ECONOMETRICS II (ECO 2401S)
University of Toronto. Department of Economics. Winter 2015
Instructor: Victor Aguirregabiria

FINAL EXAM. Monday, April 13, 2015. 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 (25 points). Consider N firms that start their operation in an industry at the same period $t = 1$. Let q_{it} represent the logarithm of output for firm i at period t. The researcher proposes the following model for the dynamics of log-output. At the first period in the firm's lifetime, $t = 1$, log-output is,

$$q_{i1} = \gamma_1 + \alpha_i + u_{i1},$$

and at any period $t \geq 2$, the model is:

$$q_{it} = \beta \; q_{i,t-1} + \gamma_t + \alpha_i + u_{it}$$

where: γ_t is a parameter (time fixed-effect) that is common to all the firms at period t; α_i is a firm-specific fixed effect that represents persistent heterogeneity between firms' productivity, with $E(\alpha_i) = 0$ and $V(\alpha_i) = \sigma_\alpha^2$; u_{it} is a productivity shock that is independently distributed over firms and over time, and independent of α_i, with $E(u_{it}) = 0$ and $V(u_{it}) = \sigma_u^2$; and $\beta \in (0, 1)$ is a parameter that captures structural state dependence in productivity, e.g., learning-by-doing.

Suppose that a researcher has a panel dataset of firms' log-output, where N is large and the sample covers only the first three periods in the firms' lifetime: $t = 1, 2, 3$. Assume that there is not attrition (i.e., no firm's exit from the sample) such that the researcher has a balanced panel of N firms over three periods of time.

(a) [5 points] In the equation in levels at $t = 3$, consider an IV estimator of β where $q_{i,t-1}$ is instrumented using $\Delta q_{i,t-1}$. Write the expression of this IV estimator.

ANSWER: The regression equation is: $q_{i3} = \beta \; q_{i2} + \gamma_3 + \alpha_i + u_{i3}$. The IV estimator of β is:

$$\widehat{\beta} = \frac{\sum_{i=1}^{N} \Delta q_{i2} \, (q_{i3} - \overline{q}_3)}{\sum_{i=1}^{N} \Delta q_{i2} \, (q_{i2} - \overline{q}_2)}$$

(b) [10 points] Obtain the limit in probability of this IV estimator as N goes to infinity. Obtain the expression of this "plim" as a function of β, σ_α^2, and σ_u^2. Is this estimator consistent? Why/Why not? Explain the result.

ANSWER: By the LLN, the limit in probability of the IV estimator is:

$$\begin{aligned} p\lim_{N\to\infty}\widehat{\beta} &= \frac{\mathbb{E}\left([\Delta q_{i2}-\mathbb{E}(\Delta q_{i2})]\ \ [q_{i3}-\mathbb{E}(q_{i3})]\right)}{\mathbb{E}\left([\Delta q_{i2}-\mathbb{E}(\Delta q_{i2})]\ \ [q_{i2}-\mathbb{E}(q_{i2})]\right)} \\ &= \beta+\frac{\mathbb{E}\left([\Delta q_{i2}-\mathbb{E}(\Delta q_{i2})]\ \ [\alpha_i+u_{i3}]\right)}{\mathbb{E}\left([\Delta q_{i2}-\mathbb{E}(\Delta q_{i2})]\ \ [q_{i2}-\mathbb{E}(q_{i2})]\right)} \end{aligned}$$

To obtain these covariances, first note that:

$$\begin{aligned} &q_1-\mathbb{E}(q_1)=\alpha+u_1 \\ &q_2-\mathbb{E}(q_2)=(1+\beta)\alpha+u_2+\beta u_1 \\ &\Delta q_2-\mathbb{E}(\Delta q_2)=\beta\alpha+u_2+(\beta-1)u_1 \end{aligned}$$

Therefore,

$$\mathbb{E}\left([\Delta q_2-\mathbb{E}(\Delta q_2)]\ \ [\alpha+u_3]\right)=\mathbb{E}\left([\beta\alpha+u_2+(\beta-1)u_1]\ \ [\alpha+u_3]\right)=\beta\ \sigma_\alpha^2$$

And:

$$\mathbb{E}\left([\Delta q_2-\mathbb{E}(\Delta q_2)]\ \ [q_2-\mathbb{E}(q_2)]\right)=\mathbb{E}\left([\beta\alpha+u_2+(\beta-1)u_1]\ \ [(1+\beta)\alpha+u_2+\beta u_1]\right)$$

$$=\beta(1+\beta)\ \sigma_\alpha^2+[1+\beta(\beta-1)]\ \ \sigma_u^2$$

Such that,

$$\begin{aligned} p\lim_{N\to\infty}\widehat{\beta} &= \beta+\frac{\beta\ \sigma_\alpha^2}{\mathbb{E}\left([\Delta q_2-\mathbb{E}(\Delta q_2)]\ \ [q_2-\mathbb{E}(q_2)]\right)} \\ &= \beta+\frac{\beta\ \sigma_\alpha^2}{\beta(1+\beta)\ \sigma_\alpha^2+[1+\beta(\beta-1)]\ \ \sigma_u^2} \end{aligned}$$

The estimator is inconsistent. This is because the instrument Δq_{i2} depends on the individual effect α_i. Note that this instrument is the same as in the Blundell-Bond (BB) estimator of dynamic panel data models. However, in this model the BB instrument is not valid because the equation at periods $t=1,2,3$ do not satisfy the stationarity condition that implies that Δq_{it} does not depend on the fixed effect.

(c) [10 points] Suppose that the panel dataset still covers three time periods but the three sample periods are $\{t-2,t-1,t\}$ with $t\to\infty$, i.e., sample observations occur many time periods after the N firms started to operate in this industry. Again, assume that the panel is balanced. Consider the IV estimator of β where $q_{i,t-1}$ is instrumented using $\Delta q_{i,t-1}$. Does this change in the starting period of the sample have any influence on the consistency of this IV estimator of β ? Why/Why not? Explain.

ANSWER: The estimator has the same definition, and by the LLN we still have that:

$$\begin{aligned} p\lim_{N\to\infty}\widehat{\beta} &= \frac{\mathbb{E}\left([\Delta q_{t-1}-\mathbb{E}(\Delta q_{t-1})]\ \ [q_t-\mathbb{E}(q_t)]\right)}{\mathbb{E}\left([\Delta q_{t-1}-\mathbb{E}(\Delta q_{t-1})]\ \ [q_{t-1}-\mathbb{E}(q_{t-1})]\right)} \\ &= \beta+\frac{\mathbb{E}\left([\Delta q_{t-1}-\mathbb{E}(\Delta q_{t-1})]\ \ [\alpha+u_t]\right)}{\mathbb{E}\left([\Delta q_{t-1}-\mathbb{E}(\Delta q_{t-1})]\ \ [q_{t-1}-\mathbb{E}(q_{t-1})]\right)} \end{aligned}$$

However, now we have that, when $t\to\infty$,

$$q_t-\mathbb{E}(q_t)=\frac{\alpha}{1-\beta}+u_t+\beta\ u_{t-1}+\beta^2\ u_{t-2}+...$$

such that $\Delta q_{t-1} - \mathbb{E}(\Delta q_{t-1})$ does not depend on α and depends only on the transitory shocks at $t-1$ and before. Therefore,

$$\mathbb{E}\left([\Delta q_{t-1} - \mathbb{E}(\Delta q_{t-1})] \;\; [\alpha + u_t] \right) = 0$$

and $p\lim_{N\to\infty} \widehat{\beta} = \beta$. For $t \to \infty$, the initial conditions of the stochastic process of q_t do not matter, Blundell-Bond stationarity conditions hold, and the estimator is consistent.

PROBLEM 2 (25 points). Consider the binary choice model $Y = 1\{Z + \beta\ X - \varepsilon \leq 0\}$, where ε is independently 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 the Smooth Maximum Score Estimator of the parameter β.

ANSWER: The SMSE is defined as the value of β that maximizes the Smooth score function. That is:

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

where $\Phi(.)$ is a continuously differentiable CDF (e.g., the CDF of the standard normal), and $b_n > 0$ is a bandwidth parameter.

(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: Define the Conditional Choice Probability (CCP) function, $P(z_0, x_0) \equiv \Pr(Y = 1|Z = z, X = x)$. Given the random sample $(y_i, x_i, z_i : i = 1, 2, ...n)$ this CCP is nonparametrically identified, i.e., can be estimated consistently without a parametric assumption on the CDF F. Then, we can treat this CCP function as known (identified) to the researcher for any value of (z_0, x_0). According to the model, we have that:

$$P(z_0, x_0) = \Pr\left(\varepsilon \leq z_0 + \beta\ x_0\right) = F\left(z_0 + \beta\ x_0\right)$$

Let $\varepsilon_0 \in \mathbb{R}$ be an arbitrary value of ε in the real line. Given this value ε_0, and given x_0, we can construct the following value of Z, that we define as $z^*(\varepsilon_0, x_0)$:

$$z^*(\varepsilon_0, x_0) \equiv \varepsilon_0 - \beta\ x_0$$

Note that $z^*(\varepsilon_0, x_0)$ is known to the researcher because β is known (i.e., it has been estimated consistently in a first step). Also, note that by construction:

$$z^*(\varepsilon_0, x_0) + \beta\ x_0 = \varepsilon_0$$

Therefore,

$$F\left(\varepsilon_0\right) = F\left(z^*(\varepsilon_0, x_0) + \beta\ x_0\right) = P(z^*(\varepsilon_0, x_0), x_0)$$

Since $P(.,.)$ and $z^*(.,.)$ are known, it is clear that $F\left(\varepsilon_0\right)$ is identified. Note that we can repeat this procedure for any value $\varepsilon_0 \in \mathbb{R}$, such that we can nonparametrically identify the CDF $F(.)$ over all its support.

PROBLEM 3 (25 points). Consider a random coefficients multinomial choice model with $J+1$ choice alternatives $\{0, 1, ..., J\}$. The utility of alternative choice j for individual n is $U_{jn} = Z_j\ \beta_n + \varepsilon_{jn}$, where: Z_j is a $1 \times K$ vector of observable attributes of alternative j; β_n is a $K \times 1$ vector of random coefficients that is i.i.d. over individuals with a normal distribution $N(\beta, \Omega)$, and Ω is a diagonal matrix with elements $\{\sigma_1^2, \sigma_2^2, ..., \sigma_K^2\}$; and ε_{jn} is an unobservable that is i.i.d. over n 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 $\{y_n : i = 1, 2, ..., N\}$ with $y_n \in \{0, 1, ..., J\}$. We are interested in using this sample to estimate the vector of parameters $\theta = \{\beta, \Omega\}$. The main challenge in estimation of this model comes from the solution of the multiple integration problem associated to the computation of choice probabilities. The researcher uses Monte Carlo simulation to deal with this problem.

(a) [5 points] Propose a simulator of the choice probability $P_j(\theta) \equiv \Pr(Y = j|\theta)$. Show that your simulator is asymptotically unbiased, and it is consistent as the number of Monte Carlo simulations, R, goes to infinity.

ANSWER:

[Description of the CCP] We can write, $\beta_n = \beta + \Omega^{1/2}\ v_n$, where $v_n = (v_{1n}, v_{2n}, ..., v_{Kn})'$ is a $K \times 1$ vector of independent standard normals. By definition of the choice probability, we have that:

$$P_j(\theta) = \int \frac{\exp\left\{Z_j\ \beta + \sum_{k=1}^{K} \sigma_k Z_{kj} v_k\right\}}{\sum_{i=0}^{J} \exp\left\{Z_i\ \beta + \sum_{k=1}^{K} \sigma_k Z_{ki} v_k\right\}} \prod_{k=1}^{K} \phi(v_k) dv_k$$

where $\phi(.)$ is the PDF of the standard normal. It is convenient to represent this expression in a compact form as:

$$P_j(\theta) = \mathbb{E}_v\left[\Lambda_j(v)\right]$$

where $\mathbb{E}_v[.]$ represents the expectation over the distribution of v_n, and $h_j(v)$ is the logit probability inside this integral.

[Description of the simulator] We approximate this K-dimensional integral by using Monte Carlo simulation. Let $\{v_{kr} : k = 1, 2, ..., K;\ r = 1, 2, ..., R\}$ be $R * K$ independent random draws from a standard normal distribution. The simulator of $P_j(\theta)$ is defined as:

$$\widetilde{P}_j^R(\theta) = \frac{1}{R} \sum_{r=1}^{R} \Lambda_j(v_r) = \frac{1}{R} \sum_{r=1}^{R} \frac{\exp\left\{Z_j\ \beta + \sum_{k=1}^{K} \sigma_k Z_{kj} v_{kr}\right\}}{\sum_{i=0}^{J} \exp\left\{Z_i\ \beta + \sum_{k=1}^{K} \sigma_k Z_{ki} v_{kr}\right\}}$$

That is, we replace the population expectation $\mathbb{E}_v\left[.\right]$ with the "empirical" expectation $\frac{1}{R}\sum_{r=1}^{R}(.)$.

[Asymptotically unbiased and consistent simulator] MATHIEU: NOTE THAT THE CLAIM IS THAT IT IS ASYMPTOTICALLY UNBIASED, NOT UNBIASED FOR FINITE R. IT WOULD BE GOOD TO MENTION THIS DURING THE EXAM BECAUSE IT IS NOT CLEAR IN THE ENUNCIATE.

It is clear that by LLN:

$$p \lim_{R \to \infty} \widetilde{P}_j^R(\theta) = p \lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} = \mathbb{E}_v\left[\Lambda_j(v_r)\right] = P_j(\theta)$$

Also, by CLT:

$$\sqrt{R}\left(\widetilde{P}_j^R(\theta) - P_j(\theta)\right) = \frac{1}{\sqrt{R}} \sum_{r=1}^{R} \left[\Lambda_j(v_r) - P_j(\theta)\right] \to_d N(0,)\left(0, Var\left[\Lambda_j(v)\right]\right)$$

(b) [10 points] Propose a Simulated Method of Moments (SMM) estimator of θ. Describe the implementation of this estimator.

ANSWER: The SMM is defined as the value of θ that solves the sample moment conditions:

$$\frac{1}{N}\sum_{n=1}^{N}\sum_{j=0}^{J} W_j(Z)\ \left[1\{y_n = j\} - \widetilde{P}_{jn}^{R}(\theta)\right] = 0$$

where $W_j(Z)$ are vectors of functions of $Z_j's$ with the same dimension as the vector of parameters θ, i.e., $2K \times 1$. To implement this estimator, we start drawing $N * R * K$ independent random draws from the standard normal that we use to construct the simulator $\widetilde{P}_{jn}^{R}(\theta)$. At each iteration in the search for the SMM estimator, we use always these same random draws. This means that during the implementation of the estimator, the simulators $\widetilde{P}_{jn}^{R}(\theta)$ are deterministic functions of θ. This is important for the numerical and statistical properties of the estimator. Then, we can use a Newton's method to solve for the solution of this system of equations.

(c) [10 points] Show that this SMM estimator is consistent as N goes to infinity and R is fixed.

ANSWER: The SMM $\widehat{\theta}_{N,R}$ satisfies the conditions:

$$\frac{1}{N}\sum_{n=1}^{N}\sum_{j=0}^{J} W_j(Z)\ \left[1\{y_n = j\} - \widetilde{P}_{jn}^{R}(\widehat{\theta}_{N,R})\right] = 0$$

For arbitrary θ, we can write the simulator as, $\widetilde{P}_{jn}^{R}(\theta) = P_j(\theta) + e_{jn}^{R}(\theta)$, where $e_{jn}^{R}(\theta)$ is the simulation error. By the properties of the simulator, described in point (a), and by the fact that we calculate an independent simulator for each observation n, we have that, for N going to infinity and fixed R is fixed.

$$\frac{1}{N}\sum_{n=1}^{N} e_{jn}^{R}(\theta) \rightarrow_p 0, \text{ uniformly in } \theta$$

Therefore, we have that:

$$\begin{aligned} & \frac{1}{N}\sum_{n=1}^{N}\sum_{j=0}^{J} W_j(Z)\ \left[1\{y_n = j\} - \widetilde{P}_{jn}^{R}(\theta)\right] \\ = & \frac{1}{N}\sum_{n=1}^{N}\sum_{j=0}^{J} W_j(Z)\ \left[1\{y_n = j\} - P_j(\theta)\right] + \frac{1}{N}\sum_{n=1}^{N} e_{jn}^{R}(\theta) \end{aligned}$$

The first term converges in probability and uniformly in θ to $E(\sum_{j=0}^{J} W_j(Z)\ [1\{y_n = j\} - P_j(\theta)])$, and the second term converges in probability and uniformly in θ to zero. Therefore:

$$\frac{1}{N}\sum_{n=1}^{N}\sum_{j=0}^{J} W_j(Z)\left[1\{y_n = j\} - \widetilde{P}_{jn}^{R}(\theta)\right] \ \rightarrow\ {}_pE\left(\sum_{j=0}^{J} W_j(Z)\ 1\left[\{y_n = j\} - P_j(\theta)\right]\right)$$

$$\text{, uniformly in } \theta$$

Given the identification assumption that θ_0 uniquely solves the system $E\left(\sum_{j=0}^{J} W_j(Z)\ 1\left[\{y_n = j\} - P_j(\theta)\right]\right) = 0$, we have that $\widehat{\theta}_{N,R}$ is consistent as N goes to infinity and R is fixed.

PROBLEM 4 (25 points). Consider a Treatment Effects model with potential outcomes (Y_0, Y_1), treatment dummy $D \in \{0,1\}$, and outcome variable Y, such that $Y = (1-D)\ Y_0 + D\ Y_1$. Let $Z \in \{0,1\}$ be a random variable that represents whether the individual is eligible to treatment. This Z variable comes from a randomized experiment such that: (1) Z is independent of potential outcomes (Y_0, Y_1); and (2) Z is correlated with treatment, i.e., $\Pr(D=1|Z=1) > \Pr(D=1|Z=0)$. The researcher has a random sample of N individuals with information on $\{y_i, d_i, z_i : i = 1, 2, ..., N\}$. The main parameter of interest is the Average Treatment Effect defined as $ATE \equiv E(Y_1 - Y_0)$.

(a) [5 points] Based on this model, show that we can write a simple linear regression equation of Y on D where the slope parameter is the ATE. What is the structure of the error term in this regression? Is the error term correlated with D? Explain.

ANSWER: We can write $Y_0 = \mu_0 + U_0$ and $Y_1 = \mu_1 + U_1$, where $\mu_0 \equiv E(Y_0)$ and $\mu_1 \equiv E(Y_1)$ such that by construction $E(U_0) = E(U_1) = 1$. Also, note that, by definition, $ATE = \mu_1 - \mu_0$. Using these definitions, we have that:

$$
\begin{aligned}
Y &= (1-D)\ Y_0 + D\ Y_1 = Y = (1-D)\ \ (\mu_0 + U_0) + D\ \ (\mu_1 + U_1) \\
&= \mu_0 + (\mu_1 - \mu_0)\ D + U_0 + (U_1 - U_0)\ D \\
&= \alpha + \beta\ D + e
\end{aligned}
$$

where $\alpha = \mu_0$, $\beta = \mu_1 - \mu_0 = ATE$, and $e = U_0 + (U_1 - U_0)\ D$. Without further assumptions, the unobservable component of the potential outcomes, U_0 and U_1, can be correlated with the treatment dummy D and this implies correlation between the error term e and the regressor D. We can derive $Cov(D, e)$ and show that it is not zero. Alternatively, we have that no correlation between e and D requires that $E(e|D=0) = E(e|D=1)$. We can show that this condition does not hold. We have that:

$$E(e|D=0) = E(U_0 + D(U_1 - U_0)\ |\ D=0) = E(U_0|D=0)$$

And

$$E(e|D=1) = E(U_0 + D(U_1 - U_0)\ |\ D=1) = E(U_1|D=1)$$

Without further assumptions, we have that U_0 and U_1 can depend on treatment D, such that $E(U_0|D=0) \neq E(U_1|D=1)$. Therefore, without further assumptions, D and e can be correlated.

(b) [10 points] Consider the IV estimation of the regression model in (a) where the treatment dummy D is instrumented using the eligibility dummy Z. Present the expression for this IV estimator. Derive the limit in probability of this IV estimator under conditions (1) and (2), and show that it is an inconsistent estimator of the ATE.

ANSWER: By definition, this IV estimator is:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(z_i - \overline{z})y_i}{\sum_{i=1}^{n}(z_i - \overline{z})d_i}$$

By the LLN, $\widehat{\beta}$ converges in probability to $\dfrac{Cov(Z,Y)}{Cov(Z,D)}$. Since $Y = \alpha + \beta\ D + e$, we have that $\widehat{\beta}$ converges in probability to $\beta + \dfrac{Cov(Z,e)}{Cov(Z,D)} = ATE + \dfrac{Cov(Z,e)}{Cov(Z,D)}$. We now derive $Cov(Z,e)$. Note that,

$$Cov(Z,e) = (1 - P_Z)\ E(e|D = 0) + P_Z\ E(e|D = 1)$$

where $P_Z = \Pr(Z = 1)$. Define $P_D(z) = \Pr(D = 1|Z = z)$. Then,

$$\begin{aligned} E(e|Z &= 0) = E(U_0 + D(U_1 - U_0) \mid Z = 0) \\ &= P_D(0)\ E(U_1 - U_0|D = 1) \end{aligned}$$

And,

$$\begin{aligned} E(e|Z &= 1) = E(U_0 + D(U_1 - U_0) \mid Z = 1) \\ &= P_D(1)\ E(U_1 - U_0|D = 1) \end{aligned}$$

Therefore,

$$\begin{aligned} Cov(Z,e) &= (1 - P_Z)\ P_D(0)\ E(U_1 - U_0|D = 1) + P_Z\ P_D(1)\ E(U_1 - U_0|D = 1) \\ &= [(1 - P_Z)\ P_D(0) + P_Z\ P_D(1)]\ \ E(U_1 - U_0|D = 1) \end{aligned}$$

That is equal to zero if and only if $E(U_1 - U_0|D = 1)$. Therefore, this is an inconsistent estimator of the ATE.

(c) [10 points] Let D_0 and D_1 be the binary variables that represent the potential treatment of an individual when she is not eligible to treatment ($Z = 0$) and when she is eligible ($Z = 1$). By definition, $D = (1 - Z)\ D_0 + Z\ D_1$. Suppose that for every individual, we have that: (3) $D_1 \geq D_0$. Show that under conditions (1) to (3), the IV estimator converges in probability to the Local Average Treatment Effect parameter defined as $LATE \equiv E\left(Y_1 - Y_0 \mid D_1 > D_0\right)$.

ANSWER: First, we rewrite the IV estimator using its representation as the Wald estimator.

$$\widehat{\beta} = \frac{\overline{y}_1 - \overline{y}_0}{\overline{d}_1 - \overline{d}_0}$$

where $\overline{y}_1$ and $\overline{d}_1$ are the sample means of Y and D, respectively, for the subsample of observations with $Z = 1$, and similarly, $\overline{y}_0$ and $\overline{d}_0$ are the sample means of Y and D for the subsample of observations with $Z = 0$. By the LLN, $\widehat{\beta}$ converges in probability to $\dfrac{E(Y|Z = 1) - E(Y|Z = 0)}{E(D|Z = 1) - E(D|Z = 0)}$. Now, we show that, under assumptions (1) to (3), this expression is equal to $LATE \equiv E\left(Y_1 - Y_0 \mid D_1 > D_0\right)$. Note that $Y = Y_0 + D\ (Y_1 - Y_0)$, and $D = D_0 + Z\ (D_1 - D_0)$. Therefore,

$$\begin{aligned} E(Y|Z = 1) &= E(Y_0 + D_1(Y_1 - Y_0)\ |Z = 1) \\ \text{(by independence of } Z \text{ with } (Y_0, Y_1, D_0, D_1)) &= E(Y_0 + D_1(Y_1 - Y_0)\ ) \end{aligned}$$

And

$$\begin{aligned} E(Y|Z = 0) &= E(Y_0 + D_1(Y_1 - Y_0)\ |Z = 0) \\ \text{(by independence of } Z \text{ with } (Y_0, Y_1, D_0, D_1)) &= E(Y_0 + D_0(Y_1 - Y_0)\ ) \end{aligned}$$

Therefore, the numerator of the PLIM of IV is:

$$\begin{aligned}\text{Numerator of PLIM of IV} &= E(Y_0 + D_1(Y_1 - Y_0)\ ) - E(Y_0 + D_0(Y_1 - Y_0)\ ) \\ &= E((D_1 - D_0)\ (Y_1 - Y_0)\ )\end{aligned}$$

By the monotonicity assumption, $(D_1 - D_0)$ can be only 0 or 1. Therefore,

$$\text{Numerator of PLIM of IV} = \Pr(D_1 - D_0 > 0)\ E(Y_1 - Y_0 \mid D_1 - D_0 > 0)$$

Similarly, for the denominator of the PLIM of IV we have that:

$$\begin{aligned}E(D|Z=1) &= E(D_0 + Z(D_1 - D_0)\ |Z=1) \\ \text{(by independence of } Z \text{ with } (D_0, D_1)) &= E(D_1)\end{aligned}$$

And

$$\begin{aligned}E(D|Z=0) &= E(D_0 + Z(D_1 - D_0)\ |Z=0) \\ \text{(by independence of } Z \text{ with } (D_0, D_1)) &= E(D_0)\end{aligned}$$

The denominator of the PLIM of IV is:

$$\text{Denominator of PLIM of IV} = E(D_1 - D_0)$$

Again, by the monotonicity assumption, $(D_1 - D_0)$ can be only 0 or 1, such that $E(D_1 - D_0) = \Pr(D_1 - D_0 > 0)$. Therefore,

$$\begin{aligned}\text{PLIM of IV} &= \frac{\Pr(D_1 - D_0 > 0)\ E(Y_1 - Y_0 \mid D_1 - D_0 > 0)}{\Pr(D_1 - D_0 > 0)} \\ &= E(Y_1 - Y_0 \mid D_1 - D_0 > 0) = LATE\end{aligned}$$