derived from 100 times of bootstrapping.

- Also, I reported mean value of bootstrapping results for the point estimates.

5 Question 5

5.1 construct the confidence region

- **STEP1** Construct the sample objective function $Q_n(\theta)$
  - Construct a simple frequency estimator $\hat{P}(y|X_m)$
  - Construct model predicted lower and upper bounds $\hat{H}_1(\theta, X_m), \hat{H}_2(\theta, X_m)$
  - Define $Q_n(\theta) = \frac{1}{m}\sum_{m=1}^{M} ||(\hat{P}(y|X_m) - \hat{H}_1(\theta, X_m))_-|| + ||(\hat{P}(y|X_m) - \hat{H}_2(\theta, X_m))_+||$
  - $*$ more details are described in the MIEQ_functions.py

- **STEP2** Discretize the parameter space $\Theta = \Delta \times M \times \Sigma$
  - Define $\Delta', M', \Sigma' = \text{linspace}(0, 3, 20)$
  - Construct $\Theta' = \Delta' \times M' \times \Sigma'$
  - In other words, I worked with 8,000 many grid points on $(0, 3)^3$ space

- **STEP3** Choose initial cutoff level $c_{(0)}$
  - Sparsely distributed grid points in Θ' will inflate the values $Q_n(\theta) - \min_{t\in\Theta} Q_n(t) \forall\theta \in \Theta'$
  - As a simple remedy, I chose $c_{(0)} = 2 \times \min_{t\in\Theta} Q_n(t)$
  - To be more precise, we should make the grid space thinner or use simulated annealing method to make Θ' more precise. In this practice, I chose to work with simpler grid space to lessen the computational burden and concentrate on implementing the method.

- **STEP4** Update $c_{(0)}$ to $c_{(1)}$
  - Filter the level set $C_n(c_{(0)}) = \{\theta \in \Theta' : Q_n(\theta) - \min_{t\in\Theta} Q_n(t) \leq \frac{c_{(0)}}{n}\}$ where $n = 250$
  - Set bootstrapping sample size $n_b = 80$ and the number of bootstrapping $B = 100$. And for each bootstrapping $b \in B$:
    1. Randomly choose n_b many market data from the initial 250 market data
    2. calculate $T_b = \max_{\theta\in\Theta'}\{n_b \cdot (Q_{b,n_b}(\theta) - \min_{t\in\Theta} Q_n(t))\}$

- Construct $T = \{T_1, T_2, \cdots T_B\}$
- Define $c_{(1)} =$ 5-quantile of the set T

- **STEP5** Following the procedure in STEP4, update $c_{(1)}$ to $c_{(2)}$
- **STEP6** Report $C_n(c_{(2)}) = \{\theta \in \Theta' : Q_n(\theta) - \min_{t \in \Theta} Q_n(t) \leq \frac{c_{(2)}}{n}\}$ where $n = 250$