ECONOMETRICS II (ECO 2401S)
University of Toronto. Department of Economics. Winter 2016
Instructor: Victor Aguirregabiria

FINAL EXAM. Thursday, April 14, 2016. From 9:00am-12:00pm (3 hours)

INSTRUCTIONS:
- **This is a closed-book exam.**
- **No study aids, including calculators, are allowed.**
- **Please, answer all the questions.**

TOTAL MARKS = 100

PROBLEM 1 (40 points). Let w_{it} be the log-wage of worker i at period t. The researcher has a panel dataset $\{w_{it} : i = 1, 2, ...N;\ t = 1, 2, ..., T\}$ where the number of workers N is large, and the number of periods T is small, e.g., $N = 5,000$ and $T = 10$. The researcher postulates the following variance-component model:

$$w_{it} = \beta_t + \alpha_i + u_{it}$$

where β_1, β_2, ..., β_T are parameters; $\mathbb{E}(\alpha_i) = \mathbb{E}(u_{it}) = 0$; $\mathbb{E}(\alpha_i\ u_{it}) = 0$; $var(\alpha_i) = \sigma^2_\alpha$; u_{it} is not serially correlated, it is homoscedastic across individuals, but its variance may vary over time, $var(u_{it}) = \sigma^2_{u,t}$. The main interest of the researcher is the analysis of wage inequality, its persistent and the evolution over time. More specifically, the researcher is interested in estimation of the variance parameters σ^2_α and $\sigma^2_{u,1}$, ..., $\sigma^2_{u,T}$.

(a) [5 points] Propose a root-N consistent estimator of the parameters β_1, β_2, ..., β_T.

ANSWER: Given the assumptions of the model, we have that for any period t, $\mathbb{E}(w_{it}\ |t) = \beta_t$. Therefore, we can estimate β_t using a Method of Moments estimator based on this moment condition. This Method of Moments estimator is:

$$\widehat{\beta}_t = \frac{1}{N} \sum_{i=1}^{N} w_{it}$$

By the LLN, this estimator converges in probability to β_t as N goes to infinity. And by the CLT, $\sqrt{N}(\widehat{\beta}_t - \beta_t)$ converges in distribution to a $N(0, Var(\alpha_i + u_{it}))$, with $Var(\alpha_i + u_{it}) = \sigma^2_\alpha + \sigma^2_{u,t}$.

(b) [15 points] Propose a root-N consistent estimator of the parameters σ^2_α, $\sigma^2_{u,1}$, ..., $\sigma^2_{u,T}$.

ANSWER: Define $\varepsilon_{it} \equiv \alpha_i + u_{it} = w_{it} - \beta_t$. Based on the model assumptions, we have that for any period t,

$$\begin{aligned}\mathbb{E}(\varepsilon_{it}\ \varepsilon_{it}\ |t) &= \sigma^2_\alpha + \sigma^2_{u,t} \\ \mathbb{E}(\varepsilon_{it}\ \varepsilon_{it-1}|t) &= \sigma^2_\alpha\end{aligned}$$

These moment conditions imply,

$$\sigma_\alpha^2 = \frac{1}{T-1}\sum_{t=2}^{T}\mathbb{E}(\varepsilon_{it}\ \varepsilon_{it-1}|t)$$

$$\sigma_{u,t}^2 = \mathbb{E}(\varepsilon_{it}\ \varepsilon_{it}\ |t) - \frac{1}{T-1}\sum_{t=2}^{T}\mathbb{E}(\varepsilon_{it}\ \varepsilon_{it-1}|t)$$

Using these moment conditions and the consistent estimators $\widehat{\beta}_t$ from Question 1a, we can construct consistent Method of Moments estimators of σ_α^2 and $\sigma_{u,t}^2$. That is,

$$\widehat{\sigma}_\alpha^2 = \frac{1}{T-1}\sum_{t=2}^{T}\left[\frac{1}{N}\sum_{i=1}^{N}\left(w_{it}-\widehat{\beta}_t\right)\left(w_{it-1}-\widehat{\beta}_{t-1}\right)\right]$$

$$\sigma_{u,t}^2 = \left[\frac{1}{N}\sum_{i=1}^{N}\left(w_{it}-\widehat{\beta}_t\right)^2\right] - \frac{1}{T-1}\sum_{t=2}^{T}\left[\frac{1}{N}\sum_{i=1}^{N}\left(w_{it}-\widehat{\beta}_t\right)\left(w_{it-1}-\widehat{\beta}_{t-1}\right)\right]$$

Under the condition that the distributions of α_i and u_{it} have finite moments of order four, these estimators are root-N consistent and asymptotically normal.

(c) [5 points] Suppose that u_{it} is serially correlated. Does this correlation affect the consistency of the estimator proposed in Question 1b? Explain.

ANSWER: If u_{it} is serially correlated, then the previous moment conditional become:

$$\mathbb{E}(\varepsilon_{it}\ \varepsilon_{it}\ |t) = \sigma_\alpha^2 + \sigma_{u,t}^2$$

$$\mathbb{E}(\varepsilon_{it}\ \varepsilon_{it-1}|t) = \sigma_\alpha^2 + \mathbb{E}(u_{it}\ u_{it-1}|t)$$

Therefore, the previous estimator of σ_α^2 is inconsistent because it not only captures the variance σ_α^2 but also the covariance $\mathbb{E}(u_{it}\ u_{it-1}|t)$. For instance, if the serial correlation is positive such that $\mathbb{E}(u_{it}\ u_{it-1}|t) > 0$, then the previous estimator $\widehat{\sigma}_\alpha^2$ over-estimates the true σ_α^2, i.e., it over-estimates the time invariant component of wage-inequality. Similarly, this bias in the estimation of $\widehat{\sigma}_\alpha^2$ implies also a bias in the estimation of $\sigma_{u,t}^2$. More specifically, the previous estimator of $\sigma_{u,t}^2$ is a consistent estimator of $\sigma_{u,t}^2 - \mathbb{E}(u_{it}\ u_{it-1}|t)$. Under positive correlation, $\widehat{\sigma}_{u,t}^2$ under-estimates the true $\sigma_{u,t}^2$.

(d) [15 points] Describe a test of the null hypothesis $E(u_{it}\ u_{i,t-1}) = 0$.

ANSWER: Due to the incidental parameters problem, we cannot obtain root-N consistent estimators of the unobservables $u_{it} = w_{it} - \beta_t - \alpha_i$. Therefore, our test of the null hypothesis $E(u_{it}\ u_{i,t-1}) = 0$ cannot be based on residuals for u_{it}. However, we can obtain root-N consistent estimates for $\Delta u_{it} \equiv u_{it} - u_{it-1} = \Delta w_{it} - \beta_t + \beta_{t-1}$. The Arellano-Bond test of serial correlation is based on the residuals:

$$\widehat{\Delta u_{it}} = \Delta w_{it} - \widehat{\beta}_t + \widehat{\beta}_{t-1}$$

Under the null hypothesis $E(u_{it}\ u_{i,t-1}) = 0$, we have that:

$$\mathbb{E}\left(\triangle u_{it}\ \ \triangle u_{it-2}\right) = 0$$

Therefore, we can (indirectly) test for no-serial correlation in u_{it} by testing for no second-order serial correlation in $\triangle u_{it}$. Let r_{2t} be the auto-covariance of order 2 at period t of $\{\triangle u_{it}\}$: i.e., $r_{2t} \equiv E\left(\triangle u_{it} \ \triangle u_{it-2}\right)$. And its sample counterpart:

$$\widehat{r}_{2t} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\Delta u}_{it} \ \widehat{\Delta u}_{it-2}$$

We can obtain $\widehat{r}_{2t}$ for any $t \in \{4, 5, ..., T\}$. Note that we need $T \geq 4$. Let $r_2 \equiv \sum_{t=4}^{T} r_{2t}$, and let $\widehat{r}_2$ be its sample counterpart. Arellano & Bond (1991) prove that under the null hypothesis $\widehat{r}_2$ is root-N asymptotically normal with mean zero, and they derive the expression for the asymptotic variance $Var(\widehat{r}_2)$. Then, under H_0: $r_2 = 0$.

$$\widehat{m}_2 \equiv \frac{\widehat{r}_2}{se(\widehat{r}_2)} \underset{a}{\sim} N(0,1)$$

PROBLEM 2 (30 points). Consider the Binary choice model, $Y_i = 1\{X_i'\beta + \alpha \ W_i + \varepsilon_i \geq 0\}$ where $1\{.\}$ is the indicator function, ε_i is independent of X_i but it may be correlated with W_i.

(a) [20 points] Describe the Rivers-Vuong approach to estimate consistently β and α (up to scale) and to test for the exogeneity of W_i. Make the assumptions of this approach explicit.

ANSWER: Consider the model:

$$\begin{array}{llll} (1) & Y & = & 1\{X'\beta + \alpha \ W + \varepsilon > 0\} \\ (2) & W & = & Z'\delta + u \end{array}$$

where ε and u are independent of X and Z, but $cov(\varepsilon, u) \neq 0$, and therefore ε and W are not independent. Suppose that (ε, u) are jointly normal. Then, we have that:

$$\varepsilon \ = \ \pi \ u + \xi$$

where (a) $\pi = \sigma_{\varepsilon u}/\sigma_u^2$; (b) ξ is normally distribution as $N(0, \sigma_\varepsilon^2\left(1 - \rho^2\right))$ where ρ is the correlation between ε and u; (c) ξ is independent of u; (d) since ε is independent of X and Z, we have that ξ is independent of X, Z, and u, and therefore it is independent of W. Then, we can write the probit model:

$$Y \ = \ 1\{X'\beta + \alpha \ W + \pi \ u + \xi > 0\}$$

And given that ξ is normally distributed and independent of X, W, and u, we have that:

$$\Pr(Y = 1|X, W, u) = \Phi\left(\frac{X'\beta + \alpha \ W + \pi \ u}{\sigma_\xi}\right)$$

We do not know u, but we can obtain a consistent estimate of u as the residual $\hat{u} = Y - Z'\hat{\delta}$.

Rivers and Vuong (1988) propose the following procedure:

Step 1. Estimate the regression of W on Z and obtain the residual $\hat{u}$;

Step 2. Run a probit for Y on X, W and $\hat{u}$.

Using this procedure we obtain consistent estimates of $\dfrac{\beta}{\sigma_\xi}$, $\dfrac{\alpha}{\sigma_\xi}$, and $\dfrac{\pi}{\sigma_\xi}$. Note that $\dfrac{\pi}{\sigma_\xi} \neq 0$ if and only if $cov(\varepsilon, u) \neq 0$. Therefore, a t-test of $H_0 : \dfrac{\pi}{\sigma_\xi} = 0$ is a test of the endogeneity of W.

(b) [10 points] Discuss how to combine the approach by Rivers-Vuong with a Maximum Score approach to obtain a consistent estimator of β and α that relaxes the parametric assumption in the distribution of ε_i.

ANSWER: Suppose that ε is independent of X and Z. Let $g(u)$ be the median of ε conditional on u, and assume that $g(.)$ is a smooth function that is unknown to the researcher. Define $\xi = \varepsilon - g(u)$. Given these assumptions and definitions, we have that $median(\xi|X, W, u) = 0$. Therefore, we have the Binary choice model:

$$Y \quad = \quad 1\{X'\beta + \alpha\ W + g(u) + \xi > 0\}$$

with $median(\xi|X, W, u) = 0$. More precisely, let $b(u) = (u, u^2, ..., u^q)'$ be a polynomial basis in u, and let π be a vector of parameters associated to the polynomial terms such that $g(u)$ is approximated using $b(u)'\pi$. Then, given residuals $\widehat{u}$ we have the model:

$$Y \quad = \quad 1\{X'\beta + \alpha\ W + b(\widehat{u})'\pi + \xi > 0\}$$

The Maximum Score Estimator is the value of (β, α, π) that maximizes the score function:

$$\begin{aligned} S(\beta, \alpha, \pi) \quad = \quad & \textstyle\sum_{i=1}^{n} y_i\ 1\left\{x_i'\beta + \alpha w_i + b(\widehat{u}_i)'\pi \geq 0\right\} \\ & +(1 - y_i)\ 1\left\{x_i'\beta + \alpha w_i + b(\widehat{u}_i)'\pi < 0\right\} \end{aligned}$$

PROBLEM 3 (30 points). Consider the Random Utility Model, $Y_n = \arg\max_{j\in\{0,1,...,J\}}$ $[X_j'\beta + Z_n'\gamma_j + \varepsilon_{nj}]$, where n is the index for individuals/observations, and j is the index for choice alternatives.

(a) [15 points] Describe the Logit model and the Maximum Likelihood estimator of the parameters of the model. Comment on the properties of this model.

ANSWER: In the Logit model ε_{jn} are i.i.d. over (n, j) Type 1 Extreme Value. For any j, we have that the CDF is $F(\varepsilon_j) = \exp\{-\exp\{-\varepsilon_j\}\}$. Under this assumption on the distribution of ε, we have the following form for the Conditional Choice Probabilities (CCPs):

$$P_j(X, Z_n; \theta) \quad = \quad \frac{\exp\{X_j'\beta + Z_n'\gamma_j\}}{\sum_{i=0}^{J} \exp\{X_i'\beta + Z_n'\gamma_i\}}$$

where $\theta = (\beta, \gamma_1, ..., \gamma_J)$ and γ_0 is normalized to zero. The log-likelihood function is:

$$l_N(\theta) = \sum_{n=1}^{N} \sum_{j=\prime}^{J} 1\{y_n = j\}\ \ln\left[\frac{\exp\{X_j'\beta + Z_n'\gamma_j\}}{\sum_{i=0}^{J} \exp\{X_i'\beta + Z_n'\gamma_i\}}\right]$$

This log-likelihood function if globally concave in θ. Furthermore, the gradient and Hessian of this function have simple closed form expressions. Therefore, the numerical computation of the MLE can be implemented in a simple way using Newton's method. In a Logit model, $\dfrac{\partial P_j}{\partial u_j} = P_j\left[1 - P_j\right]$. Taking this into account, we can show that the likelihood equations for this model are: for $\partial l_N(\theta)/\partial\beta = 0$:

$$\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{j=0}^{J} X_j\left[1\{y_n = j\} - P_j(X, Z_n;\theta)\right]\right) = 0$$

And for every $\partial l_N(\theta)/\partial\gamma_j = 0$ with $j = 1, 2, ..., J$:

$$\frac{1}{N}\sum_{n=1}^{N} Z_n\ \left[1\{y_n = j\} - P_j(X, Z_n;\theta)\right] = 0$$

Independence of Irrelevant Alternatives. The logit model imposes the restriction that the ratio between the probabilities of two alternatives, say j and i, depends ONLY on the utilities of these alternatives, and not on utilities of other alternatives:

$$\frac{P_{jn}}{P_{in}} = \frac{\exp\{X_j'\beta + Z_n'\gamma_j\}}{\exp\{X_j'\beta + Z_n'\gamma_i\}}$$

Therefore, if we change the choice set, by adding or/and removing alternatives, the ratios between probabilities should not change. This property can generate unrealistic predictions.

(b) [15 points] Describe a Nested Logit model and a two-step consistent estimator of the parameters of the model. Propose a simple approach to obtain an asymptotically efficient estimator using this two-step estimator.

ANSWER: The Nested Logit was proposed to relax the IIA property of the logit model but keeping its computational convenience. Suppose that the set $\mathcal{J} = \{0, 1, ..., J\}$ of choice alternatives is partitioned into G mutually exclusive groups of alternatives, that we index by g. Let $\mathcal{J}_g$ be the set of alternatives in group g such that: $\mathcal{J} = \bigcup_{g=1}^{G}\mathcal{J}_g$. The idea is that alternatives within a group share some common unobserved features that make them closer substitutes that alternatives in different groups. The key assumption is that the vector of unobservables $\boldsymbol{\varepsilon} = (\varepsilon_0, \varepsilon_1, ..., \varepsilon_J)$ has a Generalized Extreme Value (GEV) distribution:

$$F(\boldsymbol{\varepsilon}) = \exp\left\{-\sum_{g=1}^{G}\left[\sum\nolimits_{j\in\mathcal{J}_g}\exp\left(-\frac{\varepsilon_j}{\sigma_g}\right)\right]^{\frac{\sigma_g}{\delta}}\right\}$$

where δ, σ_1, σ_2, ..., σ_G are positive parameters, with $\delta \leq 1$. Consider the RUM $Y = \arg\max_{j\in\mathcal{J}}\{X_j'\beta + Z_n'\gamma_j + \varepsilon_{jn}\}$ where $\boldsymbol{\varepsilon}_n = (\varepsilon_{0n}, \varepsilon_{1n}, ..., \varepsilon_{Jn})$ has a GEV distribution. The CCPs of this model have the following form:

$$P_j(X, Z_n;\theta) = P_g^{(1)}(X, Z_n;\theta)\ P_{j|g}^{(2)}(X, Z_n;\theta)$$

with

$$P^{(2)}_{j|g}(X, Z_n; \theta) = \frac{\exp\left\{\dfrac{X'_j\beta + Z'_n\gamma_j}{\sigma_g}\right\}}{\sum\limits_{i\in\mathcal{J}_g} \exp\left\{\dfrac{X'_i\beta + Z'_n\gamma_i}{\sigma_g}\right\}}$$

$$P^{(1)}_{g}(X, Z_n; \theta) = \frac{\exp\left\{\dfrac{\sigma_g}{\delta}\ I_{g,n}\right\}}{\sum_{g'=1}^{G} \exp\left\{\dfrac{\sigma_{g'}}{\delta}\ I_{g',n}\right\}}$$

and $I_{g,n}$ are the group inclusive values:

$$I_{g,n} = \ln\left(\sum_{j\in\mathcal{J}_g} \exp\left\{\frac{X'_j\beta + Z'_n\gamma_j}{\sigma_g}\right\}\right)$$

The likelihood function of the model, $l(\theta) = \sum_{n=1}^{N} \ln \Pr(Y_n|X, Z_n, \theta)$ can be written as the sum of two likelihoods: $l^{(1)}(\theta) + l^{(2)}(\theta)$

$$\begin{aligned} l(\theta) &= \sum_{n=1}^{N}\sum_{g=1}^{G} 1\{y_n^{(1)} = g\} \ln P_g^{(1)}(X, Z_n; \theta) \\ &+ \sum_{n=1}^{N}\sum_{j\in\mathcal{J}_{y_n^{(1)}}} 1\{y_n^{(2)} = j\} \ln P^{(2)}_{j|y_n^{(1)}}(X, Z_n; \theta) \end{aligned}$$

where $y_n^{(1)} \in \{1, 2, ..., G\}$ represents the observed group-choice of individual n, and $y_n^{(2)} \in \mathcal{J}_{y_n^{(1)}}$ represents the observed within–group choice of individual n.

Note that $l(\theta) = l^{(1)}(\theta) + l^{(2)}(\theta)$ where: $l^{(1)}(\theta)$ is the **between-group** likelihood function for the choice variable $y_n^{(1)}$; and $l^{(2)}(\theta)$ is the **within-group** likelihood function for the choice variable $y_n^{(2)}$. We can estimate a combination of the parameters in θ by maximizing $l^{(1)}(\theta)$, and other combination of parameters θ by maximizing $l^{(2)}(\theta)$. This two-step procedure is not statistically efficient but it is computationally very convenient because each step consists of a standard MNL estimation (i.e., globally concave likelihood function).

Step 1: Maximization of within-group likelihood function $l^{(2)}(\theta)$ with probabilities:

$$P_{j|g,n} = \frac{\exp\{X_j\ \beta_g + Z_n\ \gamma_{j,g}\}}{\sum_{i\in\mathcal{J}_g} \exp\{X_i\ \beta_g + Z_n\ \gamma_{i,g}\}}$$

where the estimated parameters are: $\beta_g \equiv \dfrac{\beta}{\sigma_g}$ and $\gamma_{j,g} \equiv \dfrac{\gamma_j}{\sigma_g}$.

Step 2: Construct the estimated **inclusive values:**

$$\widehat{I}_{g,n} = \ln\left(\sum_{j\in\mathcal{J}_g} \exp\{X_j\ \widehat{\beta}_g + Z_n\ \widehat{\gamma}_{j,g}\}\right)$$

And maximization of between-group likelihood function $l^{(1)}(\theta)$ with probabilities:

$$P_{g,n} = \frac{\exp\left\{\frac{\sigma_g}{\delta}\ \widehat{I}_{g,n}\right\}}{\sum_{g'=1}^{G} \exp\left\{\frac{\sigma_{g'}}{\delta}\ \widehat{I}_{g',n}\right\}}$$

The estimated parameters are $\frac{\sigma_g}{\delta}$, with one of these parameters normalized to zero within each group.

Given this consistent two-step estimator, $\widehat{\theta}_{2-step}$, we can construct an efficient estimator, and a valid variance-covariance matrix by doing one Newton or BHHH iteration in the estimation of the full likelihood function:

$$\widehat{\theta}_{eff} = \widehat{\theta}_{2-step} - \left[\frac{\partial^2 l(\widehat{\theta}_{2-step})}{\partial\theta\partial\theta'}\right]^{-1} \frac{\partial l(\widehat{\theta}_{2-step})}{\partial\theta}$$