ECONOMETRICS II (ECO 2401S)
University of Toronto. Department of Economics. Spring 2012
Instructor: Victor Aguirregabiria

SOLUTION TO FINAL EXAM
Friday, April 13, 2012. From 9:00-12:00 (3 hours)

INSTRUCTIONS: The exam consists of three Problems. You have to answer all the questions. No study aids, including calculators, are allowed.
TOTAL MARKS = 100

PROBLEM 1 (40 points).[Differences-in-Differences in a Dynamic Panel Data model]. Consider the following linear panel data model:

$$Y_{it} = \delta\ Y_{it-1} + \beta\ D_{it} + \gamma_t + \alpha_i + u_{it}$$

for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$, with N large and $T = 3$. The regressor D_{it} is the product of two binary variables: $D_{it} = C_i * TD_t^{(3)}$, where C_i is the binary variable that indicates that individual i belongs to the "experimental group", and $TD_t^{(3)}$ is a dummy that is 1 if time period t is greater or equal than 3 and it is zero otherwise, i.e., $TD_t^{(3)} = 1\{t \geq 3\}$. We are interested in the estimation of parameters β and δ. The parameter β measures the short run effect of D on Y, and $\beta/(1-\delta)$ represents the long run effect. We are concerned about the endogeneity of D_{it}. We think that C_i can be correlated with the individual effect α_i, and $TD_t^{(3)}$ can be correlated with the aggregate effect γ_t. However, we are willing to assume that the unobservable u_{it} is not serially correlated and it is independent of C_i and $TD_t^{(3)}$.

Question 1.1. [25 points] Consider the following estimators of the parameter β: (i) OLS estimator of the model in first differences under the assumption that $\delta = 0$, i.e., this is the standard Differences-in-Differences (DID) estimator; and (ii) OLS estimator of the model in first differences without the assumption that $\delta = 0$, i.e., this estimator can be described as a 'dynamic' DID estimator.

(a) [5 points] Describe the regression equations and the sets of regressors in the standard and in the dynamic DID estimators.

ANSWER: For the standard DID estimation, the cross-sectional regression equation is:

$$\Delta Y_{i3} = \Delta\gamma_3 + \beta\ C_i + \Delta e_{i3}$$

where $\Delta Y_{i3} \equiv Y_{i3} - Y_{i2}$, $\Delta\gamma_3$ is a constant term equal to $\gamma_3 - \gamma_2$, and Δe_{i3} is an error term that should be equal to $\delta \Delta Y_{i2} + \Delta u_{i3}$. Note that the regressor C_i comes from the first difference $\Delta D_{i3} = D_{i3} - D_{i2} = C_i * 1 - C_i * 0$.

For the dynamic DID estimation, the cross-sectional regression equation is:

$$\Delta Y_{i3} = \Delta\gamma_3 + \delta\ \Delta Y_{i2} + \beta\ C_i + \Delta u_{i3}$$

with $\Delta Y_{i2} \equiv Y_{i2} - Y_{i1}$.

(b) [5 points] Show that the standard DID estimator of β in this model is inconsistent.

ANSWER: The standard DID estimator is the OLS estimator of the slope parameter β in the standard DID regression model $\Delta Y_{i3} = \Delta\gamma_3 + \beta\ C_i + \Delta e_{i3}$. Let $\widehat{\beta}_{SDID}$ be that estimator. By definition:

$$p\lim \widehat{\beta}_{SDID} = \beta + \frac{\mathbb{E}\left(\Delta e_{i3}\ C_i\right)}{Var(C_i)}$$

Given that $\Delta e_{i3} = \delta \Delta Y_{i2} + \Delta u_{i3}$ and that C_i is independent of u_{it}, we have that $\mathbb{E}\left(\Delta e_{i3}\ C_i\right) = \delta \mathbb{E}\left(\Delta Y_{i2}\ C_i\right)$. Therefore,

$$p\lim \widehat{\beta}_{SDID} = \beta + \frac{\delta\ \mathbb{E}\left(\Delta Y_{i2}\ C_i\right)}{Var(C_i)}$$

In general, $\delta\ \mathbb{E}\left(\Delta Y_{i2}\ C_i\right)$ is not zero and standard DID estimator of β in this dynamic model is inconsistent.

(c) [5 points] Prove that the standard DID estimator is consistent only if $\delta = 0$ or if $cov(\Delta Y_{i2}, C_i) = 0$.

ANSWER: Given the expression for $p\lim \widehat{\beta}_{SDID}$ that we have derived in point (b), it is clear that $p\lim \widehat{\beta}_{SDID} = \beta$ only if $\delta = 0$ or if $cov(\Delta Y_{i2}, C_i) = 0$.

(d) [5 points] Show that the 'dynamic' DID estimator of β in this model is inconsistent.

ANSWER: The dynamic DID estimator is the OLS estimator of the slopes parameter δ and β in the dynamic DID regression model $\Delta Y_{i3} = \Delta\gamma_3 + \delta\ \Delta Y_{i2} + \beta\ C_i + \Delta u_{i3}$. In general, this estimator is consistent if $\mathbb{E}\left(\Delta Y_{i2}\ \Delta u_{i3}\right) = 0$ and $\mathbb{E}\left(C_i\ \Delta u_{i3}\right) = 0$. Independence between C_i and u_{it} implies that $\mathbb{E}\left(C_i\ \Delta u_{i3}\right) = 0$. However, $\mathbb{E}\left(\Delta Y_{i2}\ \Delta u_{i3}\right)$ is not zero:

$$\begin{aligned}\mathbb{E}\left(\Delta Y_{i2}\ \Delta u_{i3}\right) &= \mathbb{E}\left(\left[(\delta - 1)\ Y_{i1} + \gamma_2 + \alpha_i + u_{i2}\right]\ \Delta u_{i3}\right) \\ &= \mathbb{E}\left(u_{i2}\ \Delta u_{i3}\right) = -\sigma_u^2 \neq 0\end{aligned}$$

(e) [5 points] Prove that the estimator is consistent only if $cov(\Delta Y_{i2}, C_i) = 0$.

ANSWER: In general, the correlation between ΔY_{i2} and Δu_{i3} will include an asymptotic bias in the OLS estimation of both δ and β. However, if the two regressors are orthogonal (that is, if $cov(\Delta Y_{i2}, C_i) = 0$) then the correlation between ΔY_{i2} and Δu_{i3} introduces and asymptotic bias in the estimation of δ but it does not have any effect on the estimation of β. More precisely, if $cov(\Delta Y_{i2}, C_i) = 0$, then:

$$\begin{aligned}p\lim \widehat{\beta}_{DDID} &= \frac{\mathbb{E}\left(C_i\ \Delta Y_{i3}\right)}{Var(C_i)} = \beta + \frac{\mathbb{E}\left(C_i\ \Delta u_{i3}\right)}{Var(C_i)} = \beta \\ p\lim \widehat{\delta}_{DDID} &= \frac{\mathbb{E}\left(\Delta Y_{i2}\ \Delta Y_{i3}\right)}{Var(\Delta Y_{i2})} = \delta + \frac{\mathbb{E}\left(\Delta Y_{i2}\ \Delta u_{i3}\right)}{Var(\Delta Y_{i2})} \neq \delta\end{aligned}$$

Question 1.2. [15 points]
(a) [5 points] Propose a feasible estimator of β in this model that is consistent when $\delta \neq 0$ and $cov(\Delta Y_{i2}, C_i) \neq 0$. Explain how to implement the estimator.

ANSWER: Under the conditions of this model, Anderson-Hsiao IV method provides consistent estimators of δ and β. In the regression equation,

$$\Delta Y_{i3} = \Delta\gamma_3 + \delta\ \Delta Y_{i2} + \beta\ C_i + \Delta u_{i3}$$

we instrument ΔY_{i2} with Y_{i1} (and C_i should not be instrumented because is not correlated with Δu_{i3}). We can implement this estimator using 2SLS where the first step consists of a regression of ΔY_{i2} on Y_{i1}, and the second step is a regression of ΔY_{i3} on C_i and $\widehat{\Delta Y}_{i2}$, where $\widehat{\Delta Y}_{i2}$ represents the fitted values from the regression in the first step.

(b) [5 points] Explain why this estimator is consistent.

ANSWER: This IV estimator is consistent if satisfies conditions: (1) $\mathbb{E}(Y_{i1}\ \Delta u_{i3}) = 0$; (2) $Cov(Y_{i1}, \Delta Y_{i2}) \neq 0$; and (3) Y_{i1} and C_i are not perfectly collinear. If u_{it} is not serially correlated, then condition (1) holds. If $\delta < 1$, then condition (2) holds. In general, condition (3) holds.

(c) [5 points] Comment on the efficiency of the estimator when δ is close to 1.

ANSWER: When δ is exactly equal 1, we have that $Cov(Y_{i1}, \Delta Y_{i2}) = 0$, and the instrument Y_{i1} does not have any explanatory power of the endogenous regressor ΔY_{i2}. Therefore, when $\delta = 1$ the IV estimator cannot identify parameters δ and β, or in other words the variance of the IV estimators of these parameters is infinite. When $\delta < 1$ but close to 1, the instrument Y_{i1} has very little explanatory power of the endogenous regressor ΔY_{i2}. This implies that, unless we have a large sample, the estimates of δ and β can be very imprecise.

PROBLEM 2 (40 points). Suppose the following Treatment Effects model.

$$Y = (1 - D)\ Y_0^* + D\ Y_1^*$$

where Y is the outcome variable, $D \in \{0, 1\}$ is the treatment dummy, and Y_1^* and Y_0^* are latent variables that represent the outcome with and without treatment, respectively. The researcher observes a random sample of Y and D and is interested in the estimation of the Average Treatment Effect, $ATE \equiv \mathbb{E}(Y_1^* - Y_0^*)$.

Suppose that the researcher can implement a randomized experiment. Let $Z \in \{0, 1\}$ be a binary indicator that represents the 'ex-ante' assignment of an individual to the treatment (experimental) group. The binary variable $D \in \{0, 1\}$ represents the actual (or 'ex-post') assignment of an individual to the treatment group. Under perfect compliance we have that $D = Z$ for every individual. In general, there is not perfect compliance such that for some individuals $D \neq Z$, though variables D and Z are positively correlated. The dataset is $\{y_i, d_i, z_i : i = 1, 2, ..., n\}$.

Question 2.1. [5 points] Write the model as a regression model where ATE is the coefficient associated to regressor D.

ANSWER: Define the mean values $\mu_0 \equiv \mathbb{E}(Y_0^*)$ and $\mu_1 \equiv \mathbb{E}(Y_1^*)$, such that $ATE = \mu_1 - \mu_0$. Also, define the deviations with respect to mean values, $e_0 \equiv Y_0^* - \mu_0$ and $e_0 \equiv Y_1^* - \mu_1$. Using these definitions, we have that $Y_0^* = \mu_0 + e_0$ and $Y_1^* = \mu_1 + e_1$. Replacing these expressions into equation $Y = (1 - D)\ Y_0^* + D\ Y_1^*$, we have that:

$$\begin{aligned} Y &= (1-D)\ \ (\mu_0 + e_0) + D\ \ (\mu_1 + e_1) \\ &= \mu_0 + ATE\ D + e \end{aligned}$$

where the error term e is equal to $e_0 + D\ (e_1 - e_0)$. Equation $Y = \mu_0 + ATE\ D + e$ is a simple regression with regressor D, parameters μ_0 and ATE, and error term e.

Question 2.2. [5 points] Show that the OLS estimator of the parameter ATE in that regression model is equal to $\overline{Y}_{(D=1)} - \overline{Y}_{(D=0)}$, where $\overline{Y}_{(D=j)} \equiv \sum_{i=1}^{n} y_i\ 1\{d_i = j\}$ / $\sum_{i=1}^{n} 1\{d_i = j\}$, for $j = 0, 1$.

ANSWER: In general, the OLS estimator of the slope parameter ATE is:

$$\widehat{ATE}_{OLS} = \frac{\sum_{i=1}^{n}\left(d_i - \overline{D}\right)\ \left(y_i - \overline{Y}\right)}{\sum_{i=1}^{n}\left(d_i - \overline{D}\right)^2}$$

where $\overline{D}$ and $\overline{Y}$ are the sample means $n^{-1}\sum_{i=1}^{n} d_i$ and $n^{-1}\sum_{i=1}^{n} y_i$, respectively. Taking into account that d_i is a binary variable and therefore $d_i^2 = d_i$, we have that the denominator of $\widehat{ATE}_{OLS}$ is:

$$\begin{aligned} \sum_{i=1}^{n}\left(d_i - \overline{D}\right)^2 &= \sum_{i=1}^{n} d_i^2 - 2\sum_{i=1}^{n} d_i\overline{D} + \sum_{i=1}^{n}\overline{D}^2 \\ &= n\ \overline{D} - 2n\ \overline{D}^2 + n\ \overline{D}^2 \\ &= n\ \overline{D}(1 - \overline{D}) \end{aligned}$$

The numerator of $\widehat{ATE}_{OLS}$ is:

$$\sum_{i=1}^{n}\left(d_i - \overline{D}\right)\ \left(y_i - \overline{Y}\right) = \sum_{i=1}^{n} d_i y_i - n\ \overline{D}\ \overline{Y}$$

Solving these expressions into the formula of $\widehat{ATE}_{OLS}$, we have that:

$$\begin{aligned} \widehat{ATE}_{OLS} &= \frac{\sum_{i=1}^{n} d_i y_i - n\ \overline{D}\ \overline{Y}}{n\ \overline{D}(1 - \overline{D})} \\ &= \frac{1}{1 - \overline{D}}\left(\frac{\sum_{i=1}^{n} d_i y_i}{\sum_{i=1}^{n} d_i} - \overline{Y}\right) \\ &= \frac{1}{1 - \overline{D}}\left(\overline{Y}_{(D=1)} - \overline{Y}\right) \end{aligned}$$

Note that:

$$\begin{aligned} \overline{Y} &= n^{-1}\sum_{i=1}^{n} d_i y_i + (1 - d_i) y_i \\ &= \overline{D}\ \overline{Y}_{(D=1)} + (1 - \overline{D})\ \overline{Y}_{(D=0)} \end{aligned}$$

And solving this expression in the equation for the $\widehat{ATE}_{OLS}$, we have that:

$$\begin{aligned} \widehat{ATE}_{OLS} &= \frac{1}{1 - \overline{D}}\left(\overline{Y}_{(D=1)} - \overline{D}\ \overline{Y}_{(D=1)} - (1 - \overline{D})\ \overline{Y}_{(D=0)}\right) \\ &= \overline{Y}_{(D=1)} - \overline{Y}_{(D=0)} \end{aligned}$$

Question 2.3. [5 points] Suppose that individuals' non-compliance to 'ex-ante' assignment is not random such that variable D is not independent of latent variables Y_0^* and Y_1^*. Show formally that the OLS estimator of the ATE in Question 2.2 is inconsistent. Obtain the expression of the asymptotic bias in terms of population expectations such as $\mathbb{E}(Y_1^*|D)$, $\mathbb{E}(Y_0^*|D)$, or/and $E(D)$.

ANSWER: By the LLN:

$$\begin{aligned} p\lim\ \widehat{ATE}_{OLS} &= p\lim\ \overline{Y}_{(D=1)} - p\lim\ \overline{Y}_{(D=0)} \\ &= \mathbb{E}(Y|D=1) - \mathbb{E}(Y|D=0) \\ &= \mathbb{E}(Y_1^*|D=1) - \mathbb{E}(Y_0^*|D=0) \\ &= ATE + \mathbb{E}(e_1|D=1) - \mathbb{E}(e_0|D=0) \end{aligned}$$

And the asymptotic bias of $\widehat{ATE}_{OLS}$ is

$$p\lim\ \widehat{ATE}_{OLS} - ATE = \mathbb{E}(e_1|D=1) - \mathbb{E}(e_0|D=0)$$

that is not zero because D is not independent of e_1 and e_0.

Question 2.4. [5 points] Consider the Instrumental Variables (IV) estimation of ATE in this regression model. More specifically, we instrument the actual treatment assignment dummy D using the ex-ante treatment assignment dummy Z. Show that the IV or 2SLS estimator is equal to:

$$\widehat{ATE}_{IV} = Wald = \frac{\overline{Y}_{(Z=1)} - \overline{Y}_{(Z=0)}}{\overline{D}_{(Z=1)} - \overline{D}_{(Z=0)}}$$

where $\overline{Y}_{(Z=j)} \equiv \sum_{i=1}^{n} y_i\ 1\{z_i = j\}\ /\ \sum_{i=1}^{n} 1\{z_i = j\}$, and $\overline{D}_{(Z=j)} \equiv \sum_{i=1}^{n} d_i\ 1\{z_i = j\}\ /\ \sum_{i=1}^{n} 1\{z_i = j\}$, for $j = 0, 1$.

ANSWER: In general, the IV estimator in this regression model is:

$$\widehat{ATE}_{IV} = \frac{\sum_{i=1}^{n}\left(z_i - \overline{Z}\right)\ \left(y_i - \overline{Y}\right)}{\sum_{i=1}^{n}\left(z_i - \overline{Z}\right)\ \left(d_i - \overline{D}\right)}$$

where $\overline{Z}$ is the sample mean $n^{-1}\sum_{i=1}^{n} z_i$. The numerator of this expression is:

$$\begin{aligned} \textstyle\sum_{i=1}^{n}\left(z_i - \overline{Z}\right)\ \left(y_i - \overline{Y}\right) &= \textstyle\sum_{i=1}^{n} z_i y_i - n\ \overline{Z}\ \overline{Y} \\ &= n\ \overline{Z}\left[\frac{\sum_{i=1}^{n} z_i y_i}{\sum_{i=1}^{n} z_i} - \overline{Y}\right] \\ &= n\ \overline{Z}\left[\overline{Y}_{(Z=1)} - \overline{Y}\right] \end{aligned}$$

And given that $\overline{Y} = n^{-1}\sum_{i=1}^{n} z_i y_i + (1 - z_i) y_i = \overline{Z}\ \overline{Y}_{(Z=1)} + (1 - \overline{Z})\ \overline{Y}_{(Z=0)}$, we have that:

$$\begin{aligned} \textstyle\sum_{i=1}^{n}\left(z_i - \overline{Z}\right)\ \left(y_i - \overline{Y}\right) &= n\ \overline{Z}\left[\overline{Y}_{(Z=1)} - \overline{Z}\overline{Y}_{(Z=1)} - (1 - \overline{Z})\overline{Y}_{(Z=0)}\right] \\ &= n\ \overline{Z}\left(1 - \overline{Z}\right)\left[\overline{Y}_{(Z=1)} - \overline{Y}_{(Z=0)}\right] \end{aligned}$$

Similarly, the denominator of this expression is:

$$\begin{aligned}
\textstyle\sum_{i=1}^{n}\left(z_i-\overline{Z}\right)\ \left(d_i-\overline{D}\right) &= \textstyle\sum_{i=1}^{n} z_i d_i - n\ \overline{Z}\ \overline{D} \\
&= n\ \overline{Z}\left[\overline{D}_{(Z=1)}-\overline{D}\right] \\
&= n\ \overline{Z}\left[\overline{D}_{(Z=1)}-\overline{Z}\overline{D}_{(Z=1)}-(1-\overline{Z})\overline{D}_{(Z=0)}\right] \\
&= n\ \overline{Z}\left(1-\overline{Z}\right)\left[\overline{D}_{(Z=1)}-\overline{D}_{(Z=0)}\right]
\end{aligned}$$

Solving these expressions into the definition of $\widehat{ATE}_{IV}$ above, we have that:

$$\begin{aligned}
\widehat{ATE}_{IV} &= \frac{n\ \overline{Z}\left(1-\overline{Z}\right)\left[\overline{Y}_{(Z=1)}-\overline{Y}_{(Z=0)}\right]}{n\ \overline{Z}\left(1-\overline{Z}\right)\left[\overline{D}_{(Z=1)}-\overline{D}_{(Z=0)}\right]} \\
&= \frac{\overline{Y}_{(Z=1)}-\overline{Y}_{(Z=0)}}{\overline{D}_{(Z=1)}-\overline{D}_{(Z=0)}}
\end{aligned}$$

Question 2.5. [5 points] Suppose that: (i) Z is independent of Y_0^* and Y_1^*; (ii) $E(D|Z=1)-E(D|Z=0)\neq 0$; and (iii) individuals are homogeneous in their treatment effects such that for every individual $Y_1^*-Y_0^*=ATE$. Show formally that the IV estimator of the ATE in Question 2.4 is consistent.

ANSWER: By the LLN:

$$p\lim\ \widehat{ATE}_{IV} = \frac{\mathbb{E}(Y|Z=1)-\mathbb{E}(Y|Z=0)}{\mathbb{E}(D|Z=1)-\mathbb{E}(D|Z=0)}$$

Taking into account that $e_1-e_0=0$ such that $Y=\mu_0+ATE\ D+e_0$, and that e_0 is independent of Z and therefore $\mathbb{E}(e_0|Z=1)=\mathbb{E}(e_0|Z=0)=0$, we have that:

$$\begin{aligned}
\mathbb{E}(Y|Z=0) &= \mu_0+ATE\ \mathbb{E}(D|Z=0) \\
\mathbb{E}(Y|Z=1) &= \mu_0+ATE\ \mathbb{E}(D|Z=1)
\end{aligned}$$

Solving these expressions into the equation of $p\lim\ \widehat{ATE}_{IV}$ above, we obtain that:

$$p\lim\ \widehat{ATE}_{IV} = \frac{ATE\ \left[\mathbb{E}(D|Z=1)-\mathbb{E}(D|Z=0)\right]}{\mathbb{E}(D|Z=1)-\mathbb{E}(D|Z=0)} = ATE$$

Question 2.6. [5 points] Suppose that we maintain conditions (i) and (ii) in Question 2.5 but we relax condition (iii). Individuals are heterogeneous in their treatment effects such that there are some individuals with $Y_1^*-Y_0^*\neq ATE$. Show formally that the IV estimator of the ATE is inconsistent. Obtain the expression of the asymptotic bias in terms of population expectations such as $E(Y_1^*|Z)$, $E(Y_0^*|D)$, $E(D|Z)$, or/and $E(Z)$.

ANSWER: It is still true that by the LLN:

$$p\lim\ \widehat{ATE}_{IV} = \frac{\mathbb{E}(Y|Z=1)-\mathbb{E}(Y|Z=0)}{\mathbb{E}(D|Z=1)-\mathbb{E}(D|Z=0)}$$

However, now we have that $Y = \mu_0 + ATE\ D + e_0 + D(e_1 - e_0)$. This implies that:

$$\begin{aligned}\mathbb{E}(Y|Z=0) &= \mu_0 + ATE\ \mathbb{E}(D|Z=0) + \mathbb{E}(D(e_1 - e_0)|Z=0) \\ &= \mu_0 + ATE\ \mathbb{E}(D|Z=0) + \mathbb{E}(D|Z=0)\ \mathbb{E}(e_1 - e_0|D=1)\end{aligned}$$

And similarly,

$$\begin{aligned}\mathbb{E}(Y|Z=1) &= \mu_0 + ATE\ \mathbb{E}(D|Z=1) + \mathbb{E}(D(e_1 - e_0)|Z=1) \\ &= \mu_0 + ATE\ \mathbb{E}(D|Z=1) + \mathbb{E}(D|Z=1)\ \mathbb{E}(e_1 - e_0|D=1)\end{aligned}$$

By independence between Z and $e_1 - e_0$, we have that $\mathbb{E}(e_1 - e_0|D=1, Z=0) = \mathbb{E}(e_1 - e_0|D=1, Z=1) = \mathbb{E}(e_1 - e_0|D=1)$. Therefore,

$$\begin{aligned}p\lim\ \widehat{ATE}_{IV} &= \frac{[\mathbb{E}(D|Z=1) - \mathbb{E}(D|Z=0)]\ \ [ATE + \mathbb{E}(e_1 - e_0|D=1)]}{\mathbb{E}(D|Z=1) - \mathbb{E}(D|Z=0)} \\ &= ATE + \mathbb{E}(e_1 - e_0|D=1)\end{aligned}$$

And the asymptotic bias of $p\lim\ \widehat{ATE}_{IV}$ is equal to $\mathbb{E}(e_1 - e_0|D=1)$, or what is equivalent $\mathbb{E}(Y_1^* - Y_0^*|D=1) - ATE$.

Question 2.7. [10 points] Consider the following model for the relationship between the ex-ante and the actual treatment assignment dummies, Z and D.

$$D = (1 - Z)\ \ D_0^* + Z\ D_1^*$$

where D_1^* and D_0^* are latent variables that represent an individual's choice of treatment with and without ex-ante assignment to treatment, respectively. Suppose that conditions (i) and (ii) in Question 2.5 hold and: (a) Z is independent of the latent variables D_0^* and D_1^*; and (b) for every individual, $D_1^* \geq D_0^*$. Prove formally that under conditions (i)-(ii) and (a)-(b) the IV estimator converges in probability to the following parameter:

$$LATE = \mathbb{E}\left(Y_1^* - Y_0^*\ |\ D_1^* > D_0^*\right)$$

ANSWER: Again, we start with the expression:

$$p\lim\ \widehat{ATE}_{IV} = \frac{\mathbb{E}(Y|Z=1) - \mathbb{E}(Y|Z=0)}{\mathbb{E}(D|Z=1) - \mathbb{E}(D|Z=0)}$$

By independence between (D_0^*, D_1^*) and Z, we have that the denominator is:

$$\begin{aligned}\mathbb{E}(D|Z=1) - \mathbb{E}(D|Z=0) &= \mathbb{E}(D_1^*|Z=1) - \mathbb{E}(D_0^*|Z=0) \\ &= \mathbb{E}(D_1^* - D_0^*)\end{aligned}$$

By the monotonicity condition (b) (i.e., $D_1^* \geq D_0^*$), there are three possible values for (D_0^*, D_1^*): $(0,0)$, $(0,1)$, and $(1,1)$. Therefore, $\mathbb{E}(D_1^* - D_0^*) = \Pr(D_0^* = 0 \text{ and } D_1^* = 1) = \Pr(D_1^* > D_0^*)$. For the numerator, we have that:

$$\begin{aligned}\mathbb{E}(Y|Z=1)-\mathbb{E}(Y|Z=0) &= \mathbb{E}((1-D_1^*)Y_0^*+D_1^*Y_1^* \mid Z=1)-\mathbb{E}((1-D_0^*)Y_0^*+D_0^*Y_1^* \mid Z=0) \\ &= \mathbb{E}((1-D_1^*)Y_0^*+D_1^*Y_1^*)-\mathbb{E}((1-D_0^*)Y_0^*+D_0^*Y_1^*) \\ &= \mathbb{E}((D_1^*-D_0^*)\ (Y_1^*-Y_0^*)) \\ &= \Pr(D_0^*=0,\, D_1^*=1)\ \mathbb{E}(Y_1^*-Y_0^* \mid D_0^*=0,\, D_1^*=1)\end{aligned}$$

Solving these expressions into the formula of $p\lim \widehat{ATE}_{IV}$, we have that:

$$\begin{aligned}p\lim\ \widehat{ATE}_{IV} &= \frac{\Pr(D_1^*>D_0^*)\ \mathbb{E}(Y_1^*-Y_0^* \mid D_1^*>D_0^*)}{\Pr(D_1^*>D_0^*)} \\ &= \mathbb{E}(Y_1^*-Y_0^* \mid D_1^*>D_0^*) = LATE\end{aligned}$$

PROBLEM 3 (20 points). Consider the (static) Panel Data Tobit model:

$$Y_{it}=\max\{\ 0\ ;\ X_{it}\beta+\alpha_i+u_{it}\}$$

for $i=1,2,...,N$ and $t=1,2,...,T$, with N large and T small. The $1\times K$ vector of regressors X_{it} is strongly exogenous with respect to the unobservable $\{u_{i1},...,u_{iT}\}$. However, we expect X_{it} to be correlated with α_i. Suppose that we are willing to assume the following Random Effects specification for the relationship between X and α_i:

$$\alpha_i=X_{i1}\ \lambda_1+X_{i2}\ \lambda_2+...+X_{iT}\ \lambda_T+e_i$$

where λ_1, λ_2, ..., λ_T are T vectors of parameters with dimension $K\times 1$ each, and e_i is an error term that is assumed to be independent of $X_i\equiv(X_{i1},\,X_{i2},\,...,X_{iT})$. Under this assumption, we can write the Tobit model as:

$$Y_{it}=\max\{\ 0\ ;\ \mathbf{X}_i\ \boldsymbol{\pi}_t+(e_i+u_{it})\}$$

where the new error term e_i+u_{it} is independent of X_i, and π_1, π_2, ..., π_T are T vectors of parameters with dimension $KT\times 1$ each.

Question 3.1. [5 points] Establish the relationship between the vectors of parameters $\{\pi_t\}$ and the vectors $\{\lambda_t\}$ and β. Show that given the vectors $\{\pi_t\}$ we can uniquely identify the vectors $\{\lambda_t\}$ and β.

ANSWER: By construction, we have that $\boldsymbol{\pi}_t=(\pi_{t1}',\pi_{t2}',...,\pi_{tT}')'$, where for any (t,s), π_{ts} is a $K\times 1$ vector with the coefficients of the regressors X_{is} in the equation of Y_{it}. And the relationship between π_{ts} and $\lambda's$ and β is:

$$\pi_{ts}=\begin{cases}\lambda_s & \text{if} \quad t\neq s \\ \\ \beta+\lambda_s & \text{if} \quad t=s\end{cases}$$

Given this relationship, it is clear that knowledge of $\pi's$ implies that $\lambda's$ and β are uniquely (over)identified. For λ_1 is identified from any π_{t1} with $t\neq 1$, λ_2 is identified from any π_{t2} with

$t \neq 2$, etc. And β is identified from any difference $\pi_{tt} - \pi_{ts}$ for any t and $s \neq t$. For instance, with $T = 2$, we have that $\lambda_1 = \pi_{21}$, $\lambda_2 = \pi_{12}$, and $\beta = \pi_{11} - \pi_{21} = \pi_{22} - \pi_{12}$.

Question 3.2. [10 points] Under the assumption that $e_i + u_{it}$ is independent of X_i, propose an adaptive or semiparametric estimator of the parameters $\{\pi_t\}$ that is consistent and asymptotically normal and it does not rely on any assumption on the functional form of the probability distribution of $e_i + u_{it}$. Describe the criterion function that this estimator optimizes.

ANSWER: Given that $e_i + u_{it}$ is independent of $\mathbf{X}_i$, we have that $median(e_i + u_{it} \mid \mathbf{X}_i) = median(e_i + u_{it}) = 0$. And given that $\max\{\ 0\ ;\ \mathbf{X}_i\ \boldsymbol{\pi}_t + (e_i + u_{it})\}$ is a monotonic function in $e_i + u_{it}$, we have that:

$$median(Y_{it} | \mathbf{X}_i) = \max\{\ 0\ ;\ \mathbf{X}_i\ \boldsymbol{\pi}_t\}$$

Based on this result we can define a Least Absolute Deviations (LAD) estimator of $\boldsymbol{\pi}_t$. Given the cross-section at period t the LAD estimator of $\boldsymbol{\pi}_t$ is defined as:

$$\widehat{\boldsymbol{\pi}}_t = \arg\min_{\boldsymbol{\pi}_t} \sum_{i=1}^{N} |Y_{it} - \max\{\ 0\ ;\ \mathbf{X}_i\ \boldsymbol{\pi}_t\}|$$

This estimator is $\sqrt{N}$-consistent and asymptotically normal.

Question 3.3. [5 points] Let $\widehat{\boldsymbol{\pi}} \equiv \{\widehat{\boldsymbol{\pi}}_t : t = 1, 2, ...T\}$ be the estimator of the parameters $\{\pi_t\}$ in Question 3.2, and let $\widehat{V}(\widehat{\boldsymbol{\pi}})$ be a consistent estimator of the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\pi}}$. The vector $\widehat{\boldsymbol{\pi}}$ contains T^2K parameters, and we want to use $\widehat{\boldsymbol{\pi}}$ to get consistent estimates of the $(T+1)K$ parameters in β and $\{\lambda_t\}$. β and $\{\lambda_t\}$ are over-identified. Explain how to use $\widehat{\boldsymbol{\pi}}$ and $\widehat{V}(\widehat{\boldsymbol{\pi}})$ to construct a consistent (and efficient) estimator of β and $\{\lambda_t\}$, and of the variance-covariance matrix of this estimator.

ANSWER: The linear relationship between the $T^2K \times 1$ vector $\boldsymbol{\pi}$ and the $(T+1)K \times 1$ vector $\theta = (\lambda', \beta')'$ described in Question 3.1 can be written as a system of linear equations:

$$\boldsymbol{\pi} = \boldsymbol{M}\ \theta$$

where $\boldsymbol{M}$ is a known matrix of zeroes and ones with dimension $T^2K \times (T+1)K$. Therefore, we can write the regression-like equation between the estimators:

$$\widehat{\boldsymbol{\pi}} = \boldsymbol{M}\ \theta + v$$

where v is a vector of error terms with the estimation error of $\widehat{\boldsymbol{\pi}}$, i.e., $v \equiv \widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}$. Given this regression-like equation, a 'natural' estimator of θ is the OLS estimator:

$$\widehat{\theta}_{OLS} = \left(\boldsymbol{M}'\boldsymbol{M}\right)^{-1} \boldsymbol{M}'\widehat{\boldsymbol{\pi}}$$

and a consistent estimator of the asymptotic variance-covariance matrix of $\widehat{\theta}_{OLS}$ is:

$$\widehat{\mathbb{V}}(\widehat{\theta}_{OLS}) = \left(\boldsymbol{M}'\boldsymbol{M}\right)^{-1} \boldsymbol{M}'\ \widehat{\mathbb{V}}(\widehat{\boldsymbol{\pi}})\ \boldsymbol{M}\ \left(\boldsymbol{M}'\boldsymbol{M}\right)^{-1}$$

But given that we have $\widehat{\mathbb{V}}(\widehat{\boldsymbol{\pi}})$, we can also construct a FGLS estimator of θ, that is more efficient than the OLS. The FGLS is defined as:

$$\widehat{\theta}_{FGLS} = \left(\boldsymbol{M}' \left[\widehat{\mathbb{V}}(\widehat{\boldsymbol{\pi}})\right]^{-1} \boldsymbol{M}\right)^{-1} \boldsymbol{M}' \left[\widehat{\mathbb{V}}(\widehat{\boldsymbol{\pi}})\right]^{-1} \widehat{\boldsymbol{\pi}}$$

and a consistent estimator of the asymptotic variance-covariance matrix of $\widehat{\theta}_{FGLS}$ is:

$$\widehat{\mathbb{V}}(\widehat{\theta}_{FGLS}) = \left(\boldsymbol{M}' \left[\widehat{\mathbb{V}}(\widehat{\boldsymbol{\pi}})\right]^{-1} \boldsymbol{M}\right)^{-1}$$