**Economics 628** **Term 1 2017/2018**

**Topics in Applied Econometrics I** **Hiro Kasahara**

# Homework 2

(Due: Monday, October 2 at the start of the class)

Note: Study groups *discussing* the problems are strongly encouraged. But please write your own answers and submit your own programs (no copy and paste from your classmate's program!).

## 1 Dynamic Discrete Choice Model

Consider a dynamic discrete choice model:

$$y_{it}^* = \alpha + \rho y_{i,t-1} + c_i + \epsilon_{it}, \quad y_{it} = 1(y_{it}^* > 0) \tag{1}$$

where $\epsilon_{it}$ is i.i.d., independent of $c_i$, and logistic distributed with unit variance such that

$$\Pr(y_{it} = 1|\{y_{i,t-s}\}_{s=1}^t, c_i) = \Pr(y_{it} = 1|y_{i,t-1}, c_i) = \Lambda(\alpha + \rho y_{i,t-1} + c_i),$$

where $\Lambda(x) = \exp(x)/(1+\exp(x))$. Further, conditional on $c_i$, $y_{i0}$ is independently drawn from the stationary distribution of the stochastic process implied by $\Pr(y_{it} = 1|y_{i,t-1}, c_i)$. The unobserved variable $c_i$ represents an individual-specific effect, independently distributed across individuals.

For example, you can think that $y_{it}$ is a discrete exporting decision while $c_i$ represents a firm's permanent characteristics that affect the export profit but are not observed from researchers. For the firms that operate in the market long enough before $t = 0$, the distribution of $y_{i0}$ would be given by the stationary distribution.

1. Denote $\lambda_{11}(\alpha, \rho, c) = \Pr(y_t = 1|y_{t-1} = 1, c) = \Lambda(\alpha+\rho+c)$ and $\lambda_{00}(\alpha, \rho, c) = \Pr(y_t = 0|y_{t-1} = 0, c) = 1 - \Lambda(\alpha + c)$. Show that

$$\Pr(y_0 = 1|\alpha, \rho, c) = \frac{1 - \lambda_{00}(\alpha, \rho, c)}{2 - \lambda_{00}(\alpha, \rho, c) - \lambda_{11}(\alpha, \rho, c)} = \frac{\Lambda(\alpha + c)}{1 - \Lambda(\alpha + \rho + c) + \Lambda(\alpha + c)},$$

so that

$$\Pr(y_{i0}|\alpha,\rho,c_i) = \left(\frac{\Lambda(\alpha+c_i)}{1-\Lambda(\alpha+\rho+c_i)+\Lambda(\alpha+c_i)}\right)^{y_{i0}} \left(\frac{1-\Lambda(\alpha+\rho+c_i)}{1-\Lambda(\alpha+\rho+c_i)+\Lambda(\alpha+c_i)}\right)^{1-y_{i0}} \tag{2}$$

for $y_{i0} \in \{0,1\}$.

2. Suppose that $c_i$ takes two values $\{\gamma^L, \gamma^H\}$ with $\Pr(c_i = \gamma^L) = \pi$ and $\Pr(c_i = \gamma^H) = 1 - \pi$. Under the parameter values $(\alpha, \rho, \gamma^L, \gamma^H, \pi) = (-1.0, 0.5, -1.0, 1.0, 0.5)$, randomly draw an observation $\{y_{it}\}_{t=0}^T$ for $i = 1, ...N$ as follows.

Step 1: For each $i$, draw $c_i$ by drawing $u_i$ ~Uniform$[0,1]$ using 'rand' command and then by setting $c_i = \gamma^L$ if $u_i \leq \pi$, or $c_i = \gamma^H$ if $u_i > \pi$.

Step 2: Given the realized value of $c_i$, compute $\Pr(y_{i0} = 1|\alpha,\rho,c_i) = \frac{\Lambda(\alpha+c_i)}{1-\Lambda(\alpha+\rho+c_i)+\Lambda(\alpha+c_i)}$. Draw $p_{i0}$ ~Uniform$[0,1]$ and set $y_{i0} = 1$ if $p_{i0} \leq \Pr(y_{i0} = 1|c_i, \alpha, \rho)$, or set $y_{i0} = 0$ otherwise.

Step 3: Given $c_i$ and $y_{i,t-1}$, generate $y_{it}$ as follows. Draw $p_{it}$ ~Uniform$[0,1]$ and set $y_{it} = 1$ $\Pr(y_{it} = 1|y_{i,t-1}, c_i) = \Lambda(\alpha + \rho y_{i,t-1} + c_i)$. Repeating this for $t = 1, ..., T$, we have $\{y_{it}\}_{t=0}^T$.

Step 4: Repeat Steps 1-4 for $i = 1, ..., N$ and store them into an $(T+1) \times N$ matrix $Y = [y_1, ..., y_N]$ with $y_i = (y_{i0}, ..., y_{iT})'$.

Generate two different data sets with $(T, N) = (2, 20000)$ and $(T, N) = (10, 4000)$.

3. The log likelihood function given the data $Y$ is

$$\mathcal{L}(\alpha,\rho,\gamma^L,\gamma^H,\pi|Y) = \sum_{i=1}^N \ln L_i(\alpha,\rho,\gamma^L,\gamma^H,\pi|y_i), \tag{3}$$

where

$$\begin{aligned} & L_i(\alpha,\rho,\gamma^L,\gamma^H,\pi|y_i) \\ = \;& \pi \Pr(y_{i0}|\alpha,\rho,\gamma^L) \prod_{t=1}^T \Pr(y_{it}|y_{i,t-1},\gamma^L) + (1-\pi)\Pr(y_{i0}|\alpha,\rho,\gamma^H) \prod_{t=1}^T \Pr(y_{it}|y_{i,t-1},\gamma^H), \end{aligned}$$

whereas $\Pr(y_{i0}|\alpha, \rho, c)$ is given by (2) while

$$\Pr(y_{it}|y_{i,t-1}, c) = [\Lambda(\alpha + \rho y_{i,t-1} + c)]^{y_{it}}[1 - \Lambda(\alpha + \rho y_{i,t-1} + c)]^{1-y_{it}} \tag{4}$$

for $c = \gamma^L$ and $\gamma^H$.

Show that, for any constant $\xi$, we have $\mathcal{L}(\alpha, \rho, \gamma^L, \gamma^H, \pi|Y) = \mathcal{L}(\alpha - \xi, \rho, \gamma^L + \xi, \gamma^H + \xi, \pi|Y)$.

4. The previous question suggests that $\alpha$, $\gamma^L$, and $\gamma^H$ are not separately identified. We normalize $\gamma^L$ and $\gamma^H$ so that $E[c_i] = 0$ or $\pi\gamma^L + (1-\pi)\gamma^H = 0$ by reparametrizing $\gamma^H$ as $\gamma^H(\gamma^L, \pi) = -\pi\gamma^L/(1-\pi)$. Furthermore, since $\pi$ takes the value between 0 and 1, we reparametrize $\pi$ as $\pi(\tau) = |\tau|/(1+|\tau|)$ where $\tau \in (-\infty, \infty)$ so that $\pi \to 0$ as $\tau \to 0$ while $\pi \to 1$ as $\tau \to \infty$ or $-\infty$ .

   The parameter vector to be estimated is $\theta = (\alpha, \rho, \gamma^L, \tau)$. Write a function m-file that computes the negative value of the loglikelihood given the value of $\theta$ and the data matrix $Y$. Estimate the parameter $\theta$ for the two data set and compute the standard errors of $(\alpha, \rho, \gamma^L, \pi)$.[1]

5. Now, suppose a researcher thought that the initial observations $y_{i0}$'s are independent of $c_i$'s. If $y_{i0}$'s are independent of $c_i$, then $\Pr(y_{i0}|c_i, \alpha, \rho) = \Pr(y_{i0}|\alpha, \rho)$, and the conditional likelihood is given by

$$\sum_{i=1}^{N} \ln\left(\pi \prod_{t=1}^{T} \Pr(y_{it}|y_{i,t-1}, \gamma^L) + (1-\pi)\prod_{t=1}^{T} \Pr(y_{it}|y_{i,t-1}, \gamma^H)\right). \tag{5}$$

   Estimate the parameter $\theta$ by maximizing the above misspecified conditional likelihood for the two data sets and compare the estimated values of $\rho$. Briefly explain your intuition on why the data set with $(T, N) = (10, 4000)$ gives an "better" estimate than the data set $(T, N) = (2, 20000)$?

[1]To compute the standard errors, you can use the delta method. Alternatively, you can write a function m-file that takes $(\alpha, \rho, \gamma^L, \pi)$, rather than $(\alpha, \rho, \gamma^L, \tau)$, as its input and generate the vector of the negative value of the individual loglikelihood as its output, then use the "OPG.m" to compute the outer-products-of-gradients estimator.

6. Suppose a researcher thought that there is no permanent unobserved heterogeneity so that $c_i = \gamma^L = \gamma^H = 0$ for all $i$'s. In this case, the loglikelihood function is given by

$$\sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} \ln \Lambda(\alpha + \rho y_{i,t-1}) + (1 - y_{it}) \ln[1 - \Lambda(\alpha + \rho y_{i,t-1})]. \tag{6}$$

Estimate the parameters $\alpha$ and $\rho$ for the two data sets and compute the standard errors of $(\alpha, \rho)$. Denote its estimate by $(\tilde{\alpha}, \tilde{\rho})$. Especially for the data set $(T, N) = (10, 4000)$, briefly explain your intuition why the value of $\tilde{\rho}$ is larger than the value of $\rho$.

7. Construct a data set from the subset of observations $i$'s in the data set $(T, N) = (10, 4000)$ with $\sum_{t=1}^{T} y_{it} = 3$. Estimate the parameter $\alpha$ and $\rho$ as in Question 6 with this data set. Similarly, estimate $\rho$ for the subset of the data set with $\sum_{t=1}^{T} y_{it} = 4$, or 5, or 6. Are these estimates of $\rho$'s conditioned on the statistic $\sum_{t=1}^{T} y_{it}$ closer to the true value 0.5 than the estimate $\tilde{\rho}$ from Question 6? Briefly explain your intuition on this result.

8. In the context of export example, we are interested in the dynamic effect of "counterfactual" policy of export subsides on a fraction of exporters. Here $y_{it} = 1$ if the firm is exporting and $y_{it} = 0$ if it is not exporting.

   Suppose that a government unexpectedly and permanently introduces export subsidies at $t = 0$ and, as a result, the value of $\alpha$ increases by one unit. Analyze how a fraction of exporters change over time after the introduction of export subsidies by generating "counterfactual" data sets under different estimated parameter values as follows.

   (a) First, we compute a fraction of exporters under the estimated model by using $(\hat{\alpha}, \hat{\rho}, \hat{\gamma}^L, \hat{\gamma}^H, \hat{\pi})$ from Question 4 for the data set with $(T, N) = (10, 4000)$. Repeat Steps 1-4 in Question 2 but using $(\hat{\alpha}, \hat{\rho}, \hat{\gamma}^L, \hat{\gamma}^H, \hat{\pi})$ in place of their true values to generate $(c_i, y_{i0})$ for $(T, N) = (10, 50000)$. Note that we use a large number of simulated observations $N = 50000$ to reduce the prediction noise due to sampling variability. Compute a fraction of exporters predicted by the estimated model for $t = 1, ..., 10$.

   (b) Second, use $(\hat{\alpha}, \hat{\rho}, \hat{\gamma}^L, \hat{\gamma}^H, \hat{\pi})$ from Question 5 for the data set with $(T, N) = (10, 4000)$ to generate a counterfactual fraction of exporters set as follows. For each of $i = 1, ...50000$, starting from $(c_i, y_{i0})$ computed in the previous question (a), repeat Step 3 under the parameter values of $(\hat{\alpha}+1, \hat{\rho}, \hat{\gamma}^L, \hat{\gamma}^H, \hat{\pi})$ to simulate a counterfactual data set $\{\{y_{it}\}_{t=1}^T\}_{i=1}^{50000}$ under export subsidies. Note that the parameter $\hat{\alpha}$ is replaced with $\hat{\alpha}+1$ to reflect the counterfactual value under export subsidies. Compute a counterfactual fraction of exporters for $t = 1, ..., 10$.

(c) Third, use the misspecified model (6) with $(\tilde{\alpha}, \tilde{\rho})$ from Question 6 to generate a predicted counterfactual data set as follows. (i) Repeat Steps 1-2 in Question 2 but using $(\tilde{\alpha}, \tilde{\rho})$ with $c_i = \gamma^L = \gamma^H = 0$ for all $i$'s to generate $(c_i, y_{i0})$ for $N = 50000$. (ii) For each of $i = 1, ...N$, starting from $y_{i0}$, repeat Step 3 under the counterfactual parameter values of $\tilde{\alpha}+1$ to simulate a counterfactual data set $\{\{y_{it}\}_{t=1}^T\}_{i=1}^N$ under export subsidies. Compute a counterfactual fraction of exporters for $t = 1, ..., 10$ under the misspecified model.

9. (optional) Make a table with four columns: (1) an estimated fraction of exporters for $t = 1, ..., 10$ simulated in (a), (2) an estimated counterfactual fraction of exporters simulated in (b), (3) a estimated counterfactual fraction of exporters simulated by the misspecified model in (c), (4) an actual fraction of exporters in the original $(T, N) = (10, 4000)$ data set for $t = 1, ..., 10$. Also, plot (1)-(4) by taking time as x-axis and actual/counterfactual fraction of exporters as y-axis. Use 'plot' command (type 'doc plot' in command line to know how to use 'plot' command).

   Briefly explain your intuition on why the counterfactual prediction under the misspecified model in (c) is different from the counterfactual prediction under the correctly specified model in (b).

10. (optional) Denote the estimated variance-covariance matrix of the MLE, $(\hat{\alpha}, \hat{\rho}, \hat{\gamma}^L, \hat{\tau})$, from Question 4, by $\hat{\Sigma}$ so that the estimated distribution of $(\hat{\alpha}, \hat{\rho}, \hat{\gamma}^L, \hat{\tau})$ is given by $N(0, \hat{\Sigma})$. Discuss how to simulate 90 percent confidence intervals for an estimated counterfactual fraction of exporters in (b) by repeatedly simulating a draw from $N(0, \hat{\Sigma})$. (You do not necessarily need

to write a code for this. Explain how to simulate confidence intervals.)

11. (optional) Read "HW2_data" into Matlab, which contains $N \times (T+1)$ panel data for firm's discrete export and import decisions, where there are $N = 718$ firms across $(T+1) = 7$ years. See "colheaders" for the variables contained in the data set. The first column contains "Plant ID" while the second column contains "Year." In this exercise, we only use the variable in the third column, "$d_e$", which takes the value of one if a firm exports and zero if it does not export. Repeat Questions 4-9 with this data set and generate the graph of actual/counterfactual fraction of exporters over seven years.

# 2 Two step estimator (Section 6 of Newey and McFadden, 1994)

Consider a parametric model specified by the two set of moment conditions:

$$E[m(Z_i;\delta^*)] = 0 \quad \text{and} \quad E[g(Z_i;\beta^*,\delta^*)] = 0,$$

where $\delta^*$ and $\beta^*$ are true parameters; $m(z;\delta)$ is a vector of moment functions with the same dimension as $\delta$ while $g(z;\beta,\delta)$ has the same dimension as $\beta$; therefore, $\delta$ and $\beta$ are just identified from the moment conditions. Given the randomly sampled data set $\{z_i\}_{i=1}^n$ from the model, suppose that we estimate $\delta$ and $\beta$ using two-step estimator as follows. In the first step, we estimate $\delta$ by

$$n^{-1}\sum_{i=1}^{n} m(z_i;\hat{\delta}) = 0.$$

In the second step, **given** $\hat{\delta}$, we estimate $\beta$ by

$$n^{-1}\sum_{i=1}^{n} g(z_i;\hat{\beta};\hat{\delta}) = 0.$$

1. Derive the asymptotic distribution of $\hat{\beta}$.

**Hint:**

$$\begin{pmatrix} G_\beta & G_\delta \\ 0 & M_\delta \end{pmatrix}^{-1} = \begin{pmatrix} G_\beta^{-1} & -G_\beta^{-1} G_\delta M_\delta^{-1} \\ 0 & M_\delta^{-1} \end{pmatrix}$$

2. Consider the following model:

$$y_i = \alpha + \beta d_i + u_i \tag{7}$$

$$d_i = 1(z_i'\gamma + v_i > 0) \tag{8}$$

where we are interested in estimating $\beta$, which represents the effect of discrete variable $d_i$ on outcome variable $y_i$. We assume that $(u_i, v_i)$ is independent of $z_i$ and jointly normally distributed as

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \overset{iid}{\sim} N(0, \Sigma), \quad \text{where } \Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}.$$

For your reference, $u_i|v_i \overset{iid}{\sim} N(\rho\frac{\sigma_u}{\sigma_v}v_i, (1-\rho^2)\sigma_u)$ and $v_i|u_i \overset{iid}{\sim} N(\rho\frac{\sigma_v}{\sigma_u}u_i, (1-\rho^2)\sigma_v)$, where $\rho = \frac{\sigma_{uv}}{\sigma_u\sigma_v}$. When $u \sim N(0,1)$, we have $E[u|u > -c] = \frac{\phi(c)}{\Phi(c)}$ and $E[u|u < -c] = -\frac{\phi(c)}{1-\Phi(c)}$.[2] **We assume that $\sigma_v = 1$.**

(a) We may write the probability of observing $(y_i, d_i)$ conditional on $z_i$ as $\Pr(y_i, d_i|z_i) = d_i \Pr(y_i, d_i = 1|z_i) + (1-d_i)\Pr(y_i, d_i = 0|z_i) = d_i \Pr(u_i = y_i - (\alpha+\beta), z_i'\gamma + v_i \geq 0|z_i) + (1-d_i)\Pr(u_i = y_i - \alpha, z_i'\gamma + v_i < 0|z_i)$. Derive the log-likelihood function of observing the randomly sampled data $\{y_i, d_i\}_{i=1}^n$ conditional on $\{z_i\}_{i=1}^n$ and define the maximum likelihood estimator of the model parameter $\theta = (\alpha, \beta, \sigma_u^2, \sigma_{uv})'$.

(b) Derive the expression for $E[y_i|d_i = 1, z_i]$ and $E[y_i|d_i = 0, z_i]$ in terms of the model parameters.

(c) Define $\lambda(d_i, z_i; \gamma) := d_i\frac{\phi(z_i'\gamma)}{\Phi(z_i'\gamma)} - (1-d_i)\frac{\phi(z_i'\gamma)}{1-\Phi(z_i'\gamma)}$. Prove that the following two-step procedure gives a consistent estimator of $(\alpha, \beta, \sigma_{uv})$: (i) estimate $\gamma$ by $\hat{\gamma} = \arg\max_\gamma \sum_{i=1}^n d_i \ln \Phi(z_i'\gamma) + (1-d_i)(1-\Phi(z_i'\gamma))$, (ii) given $\hat{\gamma}$, estimate $(\alpha, \beta, \sigma_{uv})$ by regressing $y_i$ on constant, $d_i$,

[2] This follows from $E[u|u > -c] = \int_{-c}^{\infty} u \frac{\phi(u)}{1-\Phi(-c)} du = \frac{1}{1-\Phi(-c)}\int_{-c}^{\infty} -\phi'(u) du = \frac{\phi(-c)}{1-\Phi(-c)} = \frac{\phi(c)}{\Phi(c)}$, where the third equality uses $\phi'(u) = -u\phi(u)$.

and $\lambda(d_i, z_i; \hat{\gamma})$.

(d) Derive the asymptotic distribution of the above two-step estimator. To save your time, you may refer to your result in Question 1 above but be explicit about the expression for $m(\cdot)$ and $g(\cdot)$ functions.