Problem Set 5 - Solutions

Applied Statistics and Econometrics II
Spring 2018, NYU
Ercan Karadas

[1] a) Just note that it also holds that

$$y_{t-1} = \beta_1 + \beta_2 x_{t-1} + \beta_3 y_{t-2} + \varepsilon_{t-1}$$

Therefore, y_{t-1} and ε_t are correlated so y_{t-1} is not an exogenous variables and OLS cannot produce consistent estimates in this case.

b) What the econometrician is actually estimating is this

$$y_t = \beta_1 + \beta_2 \tilde{x}_t + \varepsilon_t \qquad (\star\star)$$

where $\varepsilon_t = v_t - \beta_2 u_t$. But note that $\tilde{x}_t$ and ε_t are clearly correlated through the term u_t, so OLS cannot produce consistent estimates. But in this exercise, we are also asked to find plim b_2. To that end,

$$b_2 = \frac{\sum_{t=1}^{n}(\tilde{x}_t - \bar{\tilde{x}})(y_t - \bar{y})}{\sum_{t=1}^{n}(\tilde{x}_t - \bar{\tilde{x}})^2}$$

Now substituting the econometrician's equation into this we obtain

$$b_2 = \beta_2 + \frac{\sum_{t=1}^{n}(\tilde{x}_t - \bar{\tilde{x}})(y_t - \bar{y})}{\sum_{t=1}^{n}(\tilde{x}_t - \bar{\tilde{x}})^2}$$

$$\begin{aligned} b_2 &= \beta_2 + \frac{\sum_{t=1}^{n}(\tilde{x}_t - \bar{\tilde{x}})(y_t - \bar{y})}{\sum_{t=1}^{n}(\tilde{x}_t - \bar{\tilde{x}})^2} \\ &= \beta_2 + \frac{\frac{1}{n}\sum_{t=1}^{n}(\tilde{x}_t - \bar{\tilde{x}})(y_t - \bar{y})}{\frac{1}{n}\sum_{t=1}^{n}(\tilde{x}_t - \bar{\tilde{x}})^2} \\ \implies \text{plim } b_2 &= \beta_2 + \frac{\text{plim } \frac{1}{n}\sum_{t=1}^{n}(\tilde{x}_t - \bar{\tilde{x}})(y_t - \bar{y})}{\text{plim } \frac{1}{n}\sum_{t=1}^{n}(\tilde{x}_t - \bar{\tilde{x}})^2} \\ &= \beta_2 + \frac{E[\tilde{x}_t \varepsilon_t]}{\text{Var}(\tilde{x}_t)} \end{aligned}$$

where we used $E[\varepsilon_t] = 0$ and the fact that as $n \to \infty$ the sample moments converge to the population moments. We can compute the expressions in the last term as follows

$$\begin{aligned} E[\tilde{x}_t \varepsilon_t] &= E[(x_t + u_t)(v_t - \beta_2 u_t)] = -\beta_2 \sigma_u^2 \\ \text{Var}[\tilde{x}_t] &= \text{Var}[x_t + u_t] = \sigma_x^2 + \sigma_u^2 \end{aligned}$$

Therefore we find

$$\begin{aligned}\text{plim } b_2 &= \beta_2 \left(1 - \frac{\sigma_u^2}{\sigma_u^2 + \sigma_x^2}\right)\\ \text{or } &= \beta_2 \left(1 - \frac{1}{1 + \frac{\sigma_x^2}{\sigma_u^2}}\right)\end{aligned}$$

From this expression it's clear that if $\frac{\sigma_x^2}{\sigma_u^2}$ is small (which might happen, for instance when the variance of measurement error is large) then $b_2 \xrightarrow{p} 0$. This downward bias called *attenuation bias*. This is an example of a specification bias. By assumption the relevant model is $(\star)$ but for data reasons we have had to use $(\star\star)$. The result is a flawed estimation procedure.

c) From $(\star\star)$ we have

$$\begin{aligned}\beta_1 &= y_t - \beta_2 \tilde{x}_t - \varepsilon_t\\ \implies \beta_1 &= E[y_t - \beta_2 \tilde{x}_t]\end{aligned}$$

Using $\bar{y} = b_1 + b_2 \bar{x}$ we can also write

$$\begin{aligned}\text{plim } (b_1 - \beta_1) &= \text{plim } (\bar{y} - b_2 \bar{\tilde{x}} - E[y_t] - \beta_2 E[\tilde{x}_t])\\ &= -\text{plim } (b_2 - \beta_2) E[\tilde{x}_t]\end{aligned}$$

So, unless $E[\tilde{x}_t] = 0$, we don't have $b_1 \xrightarrow{p} \beta_1$.

[2] Consider the simple model

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

and suppose that we would like to use z_i as an instrument for x_i (but z might be neither exogenous nor strongly relevant).

a) Show that $b_{2,IV}$ is not unbiased.

$$\begin{aligned}E[b_{2,IV}] &= E\left[\frac{\sum_{i=1}^n (y_i - \bar{y})(z_i - \bar{z})}{\sum_{i=1}^n (x_t - \bar{x})(z_i - \bar{z})}\right]\\ &= \ldots\\ &= \beta_1 + E_{X,Z}\left[\frac{\sum_{i=1}^n E[\varepsilon_i | x_i, z_i](z_i - \bar{z})}{\sum_{i=1}^n (x_t - \bar{x})(z_i - \bar{z})}\right]\\ &\neq \beta_1\end{aligned}$$

Instrument exogoneity implies $E[\varepsilon_i | z_i] = 0$, but not $E[\varepsilon_i | x_i, z_i] = 0$ (this would mean that $E[\varepsilon_i | x_i] = 0$ and we would not need an instrument) b)

$$\begin{aligned} b_{2,IV} &= \frac{\text{cov}(y_i, z_i)}{\text{cov}(x_i, z_i)} \\ &= \frac{\text{cov}(\beta_1 + \beta_2 x_i + \varepsilon_i, z_i)}{\text{cov}(x_i, z_i)} \\ &= \beta_2 + \frac{\text{cov}(\varepsilon_i, z_i)}{\text{cov}(x_i, z_i)} \\ &= \beta_2 + \frac{\text{corr}(z_i, \varepsilon_i)}{\text{corr}(z_i, x_i)} \cdot \frac{\sigma_\varepsilon}{\sigma_x} \end{aligned}$$

c) Use part (b). Take $\sigma_\varepsilon = \sigma_x$ for simplicity and suppose $\text{corr}(z_i, \varepsilon_i) = 0.1 \times \text{corr}(z_i, x_i)$ so that $\text{corr}(z, \varepsilon)$ is relatively low. Then $b_{2,IV} = \beta_2 + 10$ and this bias doesn't go away.

[3] a) In this exercise,

$$\begin{aligned} \mathbf{X} &= [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{y}_2 \quad \mathbf{y}_3]_{n \times 4} \equiv [\mathbf{X}_1 \quad \mathbf{X}_2] \\ \mathbf{Z} &= [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3 \quad \mathbf{x}_4]_{n \times 4} \equiv [\mathbf{X}_1 \quad \mathbf{Z}_1] \end{aligned}$$

Here the order of elements in $\mathbf{Z}$ doesn't matter but this choice clarifies that we have exactly identified case and more importantly we can partition $\mathbf{X}$ and $\mathbf{Z}$ into two parts as in the lecture notes and then the procedure is standard:

- **Step 1.** Regress each of the variables in $\mathbf{X}$ on $\mathbf{Z}$ to obtain a matrix of fitted values $\hat{\mathbf{X}}$,

$$\hat{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 & \hat{\mathbf{X}}_2 \end{bmatrix}$$

  where $\hat{\mathbf{X}}_2 = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}_2 = [\hat{\mathbf{y}}_2 \quad \hat{\mathbf{y}}_3]$.
- **Step 2.** Regress $\mathbf{y}_1$ on $\hat{\mathbf{X}}$ to obtain the estimated $\boldsymbol{\beta} = [\alpha_2 \quad \alpha_3 \quad \beta_1 \quad \beta_2]'$ vector

$$\mathbf{b}_{2SLS} = \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1} \hat{\mathbf{X}}'\mathbf{y}_1$$

b) The composite error is $\epsilon_1 = u_1 + \alpha_2 \hat{v}_2 + \alpha_3 \hat{v}_3$ and $\sum_{i=1}^{N} \hat{\epsilon}_{1i} \hat{y}_{2i} \neq 0$, because $\sum_{i=1}^{N} \hat{y}_{2i} \hat{v}_{3i} \neq 0$. The latter does not hold because $\sum_{i=1}^{N} x_{3i} \hat{v}_{3i} \neq 0$

[4] a) There is a simultaneity bias as P_t is endogenous, which can easily be seen by writing down the reduced form equation:

$$P_t = \frac{\alpha_0 - \beta_0}{\alpha_1 + \beta_1} + \frac{\alpha_2}{\alpha_1 + \beta_1} X_t + \frac{1}{\alpha_1 + \beta_1} (u_{1t} - u_{2t})$$

for which we used the equilibrium condition $D_t = S_t$.

b)

$$\begin{aligned} b_{1,OLS} &= \frac{\text{cov}(P_t, S_t)}{\text{cov}(P_t, P_t)} \\ &= \frac{\text{cov}(P_t, \beta_0 + \beta_1 P_t + u_{2t})}{\text{var}(P_t)} \\ &= \beta_1 + \frac{\text{cov}(P_t, u_{2t})}{\text{var}(P_t)} \\ &= \beta_1 + \frac{\alpha_2}{\alpha_1 + \beta_1} \frac{\text{cov}(X_t, u_{2t})}{\text{var}(P_t)} + \frac{1}{\alpha_1 + \beta_1} \frac{\text{cov}(u_{1t} - u_{2t}, u_{2t})}{\text{var}(P_t)} \end{aligned}$$

These are all sample moments but as $n \to \infty$ they converge to the population moments where we know that $E[X_t u_{2t}] = 0$ since X_t can be treated exogenous in the system and then only the last term remains as the bias:

$$b_{1,OLS} \xrightarrow{p} \beta_1 + \frac{1}{\alpha_1 + \beta_1} \frac{E[u_{1t} u_{2t}] - \text{var}(u_{2t})}{\text{var}(P_t)}$$

[5] Recall our the formula for the IV estimator

$$\mathbf{b}_{IV} = \left(\sum_{i=1}^{N} \mathbf{z}_i \mathbf{x}_i' \right)^{-1} \sum_{i=1}^{N} \mathbf{z}_i y_i$$

Since both $\mathbf{x}$ and $\mathbf{z}$ are scalars in this case, the IV estimator simply reduces to

$$b_{IV} = \frac{\sum_{i=1}^{N} z_i y_i}{\sum_{i=1}^{N} z_i x_i} = \frac{\sum_{i=1}^{N} y_i \mathbf{I}_{z_i}}{\sum_{i=1}^{N} x_i \mathbf{I}_{z_i}}$$

where the indicator function $\mathbf{I}_{z_i} = 1$ when $z_i = 1$ and 0 otherwise.

[6] Discussed in class.

[7] a) In the second stage we are estimating the model

$$y_i = \theta \hat{x}_i^2 + v_i$$

where $\hat{x}_i = \hat{\gamma} z_i$ and

$$\hat{\gamma} = \frac{\sum_{i=1}^{N} z_i y_i}{\sum_{i=1}^{N} z_i^2}$$

is estimated in the first stage.

Therefore, the OLS estimator for θ is the estimator described in the question:

$$\begin{aligned} \hat{\beta} &= \frac{\sum_{i=1}^{N} \hat{x}_i^2 y_i}{\sum_{i=1}^{N} \hat{x}_i^4} \\ &= \frac{\sum_{i=1}^{N} z_i^2 y_i}{\hat{\gamma}^2 \sum_{i=1}^{N} z_i^4} \end{aligned}$$

Here everything is sample data as desired.

b) First note that $\hat{\gamma} \xrightarrow{p} \gamma$ as the OLS assumptions are satisfied there, in particular it is assumed that $E[z_i u_i] = 0$.

$$\begin{aligned}
\hat{\beta} &= \frac{\sum_{i=1}^{N} z_i^2 y_i}{\hat{\gamma}^2 \sum_{i=1}^{N} z_i^4} \\
&\xrightarrow{p} \frac{E[z_i^2 y_i]}{\gamma^2 E[z_i^4]} \qquad \left(WLLN \text{ and } \hat{\gamma} \xrightarrow{p} \gamma\right) \\
&= \frac{E[z_i^2 \left(\beta x_i^2 + \varepsilon_i\right)]}{\gamma^2 E[z_i^4]} \\
&= \frac{E\left[z_i^2 \left(\beta \left(\gamma z_i + u_i\right)^2 + \varepsilon_i\right)\right]}{\gamma^2 E[z_i^4]} \\
&= \beta + \frac{1}{\gamma^2 E[z_i^4]} \left(2\beta\gamma E[z_i^3 u_i] + \beta E[z_i^2 u_i^2] + E[z_i^2 \varepsilon_i]\right)
\end{aligned}$$

c) In general, it's not a consistent estimator for β. However, under certain conditions consistency can be established. A set of such conditions: z_i and u_i are independent, $E[u_i] = 0$, and $E[z_i^2 \varepsilon_i] = 0$.

[8] a) The exogeneity assumption

$$E[\mathbf{X}'\varepsilon] = 0$$

provides us a population moment condition that can be used to estimate $\boldsymbol{\beta}$.

b) First express the population moment condition as

$$E[\mathbf{X}'\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)] = 0$$

Then the corresponding sample analog should be

$$\frac{1}{N}[\mathbf{X}'\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)] = 0$$

c) In the previous part, solving for $\boldsymbol{\beta}$ gives the standard OLS estimator.

[9] a) See GMM notes

b) Here we need 2 moment conditions as we are interested in estimating two population parameters δ and γ. If we have more than two instruments we might be able to have a more efficient estimator.

c) For a given set of moment conditions means, these are the given instruments so increasing the number of instruments is not an option. However, if we can increase our sample size we might gain efficiency because remember that all the efficiency results that we have been talking are large sample results, so increasing the sample size should increase the efficiency. If instruments are very weak, then increasing the sample size wouldn't help.

d) When we have more moment conditions than the number of parameters to be estimated the model is overidentified, and those additional moment conditions are called *overidentifying restrictions.* In general, having more moments than needed is a good think. First, as we discussed in part (b) we can increase the efficiency of the estimator by exploiting the information contained in the moment conditions by computing an optimal weighting matrix. Second, remember that in our discussion of instrumental variables we said that the exogeneity of instruments can not be tested when $R = K$. However, when $R > K$ the model is overidentified and there is information available which may be used to test this assumption. These statements apply to GMM as well.

e) For this part read Sargan Test in IV notes. Basic idea is the following: when the model is exactly identified it is not possible to check the validity of the moment conditions. However, when the model is overidentified, researchers often use tests of the overidentifying restrictions to assess the validity of the moment conditions. One of such tests is Sargan test. In this test, under the null hypothesis that all instruments are uncorrelated with the error term, the test has a large-sample $\chi^2(R-K)$ distribution where $R-K$ is the number of overidentifying restrictions. If the test is rejected, the specification of the model is rejected in the sense that the sample evidence is inconsistent with the *joint validity* of all R moment conditions. Without additional information it is not possible to determine which of the moments are incorrect, that is, which of the instruments are invalid.

[10] a) Consumer's optimization problem is exactly the same as in the lecture notes where we derived Euler equation as

$$E_t\left\{\delta U'(C_{t+s+1})(1+r_{t+1+s})\right\} = U'(C_{t+s})$$

In this case, $U'(C_t) = 1/C_t$ and therefore Euler equation reads

$$E_t\left\{\delta\left(\frac{1}{C_{t+s+1}}\right)(1+r_{t+1+s})\right\} = \frac{1}{C_{t+s}}$$

b) The unconditional moment conditions are

$$E\left\{\left[\delta\left(\frac{C_t}{C_{t+1}}\right)(1+r_{t+1}) - 1\right]z_t\right\} = 0$$

The corresponding sample analog is

$$\frac{1}{T}\sum_{t=1}^{T}\left[\delta\left(\frac{C_t}{C_{t+1}}\right)^{-\gamma}(1+r_{t+1}) - 1\right]z_t = 0$$

c) In this case, we have only one parameter to estimate δ so we need only one moment condition. For example, we could simply choose $z_t = 1$ above and obtain an estimator for δ.

d) See GMM notes

[11] a) Need to show to these two conditions are satisfied: (i) $f(x) \geq 0$ in the interval that the function is nonzero, then also show that $\int_0^{2/\theta}\left(\theta - \frac{\theta^2}{2}x\right)dx = 1$.

b) $E[X] = \int_0^{2/\theta} x\left(\theta - \frac{\theta^2}{2}x\right) dx = \frac{2}{3\theta}$.

c) Find the method of moments estimator for θ. To find MOM estimator for θ, let's write the moment conditions as $E[X - \frac{2}{3\theta}] = 0$ and then the sample analog for this

$$\frac{1}{N}\sum_{i=1}^{N} x_i - \frac{2}{3\theta} = 0$$

which gives the MOM estimator as

$$\hat{\theta}_{MOM} = \frac{2N}{3\sum_{i=1}^{N} x_i}$$