AAEC/ECON 5126 FINAL EXAM: SOLUTIONS

SPRING 2014 / INSTRUCTOR: KLAUS MOELTNER

This exam is open-book, open-notes, but please work strictly on your own. Please make sure your name is on every sheet you're handing in. You have 120 minutes to complete this exam. You can collect a maximum of 50 points. Each question is scored as indicated below. Vectors are given in lower-case boldface. Matrices are written in upper-case boldface.

Question I (20 points): Measurement error

Consider the following linear regression model for individual i:

$$y_i = \beta_1 + \beta_2 x_{2i}^* + \epsilon_i \tag{1}$$

Assume that the usual properties of the classical linear regression model hold, except x_{2i}^* is measured with error. Specifically, for a sample of n observations, the relationship between the observed variable $\mathbf{x}_2$ and the true variable $\mathbf{x}_2^*$ is given via

$$\mathbf{x}_2 = \mathbf{x}_2^* + \alpha \mathbf{i}, \tag{2}$$

where α is a constant and $\mathbf{i}$ is a vector of 1's. That is, each observation on $\mathbf{x}_2^*$ is measured with the same constant error, for example due to poor calibration of the measuring instrument.

Part (a) *2 points*
For a single observation, write down the model in (8), using the relationship in (2). Let's call that the "empirical model" (the one you'd be using for analysis). Show the properties of the resulting error term (call it ω_i), i.e. its expectation and variance (you can assume that ϵ_i is homoskedastic with variance σ^2).

Solution:

$$\begin{aligned} y_i =& \beta_1 + \beta_2 \left(x_{2i} - \alpha\right) + \epsilon_i = \\ & \beta_1 + \beta_2 x_{2i} + \left(\epsilon_i - \alpha\beta_2\right) = \beta_1 + \beta_2 x_{2i} + \omega_i \end{aligned} \tag{3}$$

$$E\left(\omega_i\right) = -\alpha\beta_2, \quad V\left(\omega_i\right) = \sigma^2 \tag{4}$$

Part (b) *3 points*
Derive the covariance of x_{2i} and ω_i. Is there evidence of omitted variable bias?

Date: May 13, 2014.

Solution:

$$cov\left(x_{2i}, \omega_i\right) = cov\left(x_{2i}^* + \alpha, \epsilon_i - \alpha\beta_2\right) = 0 \tag{5}$$

since $cov\left(x_{2i}^*, \epsilon_i\right) = 0$ by the assumptions of the CLRM, and all other terms are fixed constants. So there is no indication of omitted variable bias.

Part (c) *6 points*
Write down the empirical model at the sample level. Using partitioned regression, show that the OLS estimate of β_2 is unbiased.

Solution:
See Q.3 of PS 2. This is again a situation where the error term has non-zero mean, but is well-behaved otherwise. In PS2, the nonzero expectation was μ, here it is $-\alpha\beta_2$.

Part (d) *6 points*
Now consider the same empirical model, but without an intercept. Show that in this case the estimate of β_2 is no longer unbiased.

Solution:
See Q.3 of PS 2.

Part (e) *3 points*
Using intuition (no math needed), would the estimate for β_2 be unbiased in absence of measurement error for the model without intercept? Why or why not?

Solution:
No - as long as the true model has an intercept, an empirical model that ignores the constant term will produce biased estimates for slope coefficients - as we have seen in the last part of PS2, Q.3.

Question II (30 points): Instrumental Variables / TSLS

Consider the following system of equations:

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \beta_2\mathbf{y}_2 + \boldsymbol{\epsilon}_1 \tag{6}$$

$$\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\pi} + \boldsymbol{\epsilon}_2 \tag{7}$$

where all matrices and vectors are of length n, and the column dimensions of $\mathbf{X}_1$ and $\mathbf{X}_2$ are k_1 and k_2, respectively, with $k_2 > k_1$. Your main goal is to consistently estimate $\boldsymbol{\beta} = \begin{bmatrix}\boldsymbol{\beta}_1' & \beta_2\end{bmatrix}'$, but you are concerned that the two error terms are correlated, making $\mathbf{y}_2$ an endogenous regressor in the first equation.

Your strategy is to use the OLS fitted values from the second equation (call them $\hat{\mathbf{y}}_2$) as instrument for $\mathbf{y}_2$ in a Two-Stage least Squares (TSLS) estimation of the first equation.

Throughout you can assume that $\mathbf{X}_1$ is orthogonal to $\mathbf{X}_2$ and $\mathbf{y}_2$ (i.e. $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$, $\mathbf{X}_1'\mathbf{y}_2 = \mathbf{0}$), and that neither $\mathbf{X}$ matrix is correlated with any of the error terms (i.e $E\left(\mathbf{X}_r'\boldsymbol{\epsilon}_s\right) = \mathbf{0}$, $r = 1, 2$, $s = 1, 2$).

Part (a) *5 points*
Denote the residuals from the OLS estimation of the *second* equation as $\mathbf{e}_2$. Show that they are orthogonal to $\mathbf{X}_1$. Show that the fitted values $\hat{\mathbf{y}}_2$ are also orthogonal to $\mathbf{X}_1$

Solution:

$$\begin{aligned}\mathbf{X}_1'\mathbf{e}_2 =&\mathbf{X}_1'\left(I - \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\right)\mathbf{y}_2 = \\ &\mathbf{X}_1'\mathbf{y}_2 - \mathbf{X}_1'\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{y}_2 = 0\end{aligned}$$

due to the orthogonality assumptions above. Similarly:

$$\begin{aligned}\mathbf{X}_1'\hat{\mathbf{y}}_2 =&\mathbf{X}_1'\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{y}_2 = 0 \\ &\text{since} \quad \mathbf{X}_1'\mathbf{X}_2 = 0 \quad \text{by assumption}\end{aligned}$$

Part (b) *5 points*
Show that $\hat{\mathbf{y}}_2'\hat{\mathbf{y}}_2 = \mathbf{y}_2'\hat{\mathbf{y}}_2 = \hat{\mathbf{y}}_2'\mathbf{y}_2$. (This will be needed later).

Solution:

$$\begin{aligned}\hat{\mathbf{y}}_2'\hat{\mathbf{y}}_2 =&\mathbf{y}_2'\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{y}_2 = \\ &\mathbf{y}_2'\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{y}_2 = \mathbf{y}_2'\hat{\mathbf{y}}_2 = \hat{\mathbf{y}}_2'\mathbf{y}_2\end{aligned}$$

Another correct (and faster) approach uses the projection matrix $\mathbf{P}_2 = \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'$ and its idempotency and symmetry property: $\mathbf{P}_2'\mathbf{P}_2 = \mathbf{P}_2 = \mathbf{P}_2'$.

Part (c) *5 points*
Let $\mathbf{X} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{y}_2\end{bmatrix}$ and $\mathbf{Z} = \begin{bmatrix}\mathbf{X}_1 & \hat{\mathbf{y}}_2\end{bmatrix}$.
Describe (in words or math) the regression model for the *first stage* of the TSLS procedure. Let $\hat{\mathbf{X}}$ be the fitted matrix from this model, and show that $\hat{\mathbf{X}} = \begin{bmatrix}\mathbf{X}_1 & \hat{\mathbf{y}}_2\end{bmatrix} = \mathbf{Z}$.

Solution:
Regress $\mathbf{X}$ on $\mathbf{Z}$, using OLS. Then:

$$\begin{aligned}&\hat{\mathbf{X}} = \mathbf{P}_Z\mathbf{X} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X} = \\ &\begin{bmatrix}\mathbf{X}_1 & \hat{\mathbf{y}}_2\end{bmatrix} \times \begin{bmatrix}\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\hat{\mathbf{y}}_2'\ \hat{\mathbf{y}}_2\right)^{-1}\end{bmatrix} \times \begin{bmatrix}\mathbf{X}_1'\mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{y}}_2'\mathbf{y}_2\end{bmatrix} = \\ &\begin{bmatrix}\mathbf{X}_1 & \hat{\mathbf{y}}_2\end{bmatrix} \times \begin{bmatrix}\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \left(\hat{\mathbf{y}}_2'\hat{\mathbf{y}}_2\right)^{-1}\hat{\mathbf{y}}_2'\mathbf{y}_2\end{bmatrix} = \\ &\begin{bmatrix}\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_1 & \hat{\mathbf{y}}_2\left(\hat{\mathbf{y}}_2'\hat{\mathbf{y}}_2\right)^{-1}\hat{\mathbf{y}}_2'\hat{\mathbf{y}}_2\end{bmatrix} = \\ &\begin{bmatrix}\mathbf{X}_1 & \hat{\mathbf{y}}_2\end{bmatrix}\end{aligned}$$

A clever and faster approach is to note that $\hat{\mathbf{X}} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}$, and then show that $\mathbf{Z}'\mathbf{X} = \mathbf{Z}'\mathbf{Z}$:

$$\mathbf{Z}'\mathbf{X} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{y}}_2'\mathbf{y}_2 \end{bmatrix}$$
$$\mathbf{Z}'\mathbf{Z} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{y}}_2'\hat{\mathbf{y}}_2 \end{bmatrix}$$

The two are equal since $\hat{\mathbf{y}}_2'\mathbf{y}_2 = \hat{\mathbf{y}}_2'\hat{\mathbf{y}}_2$. Then: $\hat{\mathbf{X}} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{Z} = \mathbf{Z}$.

Part (d) *5 points*
Describe, in words, the second stage of the TSLS procedure and solve for $\hat{\boldsymbol{\beta}}_{TSLS}$ in terms of $\mathbf{X}_1$,$\hat{\mathbf{y}}_2$, and $\mathbf{y}_1$.

Solution:
Regress $\mathbf{y}_1$ on $\hat{\mathbf{X}}$, that is on $\mathbf{X}_1$ and $\hat{\mathbf{y}}_2$.

$$\begin{aligned} &\hat{\boldsymbol{\beta}}_{TSLS} = \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}'\mathbf{y}_1 = \\ &\begin{bmatrix} \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\hat{\mathbf{y}}_2'\hat{\mathbf{y}}_2\right)^{-1} \end{bmatrix} \times \begin{bmatrix} \mathbf{X}_1'\mathbf{y}_1 \\ \hat{\mathbf{y}}_2'\mathbf{y}_1 \end{bmatrix} = \\ &\begin{bmatrix} \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{y}_1 \\ \left(\hat{\mathbf{y}}_2'\hat{\mathbf{y}}_2\right)^{-1}\hat{\mathbf{y}}_2'\mathbf{y}_1 \end{bmatrix} \end{aligned}$$

Part (e) *10 points*
Show that this TSLS approach for the estimation of $\boldsymbol{\beta}$ is equivalent to adding the second-equation residuals $\mathbf{e}_2$ to equation (3) as an additional regressor and using OLS, i.e. by estimating

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \beta_2\mathbf{y}_2 + \gamma\mathbf{e}_2 + \boldsymbol{\nu}_1 = \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{e}_2 + \boldsymbol{\nu}_1 \tag{8}$$

where $\mathbf{X}$ and $\boldsymbol{\beta}$ are defined as above.

Call this estimator $\tilde{\mathbf{b}}$.
(Hint: Use partitioned regression results. Start directly with the partitioned regression solution for $\tilde{\mathbf{b}}$, which will include a residual-maker matrix. Call that matrix $\mathbf{M}_e$. Note that $\mathbf{M}_e\mathbf{y}_2 = \hat{\mathbf{y}}_2$.)

Solution:

$$\begin{aligned} &\tilde{\mathbf{b}} = \left(\mathbf{X}'\mathbf{M}_e\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M}_e\mathbf{y}_1, \quad \text{where} \\ &\mathbf{M}_e = \mathbf{I} - \mathbf{e}_2\left(\mathbf{e}_2'\mathbf{e}_2\right)^{-1}\mathbf{e}_2' \end{aligned}$$

Now:

$$
\begin{aligned}
&\mathbf{M}_e\mathbf{X} = \mathbf{M}_e \begin{bmatrix}\mathbf{X}_1 & \mathbf{y}_2\end{bmatrix} = \\
&\begin{bmatrix}\mathbf{X}_1 - \mathbf{e}_2\left(\mathbf{e}_2'\mathbf{e}_2\right)^{-1}\mathbf{e}_2'\mathbf{X}_1 & \mathbf{y}_2 - \mathbf{e}_2\left(\mathbf{e}_2'\mathbf{e}_2\right)^{-1}\mathbf{e}_2'\mathbf{y}_2\end{bmatrix} = \\
&\begin{bmatrix}\mathbf{X}_1 & \hat{\mathbf{y}}_2\end{bmatrix} \quad \text{using} \quad \mathbf{y}_2 = \hat{\mathbf{y}}_2 + \mathbf{e}_2
\end{aligned}
$$

Then:

$$
\begin{aligned}
\left(\mathbf{X}'\mathbf{M}_e\mathbf{X}\right)^{-1} = \begin{bmatrix}\mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\hat{\mathbf{y}}_2 \\ \mathbf{y}_2'\mathbf{X}_1 & \mathbf{y}_2'\hat{\mathbf{y}}_2\end{bmatrix}^{-1} &= \\
\begin{bmatrix}(\mathbf{X}_1'\mathbf{X}_1)^{-1} & \mathbf{0} \\ \mathbf{0} & (\hat{\mathbf{y}}_2'\hat{\mathbf{y}}_2)^{-1}\end{bmatrix}&
\end{aligned}
$$

and

$$
\tilde{\mathbf{b}} = \begin{bmatrix}(\mathbf{X}_1'\mathbf{X}_1)^{-1} & \mathbf{0} \\ \mathbf{0} & (\hat{\mathbf{y}}_2'\hat{\mathbf{y}}_2)^{-1}\end{bmatrix} \times \begin{bmatrix}\mathbf{X}_1'\mathbf{y}_1 \\ \hat{\mathbf{y}_2}'\mathbf{y}_1\end{bmatrix}
$$

which yields the same solution as before.