Paper 10, IIB: Answers to questions asked in the microeconometrics supervisions

Eoghan O'Neill*

22 February 2020

1 Answers to questions asked by students in supervision 5

1.1 Question: In a linear regression if the variable of interest is exogenous, but one of the control variables is endogenous, is the OLS estimate of the coefficient of the variable of interest consistent?

Answer: No, unless the variable of interest is uncorrelated with the endogenous variable.

Suppose the true model is:

$$Y = X_1\beta_1 + X_2\beta_2 + Z\gamma + \varepsilon$$

Z is an unobserved variable. Let $Cov(X_1, Z) = 0$ (therefore X_1 is exogenous), and $Cov(X_2, Z) \neq 0$ (therefore X_2 is endogenous).

Then

$$plim\hat{\beta}_1 = \beta_1 + \gamma\frac{Cov(X_1^*, Z)}{Var(X_1^*)} = \beta_1$$

where $X_1^* = [I - X_2(X_2'X_2)^{-1}X_2']X_1$.

Therefore the estimator is consistent if $Cov(X_1^*, Z) = 0$. This is true if $Cov(X_1, X_2) = 0$, because $X_1^* = X_1$.

However, $Cov(X_1, X_2) = 0$ is rarely true (unless, for example X_1 is a randomized treatment variable and X_2 is something observed before the random treatment allocation). (X_2 would only be included to reduce the standard errors).

Note: If $Cov(X_1, Z) = 0$ and $Cov(X_1, X_2) = 0$, then the estimator is consistent without including the endogenous variable X_2.

*Please let me know of any mistakes or typos epo21@cam.ac.uk.

1.2 Note on the strict exogeneity assumption

Please note that the strict exogeneity assumption is different depending on whether the model explicitly contains unobserved effects.

For models with unobserved effects, the strict exogeneity assumption involves conditioning on the fixed effect. (We say that the $\{\boldsymbol{x_{it}} : t = 1, 2, ..., T\}$ are strictly exogeneous conditional on the unobserved effect c_i.

1.3 Question: Why do we condition on the time constant unobserved effect in the strict exogeneity assumption? $E[u_{it}|\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T}, c_i] = 0 \ \forall \ t, i$

Answer: Under this assumption, conditioning on c_i allows us to control for factors that are time constant, and therefore by removing the time constant unobserved effect we can identify a causal relationship between the $\mathbf{x}_i$ and the dependent variable.

It could be the case that without controlling for the time constant unobserved effect $\mathbf{x}_{i,s}$ has a partial effect on $y_{i,t}$ for $s \neq t$.

Note: Wooldridge describes the assumption $E[u_{it}|\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T}, c_i] = 0 \ \forall \ t, i$ as "the $\{\mathbf{x}_{i,t} : t = 1, 2, ..., T\}$ are strictly exogenous conditional on the unobserved effect c_i.

The assumption $E[u_{it}|\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T}, c_i] = 0 \ \forall \ t, i$ can also be written as

$E[y_{it}|\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T}, c_i] = \mathbf{x}_{i,t}\boldsymbol{\beta} + c_i \ \forall \ t, i$

As explained in Wooldridge (2010) section 10.2.2, this is a more reasonable assumption than $E[u_{it}+c_i|\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T}] = 0 \ \forall \ t, i$, because this (and a functional form assumption) would imply

$$E[y_{it}|\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T}] = \mathbf{x}_{i,t}\boldsymbol{\beta}$$

Wooldridge gives an example to show how this is unlikely to hold. Suppose y_{it} is output of a soybean farm and $\mathbf{x}_{i,t}$ is a vector of inputs. Output in year t is likely to be correlated with inputs in other years if we do not control for c_i.

The assumption $E[u_{it}|\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T}, c_i] = 0 \ \forall \ t, i$ implies (by the law of iterated expectations)

$$E[y_{it}|\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T}] = \mathbf{x}_{i,t}\boldsymbol{\beta} + E[c_i|\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T}]$$

and therefore the assumption $E[y_{it}|\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T}] = \mathbf{x}_{i,t}\boldsymbol{\beta} \ \forall \ t, i$ (which was implied by $E[u_{it} + c_i|\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T}] = 0 \ \forall \ t, i$) fails if $E[c_i|\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T}] \neq E[c_i]$.

1.4 Question: Why are the terms $\bar{\bar{y}}$, $\bar{\bar{\mathrm{x}}}$, and $\bar{\bar{u}}$ included in the within transformation that removes individual and time fixed effects?

Suppose we first remove the individual fixed effects to obtain $y_{it} - \bar{y}_i$ for all i and t.

Then we remove the average across time of $y_{it} - \bar{y}_i$ for each individual, and then rearrange.

$$y_{it}-\bar{y}_i-\frac{1}{T}\sum_{t=1}^{T}(y_{it}-\bar{y}_i)=y_{it}-\bar{y}_i-\frac{1}{T}(\sum_{t=1}^{T}y_{it})-(-\frac{1}{T}\sum_{t=1}^{T}\bar{y}_i)$$

$$=y_{it}-\bar{y}_i-\frac{1}{T}(\sum_{t=1}^{T}y_{it})+\frac{1}{T}\sum_{t=1}^{T}\bar{y}_i=y_{it}-\bar{y}_i-\frac{1}{T}(\sum_{t=1}^{T}y_{it})+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{n=1}^{N}y_{it}$$

$$=y_{it}-\bar{y}_i-\frac{1}{T}(\sum_{t=1}^{T}y_{it})+\frac{1}{NT}\sum_{t=1}^{T}\sum_{n=1}^{N}y_{it}=y_{it}-\bar{y}_t-\bar{y}_i+\bar{\bar{y}}$$

The other variables are similarly transformed, and we can obtain:

$$y_{it}-\bar{y}_t-\bar{y}_i+\bar{\bar{y}}=(\mathbf{x}_{it}-\bar{\mathbf{x}}_t-\bar{\mathbf{x}}_i+\bar{\bar{\mathbf{x}}})'\boldsymbol{\gamma}+u_{it}-\bar{u}_t-\bar{u}_i+\bar{\bar{u}},$$

where $\bar{z}_t=N^{-1}\sum_{i=1}^{N}z_{it}$, $\bar{z}_i=T^{-1}\sum_{t=1}^{T}z_{it}$ and $\bar{\bar{z}}=(NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}z_{it}$ for $z_{it}=\{y_{it},\mathbf{x}_{it},u_{it}\}$. This transformation eliminates the individual and time fixed effect.

Note that it is possible to simply include dummy variables for time periods, but the above transformation might be preferable because the estimation of a regression with a large number of variables (i.e. when T is large and dummy variables are included for time fixed effects) might be computationally infeasible.

1.5 Question: Under the Random Effects assumptions, how do we obtain estimates $\hat{\sigma}_v^2$ and $\hat{\sigma}_c^2$, where $\sigma_v^2=\sigma_c^2+\sigma_u^2$?

Answer: This is explained in Wooldridge (2010) pages 295 and 296.

Let $\check{v}_{it}$ denote the pooled OLS residuals. A consistent estimator of σ_v^2 is

$$\hat{\sigma}_v^2=\frac{1}{(NT-K)}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{v}_{it}^2$$

A consistent estimator of σ_c^2 is

$$\hat{\sigma}_c^2=\frac{1}{[NT(T-1)/2-K]}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\check{v}_{it}\check{v}_{is}$$

Then we can estimate σ_u^2 by $\hat{\sigma}_u^2=\hat{\sigma}_v^2-\hat{\sigma}_c^2$.

There are other methods for estimating σ_u^2 and σ_c^2. See Wooldridge (2010) section 10.5

1.6 Question: If the data is non-stationary, then there is potentially a spurious regression problem for the FE estimator as $T\to\infty$. Is there a potential spurious regression problem for the First Differences estimator as $T\to\infty$?

PLEASE NOTE: The properties of panel estimators as $T\to\infty$ is not an examinable topic.

Answer: This is discussed in detail in Wooldridge (2010) section 10.7.1

Suppose the strict exogeneity assumption is violated, but we maintain contemporaneous exogeneity $E[\mathbf{x}'_{it}u_{it}] = \mathbf{0}$, then the FE estimator generally has inconsistency that shrinks to zero at rate $1/T$, while the inconsistency of the FD estimator is essentially independent of T. This suggests that the FE estimator should be preferred.

However, this result depends on the idiosyncratic errors being an $I(0)$ sequence, i.e. y_{it} and $\mathbf{x}_{it}$ must be cointegrated. If there is a spurious regression problem, then the above result for the FE estimator does not hold. The problem becomes more acute for the FE estimator as $T \to \infty$. The FD approach removes any unit roots in y_{it} and $\mathbf{x}_{it}$ and so spurious regression is not an issue (but the lack of strict exogeneity might be).

1.7 Question: How do you obtain the coefficient estimates using the moment conditions given in the answer to question 2?

Answer: The coefficients can be estimated similarly to the Arellano-Bond estimator described in lecture slides 17-28 of topic 2.ii. However, the $\mathbf{z}_{it}$ vectors (see slide 21) will not include lags of the dependent variable or the first differences of the exogenous variables, but instead include $\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,T}$. This gives a $\boldsymbol{Z}_i$ matrix with $NK(T-1)$ columns corresponding to the moment conditions given in the solution to question 2.

The coefficients are estimated by $\hat{\boldsymbol{\beta}} = (\tilde{\boldsymbol{X}}'ZW_NZ'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'ZW_NZ'\Delta\mathbf{y}$.

This can also be written similarly to how Cameron and Trivedi (2005) describe the Arellano-Bond estimator. See Cameron and Trivedi (2005) section 22.5.3 and equation (22.35).

$$\hat{\boldsymbol{\beta}} = [(\sum_{i=1}^{N} \tilde{\boldsymbol{X}}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^{N} \boldsymbol{Z}_i'\tilde{\boldsymbol{X}}_i)]^{-1}(\sum_{i=1}^{N} \tilde{\boldsymbol{X}}_i'\boldsymbol{Z}_i)\boldsymbol{W}_N(\sum_{i=1}^{N} \boldsymbol{Z}_i'\Delta\mathbf{y})$$

where $\tilde{\boldsymbol{X}}_i$ is a $(T-1)\times(K+1)$ matrix with tth row $(\Delta y_{i,t-1}, \Delta\mathbf{x}'_{it})$, and $\Delta\mathbf{y}$ is a $(T-1_{\times}1)$ vector with tth row Δy_{it} and $\boldsymbol{Z}_i$ is a $T \times NK(T-1)$ matrix (but possibly with less columns if not all leads and lags of the exogenous regressors are used).

$$\boldsymbol{Z}_i = \begin{bmatrix} \mathbf{z}'_{i2} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}'_{i3} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{z}'_{iT} \end{bmatrix} = \begin{bmatrix} (\mathbf{x}'_{i,1}, ..., \mathbf{x}'_{i,T}) & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & (\mathbf{x}'_{i,1}, ..., \mathbf{x}'_{i,T}) & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & (\mathbf{x}'_{i,1}, ..., \mathbf{x}'_{i,T}) \end{bmatrix}$$

where $\mathbf{z}'_{it} = (\mathbf{x}'_{i,1}, ..., \mathbf{x}'_{i,T})\ \forall t$

Note: The moment conditions may be written as $E[Z_i\Delta\mathbf{v}_i] = 0$.

1.8 Note on independence, zero conditional mean, orthogonality and zero covariance.

For two random variables X and y, the assumption that X and Y are independent is a stronger assumption than $E[Y|X] = E[Y]$ (mean independence).

Furthermore, the assumption $E[Y|X] = E[Y]$ is a stronger assumption than $E[XY] = E[X]E[Y]$. i.e. It can be shown by the law of iterated expectations that $E[Y|X] = E[Y] \Rightarrow E[XY] = E[X]E[Y]$ however $E[XY] = E[X]E[Y] \nRightarrow E[Y|X] = E[Y]$.

Also, note that $E[XY] = E[X]E[Y] \Leftrightarrow Cov(X,Y) = 0$

$E[XY] = 0 \Rightarrow Cov(X,Y) = 0$, but $Cov(X,Y) = 0 \nRightarrow E[XY] = 0$ unless we also assume that $E[X] = 0$ or $E[Y] = 0$.

Now consider a linear model $Y = \beta_0 + \beta_1 X + u$, where $E[u] = 0$

Then X and u are independent $\Rightarrow E[u|X] = 0 \Rightarrow E[uX] = 0 \Rightarrow Cov(u,X) = 0$

but $E[uX] = 0 \nRightarrow E[u|X] = 0 \nRightarrow X$ and u are independent

For an example of variables that are mean independent, but not independent, consider the example in which X is uniform on [-1,1], and $Y = X^2$. (If you require $E[Y|X] = 0$, let $Y = (X^2) - (1/3)$).

1.9 A note on implications of strict exogeneity

For consistency of POLS, the assumptions $E[\boldsymbol{x_{is}} u_{it}] = \mathbf{0}\ \forall s,t = 1,...,T$ and $E[\boldsymbol{x_{is}} c_i] = \mathbf{0}\ \forall s,t = 1,...,T$ are required.

For RE, the assumption of strict exgeneity $E[u_{it}|\boldsymbol{x}_i, c_i] = 0\ \forall t = 1,2,...,T$ and $E[c_i|\boldsymbol{x}_i] = E[c_i] = 0$ (or the weaker assumption $E[\boldsymbol{x_{is}} c_i] = \mathbf{0}\ \forall s,t = 1,...,T$) are required for consistency. Note that for consistency of the RE estimator (a GLS estimator), a stricter assumption is required than for POLS. See section 10.4.1 of Wooldridge (2010) for an explanation.

For consistency of FE we do not require $E[c_i|\boldsymbol{x}_i] = E[c_i] = 0$, but we require strict exogeneity $E[u_{it}|\boldsymbol{x}_i, c_i] = 0\ \forall t = 1,2,...,T$.

Consider the transformed equation

$$y_{it} - \bar{y}_i = (\boldsymbol{x}_{it} - \bar{x}_i)\boldsymbol{\beta} + u_{it} - \bar{u}_u$$

or

$$\ddot{y}_{it} = \ddot{\boldsymbol{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it} \quad t = 1,2,...,T$$

For the OLS estimator to be consistent, we require the ususal assumption $E[\ddot{\boldsymbol{x}}'_{it}\ddot{u}_{it}] = \mathbf{0}$, $t = 1,2,...,T$. This is implied by the strict exogeneity assumption (therefore the FE estimator is consistent).

Note that the weaker assumption $E[\boldsymbol{x}_{it}' u_{it}] = \mathbf{0} \; \forall t$ is not sufficient for consistenct because this does not ensure that $\boldsymbol{x}_{is}$ is uncorrelated with u_{it} , $s \neq t$.

However, the following assumption is sufficient for consistency of the fixed effects estimator $E[\boldsymbol{x}_{it}'(v_{it} - \bar{v}_i)] = \mathbf{0}$, $t = 1, 2, ..., T$

1.10 Derivation of the non-matrix OLS formula from the matrix formula

Consider a linear model with one independent variable $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

The matrix formula for the OLS estimator is

$$\hat{\beta} = (X'X)^{-1} X'Y$$

Where X contains a column of 1s and a column of X_i values.

$$X'X = \begin{bmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{bmatrix}$$

$$\begin{aligned}(X'X)^{-1} &= \frac{1}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \begin{bmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{bmatrix} \\ &= \begin{bmatrix} \frac{\sum_{i=1}^n X_i^2}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} & -\frac{\sum_{i=1}^n X_i}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \\ -\frac{\sum_{i=1}^n X_i}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} & \frac{n}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \end{bmatrix}\end{aligned}$$

Also note that

$$\begin{aligned}\hat{Var}(X_i) = \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i^2 - 2\bar{X}X_i + \bar{X}^2) \\ &= \sum_{i=1}^n X_i^2 - 2\bar{X}\sum_{i=1}^n X_i + \sum_{i=1}^n \bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - 2\bar{X}n\bar{X} + n\bar{X}^2 = (\sum_{i=1}^n X_i^2) - n\bar{X}^2\end{aligned}$$

and $\sum_{i=1}^n X_i = n\bar{X}$ therefore $(\sum_{i=1}^n X_i)^2 = n^2\bar{X}^2$

Consider the term in the bottom right of $(X'X)^{-1}$

$$\begin{aligned}\frac{n}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} &= \frac{n}{n(\sum_{i=1}^n X_i^2 - \frac{1}{n}(\sum_{i=1}^n X_i)^2)} = \frac{n}{n(\sum_{i=1}^n X_i^2 - \frac{1}{n}n^2\bar{X}^2)} \\ &= \frac{1}{\sum_{i=1}^n X_i^2 - n\bar{X}^2} = \frac{1}{\widehat{Var}(X_i)}\end{aligned}$$

Also, note that

$$X'Y = \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{bmatrix}$$

Therefore the bottom right hand element of $(X'X)^{-1}X'Y$ will be

$$\hat{\beta}_1 = \begin{bmatrix} -\frac{\sum_{i=1}^n X_i}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} & \frac{n}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{bmatrix}$$

$$= -\frac{\sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} + \frac{n\sum_{i=1}^n X_i Y_i}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}$$

$$= \frac{n\sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} = \frac{\widehat{Cov}(X_i, Y_i)}{\widehat{Var}(X_i)}$$