ECONOMETRICS II (ECO 2401S)
University of Toronto. Department of Economics. Winter 2018
Instructor: Victor Aguirregabiria
SOLUTION TO FINAL EXAM
Tuesday, April 10, 2018. 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 (50 points). Consider the random-coefficients regression model, $Y_i = \alpha_i + \beta_i\ D_i$, where subindex i represents an individual, and $D_i \in \{0, 1\}$ is a binary variable. We denote Y as the outcome variable, and D as the treatment variable. The researcher observes a random sample of $\{Y, D\}$ for N individuals, $\{y_i, d_i : i = 1, 2, ..., N\}$. The random variables (α_i, β_i) are unobserved to the researcher, and they have means α^* and β^*, respectively. Random variable β_i represents the treatment effect for individual i, i.e., $\beta_i = \{Y_i|D_i = 1\} - \{Y_i|D_i = 0\}$. Therefore, β^* is the Average Treatment Effect (ATE) over the population of individuals.

Question 1.1 [5 points] Given equation $Y_i = \alpha_i + \beta_i\ D_i$, obtain a representation of this model as a regression equation where the intercept and the slope parameters are constants, i.e., constant-coefficients representation. Describe the error term in this regression equation.

ANSWER. By definition, $\alpha_i = \alpha^* + a_i$, and $\beta_i = \beta^* + b_i$, where a_i and b_i are zero mean random variables. Plugging these expressions in equation $Y_i = \alpha_i + \beta_i\ D_i$, we have that $Y_i = \alpha^* + a_i + [\beta^* + b_i]\ D_i$, or what is equivalent,

$$Y_i = \alpha^* + \beta^*\ D_i + \varepsilon_i, \quad \text{where the error term is } \varepsilon_i \equiv a_i + b_i\ D_i \qquad \blacksquare$$

Question 1.2 [5 points] Using the constant-coefficients representation in Question 1.1, consider the OLS estimator in this equation. Show that the OLS of the slope parameter can be written as $\overline{y}_1 - \overline{y}_0$, where for $d \in \{0, 1\}$, $\overline{y}_d \equiv \sum_{i=1}^N y_i\ 1\{d_i = d\}/\sum_{i=1}^N 1\{d_i = d\}$.

ANSWER. Let $\overline{y}$ and $\overline{d}$ be the sample means of Y and D, respectively. Note that, by definition, we have that: (i) $\sum_{i=1}^N y_i\ d_i = N\ \overline{y}_1\ \overline{d}$; and (ii) $\sum_{i=1}^N d_i d_i = \sum_{i=1}^N d_i d_i = N\ \overline{d}$. he OLS estimator of the slope parameter in the simple regression of Y on D is

$$\widehat{\beta}^*_{OLS} \quad = \quad \frac{\sum_{i=1}^N (y_i - \overline{y})\ (d_i - \overline{d})}{\sum_{i=1}^N (d_i - \overline{d})\ (d_i - \overline{d})} = \frac{\sum_{i=1}^N y_i\ d_i - N\ \overline{y}\ \overline{d}}{\sum_{i=1}^N d_i\ d_i - N\ \overline{d}\ \overline{d}}$$

Taking into account expressions (i) and (ii), we have that:

$$\widehat{\beta}^*_{OLS} \quad = \quad \frac{N\ \overline{y}_1 \overline{d} - N\ \overline{y}\ \overline{d}}{N\ \overline{d} - N\ \overline{d}\ \overline{d}} = \frac{\overline{y}_1 - \overline{y}}{1 - \overline{d}}$$

Now, we can write the sample mean $\overline{y}$ as,

$$\begin{aligned}\overline{y} &= N^{-1}\textstyle\sum_{i=1}^{N} y_i = N^{-1}\left[\textstyle\sum_{i=1}^{N} y_i d_i + \sum_{i=1}^{N} y_i (1-d_i)\right] \\ &= N^{-1}\left[N\ \overline{y}_1 \overline{d} + N\ \overline{y}_0\ (1-\overline{d})\right] \\ &= \overline{y}_1\ \overline{d} + \overline{y}_0\ (1-\overline{d})\end{aligned}$$

Plugging this expression into the equation above for the OLS estimator, we get:

$$\widehat{\beta}^*_{OLS} = \frac{\overline{y}_1 - \overline{y}_1\ \overline{d} - \overline{y}_0\ (1-\overline{d})}{1-\overline{d}} = \overline{y}_1 - \overline{y}_0 \qquad \blacksquare$$

Question 1.3 [5 points] Consider the OLS estimator in Question 1.2. Suppose that D_i is independent of (α_i, β_i). Prove that this OLS is a consistent estimator of $\beta^* \equiv ATE$.

ANSWER. Given a random sample of Y and D, the Law Large Numbers implies and Slutsky's Theorem imply that, if $0 < \overline{d} < 1$:

$$\widehat{\beta}^*_{OLS} = \overline{y}_1 - \overline{y}_0 \to_p \mathbb{E}(Y \mid D=1) - \mathbb{E}(Y \mid D=0)$$

Now, taking into account that $Y = \alpha^* + \beta^*\ D + \varepsilon$, with $\varepsilon \equiv a + b\ D$, and that (a,b) are mean zero and independent of D, we have that:

$$\begin{aligned}\mathbb{E}(Y \mid D=1) &= \alpha^* + \beta^* + \mathbb{E}(a+b \mid D=1) = \alpha^* + \beta^* \\ &\text{and} \\ \mathbb{E}(Y \mid D=0) &= \alpha^* + \mathbb{E}(a \mid D=0) = \alpha^*\end{aligned}$$

Therefore, $\widehat{\beta}^*_{OLS} \to_p \mathbb{E}(Y \mid D=1) - \mathbb{E}(Y \mid D=0) = \beta^*$. $\blacksquare$

Question 1.4 [5 points] Let $\{\widehat{e}_i : i = 1, 2, ..., N\}$ be the residuals from the OLS regression in Question 1.2. Under the assumption of independence between D_i and (α_i, β_i), explain how you can use these residuals to obtain a root-N consistent nonparametric estimator of the whole Cumulative Distribution Function (CDF). of the heterogeneous treatment effects β_i.

[Note for grading: The enunciate of this question does not make explicit the assumption of independence between α_i and $\alpha_i + \beta_i$ that is necessary to answer this specific question]
ANSWER. Under the assumption of independence between D_i and (α_i, β_i), we have that $\alpha_i = \{Y_i | D_i = 0\} \equiv Y_i^0$ and $\alpha_i + \beta_i = \{Y_i | D_i = 1\} \equiv Y_i^1$. Let F_α and $F_{\alpha+\beta}$ and the CDFs of the random variables $Y_i^0 = \alpha_i$ and $Y_i^1 = \alpha_i + \beta_i$, respectively. Using the sub-sample of observations with $d_i = 0$, we can estimate consistently the CDF F_α. Similarly, using the sub-sample of observations with $d_i = 1$, we can estimate consistently the CDF $F_{\alpha+\beta}$. For any constant c, the estimators of $F_\alpha(c)$ and $F_{\alpha+\beta}(c)$ are:

$$\begin{aligned}\widehat{F}_\alpha(c) &= \frac{1}{N_0} \textstyle\sum_{i=1}^{N} 1\{y_i \leq c \ \text{ and } \ d_i = 0\} \\ \widehat{F}_{\alpha+\beta}(c) &= \frac{1}{N_1} \textstyle\sum_{i=1}^{N} 1\{y_i \leq c \ \text{ and } \ d_i = 1\}\end{aligned}$$

where $N_0 \equiv \sum_{i=1}^{N} 1\{d_i = 0\}$, and $N_1 \equiv \sum_{i=1}^{N} 1\{d_i = 1\}$.

Next, we show that we can obtain the CDF F_β in terms of the CDFs F_α and $F_{\alpha+\beta}$. Remember that $\beta_i = Y_i^1 - Y_i^0$. Then, for any constant c:

$$\begin{array}{rcl} F_\beta(c) & = & \Pr\left(Y^1 - Y^0 \leq c\right) = \Pr\left(Y^1 \leq c + Y^0\right) = \Pr\left(\alpha + \beta \leq c + \alpha\right), \\ & & \text{and under independence between } \alpha+\beta \text{ and } \alpha \text{, we have:} \\ & = & \int F_{\alpha+\beta}\left(c+\alpha\right) \ f_\alpha(\alpha) \ d\alpha = \mathbb{E}_\alpha\left[F_{\alpha+\beta}\left(c+\alpha\right)\right] \end{array}$$

where $\mathbb{E}_\alpha\left[.\right]$ represents the expectation over the distribution of α.

Note that $\mathbb{E}_\alpha\left[F_{\alpha+\beta}\left(c+\alpha\right)\right]$ is the same object as $\mathbb{E}_{Y_0}\left[F_{Y_1}\left(c+Y_0\right)\right]$. Therefore, we can construct the following consistent estimator of $F_\beta(c)$.

$$\begin{array}{rcl} \widehat{F}_\beta(c) & = & \widehat{\mathbb{E}}_{Y_0}\left[\widehat{F}_{Y_1}\left(c+Y_0\right)\right] \\ & = & \dfrac{1}{N_0}\sum_{i=1}^N 1\{d_i = 0\} \ \left[\widehat{F}_{Y_1}\left(c+y_i\right)\right] \\ & = & \dfrac{1}{N_0}\sum_{i=1}^N 1\{d_i = 0\} \ \left[\dfrac{1}{N_1}\sum_{j=1}^N 1\{y_j \leq c + y_i \ \text{ and } \ d_j = 1\}\right] \qquad \blacksquare \end{array}$$

Question 1.5 [5 points] Now, consider that the treatment variable D_i is not independent of (α_i, β_i). Show that the OLS estimator of β^* is inconsistent.

ANSWER. A necessary condition for the consistency of the OLS estimator of β^* is that $\mathbb{E}(D \ \varepsilon) = 0$. But we have that,

$$\mathbb{E}\left[D \ \varepsilon\right] = \mathbb{E}\left[D \ \ (a + b \ D)\right] = \Pr(D=1) \ \mathbb{E}\left[a + b \mid D = 1\right]$$

Since D is not independent of (a, b), we have that, in general, the conditional expectation $\mathbb{E}\left[a+b \mid D=1\right]$ is different to the unconditional expectation $\mathbb{E}\left[a+b\right] = 0$. Therefore, $\mathbb{E}\left[D \ \varepsilon\right] \neq 0$ and the OLS estimator is inconsistent. $\blacksquare$

Question 1.6 [5 points] Let Z_i be a random variable with support $Z \subseteq \mathbb{R}$. Note that Z_i can be continuous. Suppose that Z_i satisfies the following conditions: (Relevance) $P_D(z) \equiv \Pr(D_i = 1|Z_i = z)$ varies with $z \in Z$; and (Independence) Z_i is independent of (α_i, β_i). Suppose that you estimate the regression equation (in Question 1.2) by Instrumental Variables (IV) using Z_i as instrument for D_i. Prove that this IV estimator is inconsistent for parameter β^*.

ANSWER. A necessary condition for the consistency of the OLS estimator of β^* is that $\mathbb{E}(Z \ \varepsilon) = 0$. Now, we have that,

$$\mathbb{E}\left[Z \ \varepsilon\right] = \mathbb{E}\left[Z \ \ (a + b \ D)\right] = \mathbb{E}\left[Z \ a\right] + \mathbb{E}\left[Z \ b \ D\right]$$

Since Z is independent of a, we have that $\mathbb{E}\left[Z \ a\right] = \mathbb{E}\left[a\right] = 0$. Then,

$$\begin{array}{rcl} \mathbb{E}\left[Z \ \varepsilon\right] & = & \mathbb{E}_Z\left[Z \ \mathbb{E}\left(b \ D \mid Z\right)\right] = \mathbb{E}_Z\left[Z \ \Pr(D=1|Z) \ \mathbb{E}\left(b \mid Z, D=1\right)\right] \\ & = & \mathbb{E}_Z\left[Z \ P_D(Z) \ \mathbb{E}\left(b \mid Z, D=1\right)\right] \end{array}$$

Since Z is NOT independent of b, we have that the conditional expectation $\mathbb{E}\left(b \mid Z, D=1\right)$ is different to $\mathbb{E}\left(b \mid Z\right) = \mathbb{E}\left(b\right) = 0$. Therefore, $\mathbb{E}\left[Z \ \varepsilon\right] \neq 0$ and the IV estimator of β^* is inconsistent. $\blacksquare$

Question 1.7 [10 points] Suppose that we consider the following model for the treatment variable: $D_i = 1\{u_i \leq \gamma(Z_i)\}$, where u_i is a zero mean random variable that is independent

of Z_i, and $\gamma(Z_i)$ is some function of Z_i. Suppose that (α_i, β_i, u_i) are normally distributed and independent of Z_i.

(a) Obtain the expression for $E(Y|Z{,}D=1)$ in terms of α^*, β^*, and inverse Mills ratio. Similarly, obtain the expression for $E(Y \mid Z,\, D=0)$ in terms of α^*, β^*, and inverse Mills ratio.

ANSWER. If ε and u are normally distributed random variables with mean zero, we have that for any constant c:

$$E(\varepsilon \mid u \leq c) \;=\; -\frac{\sigma_{\varepsilon u}}{\sigma_u} \frac{\phi\left(\dfrac{c}{\sigma_u}\right)}{\Phi\left(\dfrac{c}{\sigma_u}\right)}$$

$$E(\varepsilon \mid u \geq c) \;=\; \frac{\sigma_{\varepsilon u}}{\sigma_u} \frac{\phi\left(\dfrac{c}{\sigma_u}\right)}{1-\Phi\left(\dfrac{c}{\sigma_u}\right)}$$

Taking into account this property, we have that:

$$\begin{aligned}
\mathbb{E}(Y \mid Z,\, D=1) &= \alpha^* + \beta^* + \mathbb{E}(a + b\ D \mid Z,\, D=1) \\
&= \alpha^* + \beta^* + \mathbb{E}(a + b \mid Z,\, u \leq \gamma(Z)) \\
&= \alpha^* + \beta^* - \frac{\sigma_{(a+b)u}}{\sigma_u} \frac{\phi\left(\dfrac{\gamma(Z)}{\sigma_u}\right)}{\Phi\left(\dfrac{\gamma(Z)}{\sigma_u}\right)}
\end{aligned}$$

And,

$$\begin{aligned}
\mathbb{E}(Y \mid Z,\, D=0) &= \alpha^* + \mathbb{E}(a + b\ D \mid Z,\, D=0) \\
&= \alpha^* + \mathbb{E}(a \mid Z,\, u > \gamma(Z)) \\
&= \alpha^* + \frac{\sigma_{a,u}}{\sigma_u} \frac{\phi\left(\dfrac{\gamma(Z)}{\sigma_u}\right)}{1-\Phi\left(\dfrac{\gamma(Z)}{\sigma_u}\right)} \qquad \blacksquare
\end{aligned}$$

(b) Using the derivations in (a), propose a two-step method for consistent estimation of β^*.

ANSWER. First, note that $\mathbb{E}(Y \mid Z) = P_D(Z)\, \mathbb{E}(Y \mid Z,\, D=1) + [1-P_D(Z)]\, \mathbb{E}(Y \mid Z,\, D=0)$, and taking into account the expressions in (a), we have:

$$\mathbb{E}(Y \mid Z) = \alpha^* + \beta^*\ \Phi\left(\frac{\gamma(Z)}{\sigma_u}\right) + \left[\frac{\sigma_{a,u} - \sigma_{a+b,u}}{\sigma_u}\right]\ \phi\left(\frac{\gamma(Z)}{\sigma_u}\right)$$

Or, in a regression-like equation:

$$Y = \alpha^* + \beta^*\ \Phi_Z + \delta\ \phi_Z + e$$

where $\Phi_Z \equiv \Phi\left(\dfrac{\gamma(Z)}{\sigma_u}\right)$, $\phi_Z \equiv \phi\left(\dfrac{\gamma(Z)}{\sigma_u}\right)$, $\delta \equiv \dfrac{\sigma_{a,u} - \sigma_{a+b,u}}{\sigma_u}$, and e is an error term that is mean zero and mean independent of Z, i.e., $E(e|Z) = 0$.

Given this regression representation, we can estimate $(\alpha^*, \beta^*, \delta)$ using the following two-step method. In the first step, we estimate Probit model where the dependent variable is D and the explanatory variables are the terms of a polynomial in Z that approximate function $\gamma(Z)/\sigma_u$. Given the estimated Probit model, we construct the estimated values $\widehat{\Phi_i} \equiv \Phi\left(\widehat{\frac{\gamma(z_i)}{\sigma_u}}\right)$ and $\widehat{\phi_i} \equiv \phi\left(\widehat{\frac{\gamma(z_i)}{\sigma_u}}\right)$ for every observation i in the sample. In the second step, we run an OLS regression for y_i on $[1, \widehat{\Phi_i}, \widehat{\phi_i}]$. This OLS estimation provides a consistent estimator of $(\alpha^*, \beta^*, \delta)$. ■

Question 1.8 [10 points] Consider the model in Question 1.7, but now we do not assume that (α_i, β_i, u_i) are normally distributed We still assume that (α_i, β_i, u_i) are independent of Z_i, but their distribution is Nonparametrically specified. Suppose that the support of Z_i is the whole real line, and $P_D(z) \equiv F_u(\gamma(z))$ is strictly monotonic in z.

(a) Obtain the expression for $E(Y \mid Z, D=1)$ in terms of α^*, β^*, and a selection term. Similarly, obtain the expression for $E(Y \mid Z, D=0)$ in terms of α^*, β^*, and a selection term.

ANSWER. Under the assumption that u is independent of Z we have that $P_D(Z) \equiv F_u(\gamma(Z))$ and there is a one-to-one relationship between $P_D(Z)$ and $\gamma(Z)$. Furthermore, under the assumption of independence between Z and (α, β, u), we have that:

$$\begin{aligned}\mathbb{E}(Y \mid Z, D=1) &= \alpha^* + \beta^* + \mathbb{E}(a + b \mid Z, u \leq \gamma(Z)) \\ &= \alpha^* + \beta^* + S_{a+b}(\gamma(Z))\end{aligned}$$

And

$$\begin{aligned}\mathbb{E}(Y \mid Z, D=0) &= \alpha^* + \beta^* + \mathbb{E}(a \mid Z, u > \gamma(Z)) \\ &= \alpha^* + S_a(\gamma(Z))\end{aligned}$$

Finally, given the one-to-one relationship between $P_D(Z)$ and $\gamma(Z)$ we can represent the selection terms $S_{a+b}(\gamma(Z))$ and $S_a(\gamma(Z))$ as functions of the propensity score $P_D(Z)$. That is:

$$\begin{aligned}\mathbb{E}(Y \mid Z, D=1) &= \alpha^* + \beta^* + s_{a+b}(P_D(Z)) \\ &\text{and} \\ \mathbb{E}(Y \mid Z, D=1) &= \alpha^* + s_a(P_D(Z)) \quad ■\end{aligned}$$

(b) Using the equations derived in (a), show that β^* is identified. Propose a two-step method for the consistent estimation of β^*.

ANSWER. First, note that $\mathbb{E}(Y \mid Z) = P_D(Z)\, \mathbb{E}(Y \mid Z, D=1) + [1 - P_D(Z)]\, \mathbb{E}(Y \mid Z, D=0)$, and taking into account the expressions in (a), we have:

$$\mathbb{E}(Y \mid Z) = \alpha^* + \beta^*\ P_D(Z) + s(P_D(Z))$$

with $s(p) \equiv p\ s_{a+b}(p) + (1-p)\ s_a(p)$. Or, in a regression-like equation:

$$Y = \alpha^* + \beta^*\ P_D + s(P_D) + e$$

where $P_D \equiv P_D(Z)$, $s(P_D) \equiv s(P_D(Z))$, and e is an error term that is mean zero and mean independent of Z, i.e., $E(e|Z) = 0$.

Given this regression representation, consider the following two-step nonparametric method. In the first step, we estimate a nonparametric regression for the dependent variable D on the instrument Z. Given this estimated NP regression, we construct the estimated values $\widehat{P}_i \equiv P_D(z_i)$ for every observation i in the sample. In the second step, we run an Partially Linear Model for y_i on $\widehat{P}_i$ and $s(\widehat{P}_i)$. Unfortunately, it is clear that in this second step we cannot identify the parameter β^* from the nonparametric selection term $s(\widehat{P}_i)$. The ATE is not identified in this model. ■

———————————————————————————————————————

PROBLEM 2 (25 points). Consider the binary choice model $Y = 1\{Z + \beta\ X - \varepsilon \leq 0\}$, where ε is indenpendently distributed of (Z, X) with CDF $F(.)$ that is continuously differentiable over the real line. The explanatory variable Z has support over the whole real line, while the explanatory variable X is binary with support $\{0, 1\}$.

(a) [10 points] Describe Maximum Score Estimator (MSE) and the Smooth Maximum Score Estimator of the parameter β. Explain why the SMSE deals with some of the limitations in the MSE.

ANSWER. Given that ε is independent of (Z, X) is also median independent. Therefore, $median(\varepsilon|Z, X) = median(\varepsilon) = 0$. Then, we have that:

$$median(Y \mid Z, X) = 1\{Z + \beta\ X \leq 0\}$$

Suppose that we use the this conditional median, $1\{Z + \beta\ X \leq 0\}$, to predict the outcome $Y = 1$. The Maximum Score Estimator (MSE) of β is defined as the value of β that maximizes the score function that counts the number of correct predictions when we predict $Y = 1$ iff $1\{Z + \beta\ X \leq 0\}$ and predict $Y = 0$ iff $1\{Z + \beta\ X > 0\}$. That is,

$$\hat{\beta}_{MSE} = \arg\max_{\beta}\ S(\beta) = \sum\nolimits_{i=1}^{n} y_i\ 1\{z_i + \beta\ x_i \leq 0\} + (1 - y_i)\ 1\{z_i + \beta\ x_i > 0\}$$

This is estimator is consistent but it has several limitations: (1) it is NOT root-n consistent. Its rate of convergence to the true β is $n^{1/3}$. (2) It is not asymptotically normal. It has a not standard distribution. (3) We cannot use the standard gradient based methods to search for the MSE. And (4) if the sample size is not large enough, there may not be a unique value of β that maximizes $S(\beta)$. The maximizer of $S(\beta)$ can be a whole (compact) set in the space of β.

All these limitations of the MSE are related to the fact that the criterion function $S(\beta)$ is a discontinuous step-function. Based on this, Horowitz proposes an estimator that is based on a "smooth" score function. First, note that score function $S(\beta)$ can be written as follows:

$$\begin{aligned}
S(\beta) &= \sum_{i=1}^{n} y_i\ 1\{z_i + \beta\ x_i \leq 0\} + (1 - y_i)\ 1\{z_i + \beta\ x_i > 0\} \\
&= \sum_{i=1}^{n} y_i\ 1\{z_i + \beta\ x_i \leq 0\} + (1 - y_i)\left(1 - 1\{z_i + \beta\ x_i \leq 0\}\right) \\
&= \sum_{i=1}^{n} (1 - y_i) + \sum_{i=1}^{n} (2y_i - 1)\ \ 1\{z_i + \beta\ x_i \leq 0\}
\end{aligned}$$

Therefore, maximizing $S(\beta)$ is equivalent to maximizing $\sum\nolimits_{i=1}^{n} (2y_i - 1)\ 1\{z_i + \beta\ x_i \leq 0\}$, and:

$$\hat{\beta}_{MSE} = \arg\max_{\beta}\ \ \sum\nolimits_{i=1}^{n} (2y_i - 1)\ \ 1\{z_i + \beta\ x_i \leq 0\}$$

Horowitz proposes to replace $1\{z_i + \beta\ x_i \leq 0\}$ by a function $\Phi\left(\dfrac{z_i + \beta\ x_i}{b_n}\right)$, where $\Phi(.)$ is the CDF of the standard normal, and b_n is a bandwidth parameter such that: (i) $b_n \to 0$ as $n \to \infty$; and (ii) $nb_n \to \infty$ as $n \to \infty$. That is, b_n goes to zero but more slowly than $1/n$. The Smooth-MSE is defined as:

$$\hat{\beta}_{SMSE} = \arg\max_{\beta} \quad \sum_{i=1}^{n} (2y_i - 1)\ \ \Phi\left(\frac{z_i + \beta\ x_i}{b_n}\right)$$

As $n \to \infty$, and $b_n \to 0$, the function $\Phi\left(\dfrac{z_i + \beta\ x_i}{b_n}\right)$ converges uniformly to $1\{z_i + \beta\ x_i \leq 0\}$, and the criterion function converges uniformly to the Score function. This implies the consistency of $\hat{\beta}_{SMSE}$.

Under the additional condition that $nb_n \to \infty$ as $n \to \infty$, and the Kernel function has enough smooth derivatives (e.g., Normal CDF) this estimator is n^{δ} consistent and asymptotically normal, with $2/5 \leq \delta < 1/2$. It can be computed using standard gradient search methods because the criterion function is continuously differentiable. ■

(b) [15 points] Suppose that β is known (or consistently estimated). Provide a constructive proof of the identification of the distribution function $F(\varepsilon_0)$ at any value ε_0 in the real line.

ANSWER. The CCP function $P(z,x) = \Pr(Y = 1|Z = z,\ X = x)$ is nonparametrically identified from the data at every (z, x). Suppose that β has been identified/estimated (e.g., SMSE estimator). For arbitrary values of $x \in \{0, 1\}$ and $\varepsilon \in \mathbb{R}$, say (x_0, ε_0), we want to estimate $F_{\varepsilon|x_0}(\varepsilon_0)$. Let z_0 be the value $z_0 = \varepsilon_0 - \beta\ x_0$, and let $P(z_0, x_0)$ be the CCP evaluated at (z_0, x_0). Then:

$$\begin{aligned} P(z_0, x_0) &\equiv \Pr(Y = 1 \mid Z = z_0,\ X = x_0) \\ &= \Pr(\varepsilon \geq z_0 + \beta\ x_0) \\ &= 1 - F_{\varepsilon|x_0}(z_0 + \beta\ x_0) = 1 - F_{\varepsilon|x_0}(\varepsilon_0) \end{aligned}$$

Or equivalently, $F_{\varepsilon|x_0}(\varepsilon_0) = 1 - P(z_0, x_0)$. That is, for any $(x_0$,$\varepsilon_0)$ we can always define a value z_0 such that $F_{\varepsilon|x_0}(\varepsilon_0)$ is equal to $1 - P(z_0, x_0)$. ■

PROBLEM 3 (25 points). Consider the dynamic panel data model, $y_{it} = \beta\ y_{it-1} + \alpha_i + u_{it}$, where u_{it} is *i.i.d.*

Question 3.1 [10 points] Define Anderson-Hsiao IV estimator of β. Derive the expression of its asymptotic variance. Explain the weak instruments problem as β approaches 1.

ANSWER. For the AR(1)-PD model without other regressors, the Anderson-Hsiao estimator is defined as:

$$\hat{\beta}_{AH} = \frac{\displaystyle\sum_{t=3}^{T} \left[\sum_{i=1}^{N} y_{i,t-2}\ \Delta y_{it}\right]}{\displaystyle\sum_{t=3}^{T} \left[\sum_{i=1}^{N} y_{i,t-2}\ \Delta y_{i,t-1}\right]}$$

Notice that we need $T \geq 3$ to implement this estimator. Anderson-Hsiao estimator is an IV estimator in the equation in FD where the FD of the lagged endogenous variable is instrumented with y_{it-2}.

To derive the asymptotic variance, notice that:

$$\sqrt{N}\left(\hat{\beta}_{AH}-\beta\right) = \sum_{t=3}^{T}\left[\frac{1}{N}\sum_{i=1}^{N} y_{i,t-2}\ \Delta y_{i,t-1}\right]^{-1}\sum_{t=3}^{T}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i,t-2}\ \Delta u_{it}\right]$$

As $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N} y_{i,t-2}\ \Delta y_{i,t-1} \to_p \mathbb{E}\left(y_{t-2}\ \Delta y_{t-1}\right)$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} y_{i,t-2}\ \Delta u_{it} \to_d N\left(0, Var\left(y_{i,t-2}\ \Delta u_{it}\right)\right)$$

By Mann-Wald Theorem, as $N \to \infty$,

$$\sqrt{N}\left(\hat{\beta}_{AH}-\beta\right) \to_d N\left(0, V_{AH}\right)$$

$$\text{with } V_{AH} = \frac{Var\left(y_{t-2}\ \Delta u_t\right)}{\left[\mathbb{E}\left(y_{t-2}\ \Delta y_{t-1}\right)\right]^2}$$

To obtain the expression of V_{AH} in terms of the parameters of the model, we need to derive the expressions for $Var\left(y_{t-2}\ \Delta u_t\right)$ and $\mathbb{E}\left(y_{t-2}\ \Delta y_{t-1}\right)$.

(A) $Var\left(y_{t-2}\ \Delta u_t\right)$. First, note that $Var(y_{t-2}\ \Delta u_t) = E(y_{t-2}^2\ \Delta u_t^2)$. Note that y_{t-2} depends on transitory shocks u at periods $t-2$ and before. Therefore, Δu_t is independent of y_{t-2}. Applying the law of iterative expectations and the independence between Δu_t and y_{t-2}, we have that $E(y_{it-2}^2\ \Delta u_t^2) = E(y_{t-2}^2)$ $E(\Delta u_t^2)$. Given that $E(\Delta u_t^2) = 2\sigma_u^2$, and $E\left(y_{t-2}^2\right) = \dfrac{\sigma_\alpha^2}{(1-\beta)^2} + \dfrac{\sigma_u^2}{1-\beta^2}$, we have that:

$$Var(y_{t-2}\ \Delta u_t) = 2\sigma_u^2\left[\frac{\sigma_\alpha^2}{(1-\beta)^2} + \frac{\sigma_u^2}{1-\beta^2}\right].$$

(B) $\mathbb{E}\left(y_{t-2}\ \Delta y_{t-1}\right)$. We have that $\mathbb{E}\left(y_{t-2}\ \Delta y_{t-1}\right) = \mathbb{E}\left(y_1\ \Delta y_2\right)$, and

$$\begin{aligned}
\mathbb{E}\left(y_1\ \Delta y_2\right) &= E\left([u_1+\beta u_0+...]\ \ [u_2+(\beta-1)u_1+\beta(\beta-1)u_0+...]\right) \\
&= (\beta-1)\ \left[\sigma_u^2+\beta^2\ \sigma_u^2+\beta^4\ \sigma_u^2+\beta^6\ \sigma_u^2+...\right] \\
&= \frac{(\beta-1)\ \sigma_u^2}{1-\beta^2} = \frac{-\sigma_u^2}{1+\beta}
\end{aligned}$$

Then, combining (A) and (B), we have that:

$$\begin{aligned}
V_{AH} &= \frac{2\sigma_u^2\ \left[\dfrac{\sigma_\alpha^2}{(1-\beta)^2}+\dfrac{\sigma_u^2}{1-\beta^2}\right]}{\left(\dfrac{-\sigma_u^2}{1+\beta}\right)^2} \\
&= \frac{2(1+\beta)^2}{\sigma_u^2}\frac{1}{1-\beta}\left[\frac{\sigma_\alpha^2}{1-\beta}+\frac{\sigma_u^2}{1+\beta}\right] \\
&= 2\ \left(\frac{1+\beta}{1-\beta}\right)\left[1+\left(\frac{1+\beta}{1-\beta}\right)\frac{\sigma_\alpha^2}{\sigma_u^2}\right]
\end{aligned}$$

We can see that as β approaches to 1 the variance of the AH estimator goes to infinity. This is due to a weak instruments problem. As β approaches to 1, the proportion of the variance of Δy_{it-1} explained by the instrument y_{it-2} (i.e., the R-square in the auxiliary regression of Δy_{it-1} on y_{it-2}) goes to zero. ■

Question 3.2 [15 points] Suppose that $|\beta| < 1$ and for every individual i there is a time period $t_i^* < 1$ such that $y_{it_i^*} = \alpha_i/1-\beta$. Show that under these conditions we can obtain an IV estimator of β that does not suffer of a weak instruments problem as β approaches 1.

ANSWER. Suppose that $|\beta| < 1$, and that at some period $t^* \leq 0$ in the past (that can be individual specific), the process $\{y_{it}\}$ visited his individual-specific mean,

$$y_{it^*} = \mu_{y,i} \equiv \mathbb{E}\left(y_{it} \mid i\right)$$

Stationarity implies that $\mu_{y,i} = \beta\ \mu_{y,i} + \alpha_i$, and

$$\mu_{y,i} = \frac{\alpha_i}{1-\beta}$$

Under this stationarity assumption, we have that, one period after t^*:

$$\begin{aligned} y_{i,t^*+1} &= \beta\ \mu_{y,i} + \alpha_i + u_{i,t^*+1} \\ &= \mu_{y,i} - (1-\beta)\mu_{y,i} + \alpha_i + u_{i,t^*+1} \\ &= \mu_{y,i} - (1-\beta)\left[\frac{\alpha_i}{1-\beta}\right] + \alpha_i + u_{i,t^*+1} \\ &= \mu_{y,i} + u_{i,t^*+1} \end{aligned}$$

And for any $t > t^*$, we have that:

$$y_{it} = \mu_{y,i} + \sum_{j=0}^{t-t^*} \beta^j\ u_{i,t-j}$$

We have that:

$$\Delta y_{it} = \sum_{j=0}^{t-t^*} \beta^j\ \Delta u_{i,t-j}$$

And this implies that Δy_{it} does NOT dependent of α_i. This property implies that there are valid instruments for the equation in levels.

$$y_{it} = \beta\ y_{i,t-1} + \alpha_i + u_{it}$$

If u_{it} is not serially correlated, then $\triangle y_{i,t-1}$, $\triangle y_{i,t-2}$, ... are not correlated with $(\alpha_i + u_{it})$. For instance, we can estimate β by IV using $\Delta y_{i,t-1}$ as instrument of $y_{i,t-1}$. This IV estimator is:

$$\hat{\beta} = \frac{\displaystyle\sum_{i=1}^{N}\sum_{t=3}^{T} \Delta y_{i,t-1}\ y_{it}}{\displaystyle\sum_{i=1}^{N}\sum_{t=3}^{T} \Delta y_{i,t-1}\ y_{i,t-1}}$$

This IV estimator does not suffer of weak instruments as $\beta \to 1$. Notice that:

$$\begin{aligned} \sqrt{N}\left(\hat{\beta} - \beta\right) &= \frac{N^{-1/2}\sum_{i=1}^{N} \Delta y_{i,t-1}\ \ (\alpha_i + u_{it})}{N^{-1}\sum_{i=1}^{N} \Delta y_{i,t-1}\ y_{i,t-1}} \\ \to_d &\qquad N\left(0, V_{BB}\right) \end{aligned}$$

where $V_{BB} = \dfrac{Var\left(\Delta y_{t-1}\left(\alpha + u_t\right)\right)}{\mathbb{E}\left(\Delta y_{t-1}\ y_{t-1}\right)^2}$. Given that: $Var\left(\Delta y_{t-1}\left(\alpha_i + u_{it}\right)\right) = 2\sigma_u^2\ \left(\sigma_u^2 + \sigma_\alpha^2\right)/\left(1+\beta\right)$, and $\mathbb{E}\left(\Delta y_{i,t-1}\ y_{i,t-1}\right) = \sigma_u^2/\left(1+\beta\right)$, we have that:

$$\begin{aligned} V_{BB} &= \frac{2\sigma_u^2\left(\sigma_u^2+\sigma_\alpha^2\right)/\left(1+\beta\right)}{\sigma_u^4/\left(1+\beta\right)^2} \\ &= 2\left(1+\beta\right)\left(1+\frac{\sigma_\alpha^2}{\sigma_u^2}\right) \end{aligned}$$

In contrast with the AH estimator, the variance of this estimator does not goes to infinity as β goes to one. ■

——————————————————————————————

END OF THE EXAM