ECONOMETRICS II (ECO 2401S)
University of Toronto. Department of Economics. Winter 2014
Instructor: Victor Aguirregabiria

SOLUTION TO FINAL EXAM
Monday, April 14, 2014. From 9:00am-12:00pm (3 hours)

INSTRUCTIONS: This is a closed-book exam. No study aids, including calculators, are allowed. The exam consists of five sets of questions. Try to answer all the questions.

TOTAL MARKS = 100

PROBLEM 1 (25 points). Consider the Censored Linear Regression Model (CLRM), $Y = \max\{X\ \beta + \varepsilon\ ;\ 0\}$ where ε is independent of X and it has a pdf $f(.)$ and a cdf $F(.)$ that is strictly monotonically increasing over the real line.

(a) [5 points] Obtain the expression of the selection bias function $s(X\ \beta) \equiv E(\max\{X\beta + \varepsilon; 0\} \mid X) - X\beta$.

ANSWER: For notational simplicity, let $z \equiv X\beta$. Then, $s(z) = E(\max\{\varepsilon; -z\}) = \Pr(\varepsilon > -z)\ E(\varepsilon \mid \varepsilon > -z) + \Pr(\varepsilon < -z)\ (-z)$, such that:

$$\begin{aligned} s(z) &= [1 - F(-z)]\ \int_{-z}^{+\infty} \varepsilon \frac{f(\varepsilon)}{1 - F(-z)} d\varepsilon + F(-z)\ (-z) \\ &= -z\ F(-z) + \int_{-z}^{+\infty} \varepsilon\ f(\varepsilon) d\varepsilon \end{aligned}$$

(b) [5 points] Show that the selection bias function $s(X\ \beta)$ is a strictly decreasing.

ANSWER: Given the conditions on the CDF F, the sample selection function $s(z)$ is continuously differentiable. The derivative function of $s(z)$ is:

$$s'(z) = -F(-z) + z\ f(-z) - z\ f(-z) = -F(-z) < 0$$

(c) [5 points] Suppose that this CLRM is such that $X\beta = \beta_0 + \beta_1 X_1$, where $X_2 \in \{0, 1\}$ is a binary variable. Obtain the expression of the parameter $\Delta \equiv E(Y|X_1 = 1) - E(Y|X_1 = 0)$ in terms of β_1 and $s(.)$.

ANSWER: By definition of the selection function, we have that $E(Y|X) = X\beta + s(X\beta)$. Then, in this specific model we have that:

$$\begin{aligned} \Delta &= E(Y|X_1 = 1) - E(Y|X_1 = 0) \\ &= [\beta_0 + \beta_1 + s(\beta_0 + \beta_1)] - [\beta_0 + s(\beta_0)] \\ &= \beta_1 + s(\beta_0 + \beta_1) - s(\beta_0) \end{aligned}$$

(d) [5 points] In the model of question (c), show that the OLS estimator of β_1 in the linear regression $Y = \beta_0 + \beta_1 \ X_1 + e$ (that ignores the selection problem) is a consistent estimator of Δ.

ANSWER: The OLS estimator of β_1 is $\sum_{i=1}^{n}(x_{1i} - \overline{x}_1)y_i / \sum_{i=1}^{n}(x_{1i} - \overline{x}_1)^2$, that by the LLN converges in probability to $E((X_1 - E(X_1))Y) \ / \ V(X_1)$. Given that X_1 is binary, we have that: (1) $E(X_1) = p$ where $p \equiv \Pr(X_1 = 1)$; (2) $V(X_1) = p(1-p)$; and (3) $E((X_1 - E(X_1))Y) = p \ (1-p) \ E(Y|X_1 = 1) - (1-p) \ p \ E(Y|X_1 = 0)$. Therefore, $E((X_1 - E(X_1))Y) \ / \ V(X_1) = E(Y|X_1 = 1) - \ E(Y|X_1 = 0) = \Delta$.

(e) [5 points] Using the expressions that you have derived in (c) and (d), obtain the sign of the bias of the OLS estimator of β_1 when $\beta_1 > 0$, and when $\beta_1 < 0$.

ANSWER: Given that $s(.)$ is strictly decreasing, we have that: (1) if $\beta_1 > 0$, then $s(\beta_0 + \beta_1) - s(\beta_0) < 0$ and $\Delta < \beta_1$, i.e., the OLS estimator provides a downward bias estimate of the true β_1; and (2) if $\beta_1 < 0$, then $s(\beta_0 + \beta_1) - s(\beta_0) > 0$ and $\Delta > \beta_1$, i.e., the OLS estimator provides an upward bias estimate of the true β_1. In both cases we have that $|\Delta| < |\beta_1|$, i.e., the OLS estimator implies an attenuation bias.

PROBLEM 2 (20 points). Let θ be a vector $S \times 1$ parameters, and let β be a vector of $R \times 1$ parameters, where $R > S$. Suppose that an econometric model implies the following relationship between these two vector of parameters $\beta = A \ \theta$, where A is a $R \times S$ matrix of constants that is full column rank. Let $\widehat{\beta}$ be a root-n consistent and asymptotically normal estimator of β, such that $\sqrt{n}(\widehat{\beta} - \beta) \rightarrow_d N(0, \mathbf{V}_\beta)$, and let $\widehat{\mathbf{V}}_\beta$ be a root-n consistent estimator of $\mathbf{V}_\beta$.

(a) [5 points] Define the Minimum Distance Estimator (MDE) of θ when the restrictions $\beta = A \ \theta$ are equally weighted.

ANSWER:

$$\widehat{\theta}_{EMD} = \arg\min_{\theta} \left[\widehat{\beta} - \mathbf{A} \ \theta\right]' \left[\widehat{\beta} - \mathbf{A} \ \theta\right]$$

This a quadratic criterion function that is globally convex in θ. The estimator is characterized by first order conditions of optimality $-2\mathbf{A}' \left[\widehat{\beta} - \mathbf{A} \ \widehat{\theta}_{EMD}\right] = 0$, and solving for $\widehat{\theta}_{EMD}$ we get:

$$\widehat{\theta}_{EMD} = \left[\mathbf{A}'\mathbf{A}\right]^{-1} \mathbf{A}'\widehat{\beta}$$

(b) [5 points] Define the MDE of θ when the restrictions are optimally weighted.

ANSWER:

$$\widehat{\theta}_{EMD} = \arg\min_{\theta} \left[\widehat{\beta} - \mathbf{A} \ \theta\right]' \widehat{\mathbf{V}}_\beta^{-1} \left[\widehat{\beta} - \mathbf{A} \ \theta\right]$$

This a quadratic criterion function that is globally convex in θ. The estimator is characterized by first order conditions of optimality $-2\mathbf{A}'\widehat{\mathbf{V}}_{\beta}^{-1}\left[\widehat{\beta}-\mathbf{A}\ \widehat{\theta}_{OMD}\right]=0$, and solving for $\widehat{\theta}_{OMD}$ we get:

$$\widehat{\theta}_{OMD}=\left[\mathbf{A}'\ \widehat{\mathbf{V}}_{\beta}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}'\widehat{\mathbf{V}}_{\beta}^{-1}\widehat{\beta}$$

(c) [10 points] Derive the asymptotic variance of the estimator of θ, both for the Equally Weighted and for the Optimally Weighted MDE.

ANSWER: For the EMD

$$\begin{aligned}\sqrt{n}(\widehat{\theta}_{EMD}-\theta) &= \sqrt{n}\left[\mathbf{A}'\mathbf{A}\right]^{-1}\mathbf{A}'\widehat{\beta}-\sqrt{n}\left[\mathbf{A}'\mathbf{A}\right]^{-1}\mathbf{A}'\mathbf{A}\theta \\ &= \sqrt{n}\left[\mathbf{A}'\mathbf{A}\right]^{-1}\mathbf{A}'\widehat{\beta}-\sqrt{n}\left[\mathbf{A}'\mathbf{A}\right]^{-1}\mathbf{A}'\beta \\ &= \left[\mathbf{A}'\mathbf{A}\right]^{-1}\mathbf{A}'\sqrt{n}(\widehat{\beta}-\beta)\end{aligned}$$

Therefore,

$$\sqrt{n}(\widehat{\theta}_{EMD}-\theta)\rightarrow_d\ N(0,\left[\mathbf{A}'\mathbf{A}\right]^{-1}\mathbf{A}'\mathbf{V}_{\beta}\mathbf{A}\left[\mathbf{A}'\mathbf{A}\right]^{-1})$$

such that

$$Avar(\widehat{\theta}_{EMD})=\left[\mathbf{A}'\mathbf{A}\right]^{-1}\mathbf{A}'\mathbf{V}_{\beta}\mathbf{A}\left[\mathbf{A}'\mathbf{A}\right]^{-1}$$

For the OMD

$$\begin{aligned}\sqrt{n}(\widehat{\theta}_{OMD}-\theta) &= \sqrt{n}\left[\mathbf{A}'\ \widehat{\mathbf{V}}_{\beta}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}'\widehat{\mathbf{V}}_{\beta}^{-1}\widehat{\beta}-\sqrt{n}\left[\mathbf{A}'\widehat{\mathbf{V}}_{\beta}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}'\widehat{\mathbf{V}}_{\beta}^{-1}\mathbf{A}\theta \\ &= \sqrt{n}\left[\mathbf{A}'\widehat{\mathbf{V}}_{\beta}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}'\widehat{\mathbf{V}}_{\beta}^{-1}\widehat{\beta}-\sqrt{n}\left[\mathbf{A}'\widehat{\mathbf{V}}_{\beta}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}'\widehat{\mathbf{V}}_{\beta}^{-1}\beta \\ &= \left[\mathbf{A}'\widehat{\mathbf{V}}_{\beta}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}'\widehat{\mathbf{V}}_{\beta}^{-1}\sqrt{n}(\widehat{\beta}-\beta)\end{aligned}$$

Therefore, by the consistency of $\widehat{\mathbf{V}}_{\beta}^{-1}$ (Slutsky's theorem):

$$\begin{aligned}\sqrt{n}(\widehat{\theta}_{OMD}-\theta) &\rightarrow\ _d\ N(0,\left[\mathbf{A}'\mathbf{V}_{\beta}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}'\mathbf{V}_{\beta}\mathbf{A}\left[\mathbf{A}'\mathbf{V}_{\beta}^{-1}\mathbf{A}\right]^{-1}) \\ &= N(0,\left[\mathbf{A}'\mathbf{V}_{\beta}^{-1}\mathbf{A}\right]^{-1})\end{aligned}$$

such that

$$Avar(\widehat{\theta}_{OMD})=\left[\mathbf{A}'\mathbf{V}_{\beta}^{-1}\mathbf{A}\right]^{-1}$$

PROBLEM 3 (30 points). Consider a random coefficients multinomial choice model with $J+1$ choice alternatives $\{0,1,...,J\}$. The "utility" of alternative $j=0$ is normalized to zero. The utility of alternative choice $j>0$ for individual i is $U_{ij}=\alpha+\beta_i\ Z_j+\varepsilon_{ij}$, where: α is a parameter; Z_j is an observable attribute of alternative j; β_i is a random coefficient with that is i.i.d. over individuals with a normal distribution $N(\mu_\beta,\sigma_\beta^2)$; and ε_{ij} is an unobservable that is i.i.d. over i and over j with an extreme value type 1 distribution. The researcher observes product attributes $\{Z_1,\ Z_2,\ ...,\ Z_J\}$ and a random

sample of n individuals with their optimal choices $\{y_i : i = 1, 2, ..., n\}$ with $y_i \in \{0, 1, ..., J\}$. We are interested in using this sample to estimate the parameters of the $\theta = \{\alpha, \mu_\beta, \sigma_\beta\}$.

(a) [5 points] Write the expression of the choice probability $P_j(\theta) \equiv \Pr(Y = j|\theta)$ in terms of the primitives of the model.

ANSWER: We can write $\beta_i = \mu_\beta + \sigma_\beta \ v_i$, where v_i is *iid* standard normal with pdf $\phi(.)$. Then:

$$P_j(\theta) = \int \frac{\exp\left\{\alpha + \left[\mu_\beta + \sigma_\beta v_i\right] \ Z_j\right\}}{1 + \sum_{k=1}^{J} \exp\left\{\alpha + \left[\mu_\beta + \sigma_\beta v_i\right] \ Z_k\right\}} \phi(v_i) \ dv_i$$

(b) [5 points] Write the log-likelihood function of this model and data.

ANSWER:

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{n} \sum_{j=0}^{J} 1\{y_i = j\} \ \ln P_j(\theta) \\ &= \sum_{j=0}^{J} n_j \ \ln P_j(\theta) \end{aligned}$$

where $n_j \equiv \sum_{i=1}^{n} 1\{y_i = j\}$.

(c) [5 points] Show that if we do not impose the model restrictions on the choice probabilities, the MLE of P_j is the frequency estimator.

ANSWER: Without restrictions on the choice probabilities, we have a multinomial model with J free parameters/probabilities $\mathbf{P} = \{P_j : j = 1, 2, ..., J\}$. The log likelihood function is:

$$\begin{aligned} l(\mathbf{P}) &= \sum_{i=1}^{n} \sum_{j=1}^{J} 1\{y_i = j\} \ \ln P_j + 1\{y_i = 0\} \ \ln(1 - P_1 - ... - P_J) \\ &= \sum_{j=1}^{J} n_j \ \ln P_j + n_0 \ \ln(1 - P_1 - ... - P_J) \end{aligned}$$

The likelihood equations are:

$$\frac{\partial l(\mathbf{P})}{\partial P_j} = \frac{n_j}{\widehat{P}_j} - \frac{n_0}{\widehat{P}_0} = 0$$

Or $n_j \ \widehat{P}_0 = n_0 \ \widehat{P}_j$. Summing this expression over $j > 0$, we have that $(n - n_0) \ \widehat{P}_0 = n_0(1 - \widehat{P}_0)$, and solving for $\widehat{P}_0$, we get $\widehat{P}_0 = n_0/n$. Plugging this expression into the likelihood equations we get that $\widehat{P}_j = n_j/n$, and this is the frequency estimator.

(d) [5 points] Propose a simulator of the choice probability $P_j(\theta)$. Define the simulated log-likelihood function of this model, and the simulated MLE.

ANSWER: Let $\{v^r : r = 1, 2, ...R\}$ be R independent random draws from a standard normal distribution. Then, the simulator of $P_j(\theta)$ is:

$$\widetilde{P}_j^R(\theta) = \frac{1}{R}\sum_{r=1}^{R} \frac{\exp\left\{\alpha + \left[\mu_\beta + \sigma_\beta\ v^r\right]\ Z_j\right\}}{1 + \sum_{k=1}^{J} \exp\left\{\alpha + \left[\mu_\beta + \sigma_\beta\ v^r\right]\ Z_k\right\}}$$

The simulated log-likelihood function is:

$$\widetilde{l}^R(\theta) = \sum_{j=0}^{J} n_j\ \ln \widetilde{P}_j^R(\theta)$$

The simulated MLE is the value of θ that maximizes $\widetilde{l}^R(\theta)$.

(e) [5 points] Obtain the simulated likelihood equations with respect to α, μ_β, and σ_β.

ANSWER: The likelihood equations are $\sum_{j=0}^{J} n_j\ \dfrac{\partial \widetilde{P}_j^R(\theta)}{\partial\theta}\ \dfrac{1}{\widetilde{P}_j^R(\theta)} = 0$, where

$$\frac{\partial \widetilde{P}_j^R(\theta)}{\partial\theta} = \left(\frac{\partial \widetilde{P}_j^R(\theta)}{\partial\alpha}, \frac{\partial \widetilde{P}_j^R(\theta)}{\partial\mu_\beta}, \frac{\partial \widetilde{P}_j^R(\theta)}{\partial\sigma_\beta}\right)'$$

and

$$\begin{aligned}
\frac{\partial \widetilde{P}_j^R(\theta)}{\partial\alpha} &= \frac{1}{R}\textstyle\sum_{r=1}^{R} \Lambda_j(v^r,\theta)\ \left[1 - \Lambda_j(v^r,\theta)\right] \\
\frac{\partial \widetilde{P}_j^R(\theta)}{\partial\mu_\beta} &= \frac{1}{R}\textstyle\sum_{r=1}^{R} \Lambda_j(v^r,\theta)[Z_j - \sum_{k=1}^{J} Z_k\Lambda_k(v^r,\theta)] \\
\frac{\partial \widetilde{P}_j^R(\theta)}{\partial\sigma_\beta} &= \frac{1}{R}\textstyle\sum_{r=1}^{R} v^r\Lambda_j(v^r,\theta)[Z_j - \sum_{k=1}^{J} Z_k\Lambda_k(v^r,\theta)]
\end{aligned}$$

with

$$\Lambda_j(v^r,\theta) \equiv \frac{\exp\left\{\alpha + \left[\mu_\beta + \sigma_\beta\ v^r\right]\ Z_j\right\}}{1 + \sum_{k=1}^{J} \exp\left\{\alpha + \left[\mu_\beta + \sigma_\beta\ v^r\right]\ Z_k\right\}}$$

(f) [5 points] Describe a simulated-based version of the BHHH algorithm to compute the simulated MLE.

ANSWER: We start with an initial vector $\widehat{\theta}_0$. Then, we generate the recursively the sequence $\{\widehat{\theta}_K : K \geq 1\}$ where at iteration $K \geq 1$, we obtain update $\widehat{\theta}_K$ using the formula:

$$\widehat{\theta}_K = \widehat{\theta}_{K-1} + \left[\sum_{i=1}^{n} \frac{\partial \widetilde{l}_i^R(\widehat{\theta}_{K-1})}{\partial\theta}\ \frac{\partial \widetilde{l}_i^R(\widehat{\theta}_{K-1})}{\partial\theta'}\right]^{-1} \left[\sum_{i=1}^{n} \frac{\partial \widetilde{l}_i^R(\widehat{\theta}_{K-1})}{\partial\theta}\right]$$

where $\dfrac{\partial \widetilde{l}_i^R(\theta)}{\partial\theta}$ is the simulated score based with

$$\widetilde{l}_i^R(\theta) = \sum_{j=0}^{J} 1\{y_i = j\}\ \ln \widetilde{P}_j^R(\theta)$$

such that

$$\frac{\partial \widetilde{l}_i^R(\theta)}{\partial \theta} = \sum_{j=0}^{J} 1\{y_i = j\} \ \frac{\partial \widetilde{P}_j^R(\theta)}{\partial \theta} \frac{1}{\widetilde{P}_j^R(\theta)}$$

PROBLEM 4 (5 points). Consider the single-equation econometric model $Y = g(X, \varepsilon; \theta)$ where g is a known real valued function, X is a vector of observable explanatory variables, ε is a vector of unobservables, and θ is a vector of unknown parameters. Function g is continuous in all its arguments and monotonic in ε, though may not be strictly monotonic. The unobservables ε are independent of X and have a continuous and strictly increasing CDF over the Euclidean space. The distribution of ε is unknown to the researcher, i.e., the model is semiparametric. Let $\{y_i, x_i : i = 1, 2, ..., n\}$ be a random sample of Y and X.

(a) [5 points] Propose a consistent estimator of θ.

ANSWER: Under the assumption that the unobservables have medians independent of X, the Least Absolute Deviations (LAD) estimator provides a consistent estimator of θ $Y = g(X, \varepsilon; \theta)$ in a model with non-additively-separable unobservables such as $Y = g(X, \varepsilon; \theta)$. This estimator is defined as:

$$\widehat{\theta}_{LAD} = \arg \min_{\theta} \sum_{i=1}^{n} |y_i - g(x_i, 0; \theta)|$$

PROBLEM 5 (15 points). Consider the static linear panel data model $Y_{it} = \alpha_i + \beta_i$ X_{it} $+u_{it}$, where X_{it} is a explanatory variable that is strictly exogenous with respect to u_{it}.

(a) [5 points] Suppose that β_i and X_{it} are independently distributed, but α_i and X_{it} are not independent. The researcher does not want to make any assumption about the joint distribution of α_i and X_{it}. Propose an estimator of $\overline{\beta} \equiv E(\beta_i)$ that is consistent as $N \to \infty$ and T is fixed.

ANSWER: Define the random variable $v_i \equiv \beta_i - \overline{\beta}$, that by construction has zero mean, and by assumption is independent of X_{it}. We can write the model as $Y_{it} = \alpha_i + \overline{\beta}\ X_{it}\ +e_{it}$, where $e_{it} \equiv u_{it} + v_i\ X_{it}$. Note that for any two periods t and s we have that:

$$\begin{aligned} E(X_{it}\ e_{is}) &= E\left(X_{it}\ \left[u_{is} + v_i\ X_{is}\right]\right) \\ &= E\left(X_{it}\ u_{is}\right) + E\left(v_i\ X_{it}\ X_{is}\right) \\ &= 0 + 0 = 0 \end{aligned}$$

Therefore, X_{it} is strictly exogenous with respect to e_{it} in the model $Y_{it} = \alpha_i + \overline{\beta}\ X_{it}\ +e_{it}$. Under this condition, we know that the OLS estimator in the first differences transformed equation, or the OLS estimator in the within-groups transformed equation are consistent as $N \to \infty$ and T is fixed.

(b) [5 points] Prove that the estimator proposed in (a) is consistent.

ANSWER: For instance, in the first differences transformed equation,

$$Y_{it} - Y_{it-1} = \overline{\beta}\ \ (X_{it} - X_{it-1}) + (e_{it} - e_{it-1})$$

and $E([X_{it} - X_{it-1}]\ [e_{it} - e_{it-1}]) = 0$. Under this condition, and given that there are not incidental parameters, the OLS estimator is consistent.

(c) [5 points] Suppose that both α_i and β_i are NOT independently distributed of X_{it}. The researcher does not want to make any assumption on the joint distribution of (α_i, β_i) and X_{it}. Propose an estimator of $\overline{\beta} \equiv E(\beta_i)$ that is consistent as $N \to \infty$ and T is fixed, and prove its consistency.

ANSWER: OLS in first difference or OLS in within-groups transformed model are inconsistent because now the error term $e_{it} \equiv u_{it} + v_i\ X_{it}$ is correlated with X_{it}. However, we can consider the following transformation of the model. First, we take first differences:

$$Y_{it} - Y_{it-1} = \beta_i\ \ (X_{it} - X_{it-1}) + (u_{it} - u_{it-1})$$

And provided that $(X_{it} - X_{it-1}) \neq 0$ (this has zero probability mass if X_{it} is a continuous random variable) we can divide right-hand-side and left-hand-side by $(X_{it} - X_{it-1})$ to get:

$$\frac{Y_{it} - Y_{it-1}}{X_{it} - X_{it-1}} = \beta_i + \frac{u_{it} - u_{it-1}}{X_{it} - X_{it-1}}$$

or

$$\frac{Y_{it} - Y_{it-1}}{X_{it} - X_{it-1}} = \overline{\beta} + \xi_{it}$$

where $\xi_{it} \equiv v_i + \dfrac{u_{it} - u_{it-1}}{X_{it} - X_{it-1}}$.

It is clear that $E(\xi_{it}) = 0$. Therefore, the sample mean of $\frac{Y_{it} - Y_{it-1}}{X_{it} - X_{it-1}}$ is a consistent estimator of $\overline{\beta}$ as $N \to \infty$ and T is fixed, and prove its consistency. For instance, suppose that $T = 2$:

$$\widehat{\beta} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{Y_{i2} - Y_{i1}}{X_{i2} - X_{i1}} \right)$$

It is clear that by LLN $\widehat{\beta}$ converges in probability to $E\left(\frac{Y_{i2} - Y_{i1}}{X_{i2} - X_{i1}}\right) = E(\overline{\beta} + \xi_{i2}) = \overline{\beta}$.