SGPE Econometrics 1 Tutorial 2 Answer Guide

October 27, 2021

1 Question 1

The OLS sampling error is, in data matrix form, $\left(\hat{\beta}_{OLS} - \beta\right) = (\mathrm{X}'\mathrm{X})^{-1}\,\mathrm{X}'\varepsilon$. Rewrite this in sample moment form. Use it to show that the OLS estimator $\hat{\beta}_{oLS}$ is consistent.

Answer:
We know from the first the sampling error $\hat{\beta} - \beta$ is

$$\hat{\beta} - \beta = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\epsilon \tag{1}$$

We first starting writing the in a sample moment form

$$\hat{\beta} - \beta = \left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\frac{1}{n}\mathbf{X}'\epsilon \tag{2}$$

From the lecture notes we can define

$$\frac{1}{n}\mathbf{X}'\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' = \mathbf{S}_{xx}, \quad \left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1} = \mathbf{S}_{xx}^{-1}$$
$$\frac{1}{n}\mathbf{X}'\epsilon = \overline{\mathbf{g}}_n$$

Thus, the OLS sampling error, in sample moment form, is

$$\hat{\beta} - \beta = \mathbf{S}_{xx}^{-1}\overline{\mathbf{g}}_n$$

Now want to prove that the OLS estimator is consistent which entails

$$p\lim_{n\longrightarrow\infty}(\hat{\beta} - \beta) = 0 \tag{3}$$

That is:

$$p\lim_{n\longrightarrow\infty}(\mathbf{S}_{xx}^{-1}\overline{\mathbf{g}}_n) = 0 \tag{4}$$

The first step is apply Hayashi Lemma 2.3(a) (or Slutsky's theorem), for example if we have two scalar sequences $\{x_n\} \underset{p}{\longrightarrow} \delta$ and $\{y_n\} \underset{p}{\longrightarrow} \gamma$, then $x_n y_n \underset{p}{\longrightarrow} \delta\gamma$. Using Slutsky's theorem, we can get:

$$p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) = p \lim_{n \longrightarrow \infty} (\mathbf{S}_{xx}^{-1} \overline{\mathbf{g}}_n) = p \lim_{n \longrightarrow \infty} \left(\mathbf{S}_{xx}^{-1}\right) p \lim_{n \longrightarrow \infty} \left(\overline{\mathbf{g}}_n\right)$$

Next, we apply the strong law of large numbers (SLLN) where the sample mean converges to the population mean $\left(\Sigma_{xx}^{-1}\right.$ and $\left.E\left(\mathbf{g}_i\right)\right)$ and the continuous mapping theorem, that is $x_n \underset{p}{\longrightarrow} x \Rightarrow f\left(x_n\right) \underset{p}{\longrightarrow} f(x)$, (note we assumes that both $\mathbf{y}$ and $\mathbf{X}$ are jointly stationary and ergodic)

$$\mathbf{S}_{xx}^{-1} \underset{a.s}{\longrightarrow} \Sigma_{xx}^{-1}, \Rightarrow \mathbf{S}_{xx}^{-1} \underset{p}{\longrightarrow} \Sigma_{xx}^{-1}$$

$$\overline{\mathbf{g}}_n \underset{a.s}{\longrightarrow} E\left(\overline{\mathbf{g}}_i\right), \Rightarrow \overline{\mathbf{g}}_n \underset{p}{\longrightarrow} E\left(\mathbf{g}_i\right)$$

Note remember **Almost sure convergence** implies **Convergence in probability** which is what the above two terms is showing. Thus,

$$p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) = \Sigma_{xx}^{-1} E\left(\mathbf{g}_i\right) \tag{5}$$

Note, under the large sample distribution assumption for OLS, we assumed that Σ_{xx} is non-singular (which implies the inverse exist) and weak exogeneity, that is $E\left(\mathbf{g}_i\right) = 0$ Therefore,

$$p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) = \Sigma_{xx}^{-1} 0$$

$$p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) = 0 \Leftrightarrow, \hat{\beta} \underset{p}{\longrightarrow} \beta$$

Hence, the OLS estimator is consistent.

2 Question 2

Show that the asymptotic variance of the OLS estimator is $\Sigma_{xx}^{-1} \mathbf{S} \Sigma_{xx}^{-1}$. Include in your answer an explanation of what Σ_{xx} and S are. Show how the expression for AVar $\left(\widehat{\beta}_{OLS}\right)$ simplifies under conditional homoskedasticity and independence.

Answer:
We want to prove that

$$\sqrt{n}(\hat{\beta} - \beta) \underset{d}{\longrightarrow} N\left(0, \Sigma_{xx}^{-1} \mathbf{S} \Sigma_{xx}^{-1}\right)$$

The first step of this proof is we want to use the sampling error

$$\hat{\beta} - \beta = \mathbf{S}_{xx}^{-1} \overline{\mathbf{g}}_n$$

Next, we multiply both sides with $\sqrt{n}$

$$\sqrt{n}(\hat{\beta}-\beta)=\sqrt{n}\mathbf{S}_{xx}^{-1}\overline{\mathbf{g}}_n$$

Note, multiplying $\sqrt{n}$ on both sides ensure that we do not converge to a degenerate distribution, and it converges to a well-defined distribution.

Note, from the proof in consistency we already know that

$$\mathbf{S}_{xx}^{-1}\underset{p}{\longrightarrow}\Sigma_{xx}^{-1}$$

Now, for $\sqrt{n}\overline{\mathbf{g}}_n$, we apply assumption 2.5 and using the central limit theorem we get

$$\sqrt{n}\overline{\mathbf{g}}_n\underset{d}{\longrightarrow}N(0,\mathbf{S})$$

where $\mathbf{S}=E\left[\mathbf{g}_i\mathbf{g}_i'\right]$ is the asymptotic variance of $\overline{\mathbf{g}}_n$.

Then applying the Slutsky theorem (or Hayashi Lemma 2.4) where $\mathbf{x}_n\underset{d}{\longrightarrow}\mathbf{x},\mathbf{x}\sim N(0,\Sigma)$, and $\mathbf{A}_n\underset{p}{\longrightarrow}\mathbf{A}$ then $\mathbf{A}_n\mathbf{x}_n\underset{d}{\longrightarrow}N\left(0,\mathbf{A}\Sigma\mathbf{A}'\right)$. In this proof, we can assume $\mathbf{A}_n=\mathbf{S}_{xx}^{-1}$ and $\mathbf{x}_n=\sqrt{n}\overline{\mathbf{g}}_n$. Then:

$$\sqrt{n}\mathbf{S}_{xx}^{-1}\overline{\mathbf{g}}_n\underset{d}{\longrightarrow}N\left(0,\Sigma_{xx}^{-1}\mathbf{S}\Sigma_{xx}^{-1}\right)$$

since where assume that the Σ_{xx} is symmetric matrix, then $\Sigma_{xx}^{-1}=\Sigma_{xx}^{-1'}$. Thus, we have proved that

$$\sqrt{n}(\hat{\beta}-\beta)\underset{d}{\longrightarrow}N\left(0,\Sigma_{xx}^{-1}\mathbf{S}\Sigma_{xx}^{-1}\right)$$

If we assume conidtional homoskedasticity and independence, then

$$\mathbf{S}=E\left[\mathbf{g}_i\mathbf{g}_i'\right]=E\left[\epsilon_i^2\mathbf{x}_i\mathbf{x}_i'\right]=E\left[\epsilon_i^2\right]E\left[\mathbf{x}_i\mathbf{x}_i'\right]=\sigma^2\Sigma_{xx}$$

Thus,

$$\sqrt{n}(\hat{\beta}-\beta)\underset{d}{\longrightarrow}N\left(0,\Sigma_{xx}^{-1}\sigma^2\Sigma_{xx}\Sigma_{xx}^{-1}\right)$$

since $\Sigma_{xx}\Sigma_{xx}^{-1}=\mathbf{I}$

$$\sqrt{n}(\hat{\beta}-\beta)\underset{d}{\longrightarrow}N\left(0,\sigma^2\Sigma_{xx}^{-1}\right)$$

3 Question 3

Derive the OLS estimator $\widehat{\beta}_{OLS}=\left(\mathrm{X}'\mathrm{X}\right)^{-1}\mathrm{X}'\mathrm{y}$ as a Method of Moments estimator.

Answer:

The basic intuition behind the method of moments (MM) is that you want to set the population parameter moments to the sample moments. In this section, we want to derive the OLS estimator using MM. Let's first start by recappping the weak exogeneity assumption

$$E(\mathbf{x}_i\epsilon_i) = E(\mathbf{x}_i(y_i - \mathbf{x}_i\beta)) = E(\mathbf{g}_i) = 0$$

This weak exogeneity assumption can also called the population moment condition. Next, we want to derive the corresponding sample moment for the above equation. Let's first define some terms

$$\hat{\epsilon}_i = y_i - \mathbf{x}_i\hat{\beta}_0$$

where $\hat{\epsilon}_i$ is the estimated residuals and $\hat{\beta}_0$ is an estimator of β. Using the above term, we can derive the sample moment

$$\overline{\mathbf{g}}\left(\hat{\beta}_0\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\left(\hat{\beta}_0\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\hat{\epsilon}_i = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\left(y_i - \mathbf{x}_i\hat{\beta}_0\right)$$

We can rewrite the above term is matrix form

$$\overline{\mathbf{g}}\left(\hat{\beta}_0\right) = \frac{1}{n}\mathbf{X}'\hat{\epsilon} = \frac{1}{n}\mathbf{X}'\left(\mathbf{y} - \mathbf{X}\hat{\beta}_0\right) = \frac{1}{n}\mathbf{X}'\mathbf{y} - \frac{1}{n}\mathbf{X}'\mathbf{X}\hat{\beta}_0$$

Now for the MM, want to set

$$\overline{\mathbf{g}}\left(\hat{\beta}_0\right) = E(\mathbf{g}_i)$$

Given the population moment condition we have

$$\overline{\mathbf{g}}\left(\hat{\beta}_0\right) = 0$$

$$\frac{1}{n}\mathbf{X}'\mathbf{y} - \frac{1}{n}\mathbf{X}'\mathbf{X}\hat{\beta}_0 = 0$$

$$\frac{1}{n}\mathbf{X}'\mathbf{y} = \frac{1}{n}\mathbf{X}'\mathbf{X}\hat{\beta}_0$$

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\hat{\beta}_0$$

Pre-multiply both sides by $(\mathbf{X}'\mathbf{X})^{-1}$,

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\beta}_0$$

Since $(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X}) = \mathbf{I}$

$$\hat{\beta}_0 = \hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

which is the OLS estimator.

4 Question 4

Derive the IV estimator in the exactly-identified case $\widehat{\beta}_{IV} = (Z'\mathrm{X})^{-1} Z'y$ as a Method of Moments estimator.

Answer:
With instrumental variables (IV), we have some linear regression function form, but now have a $n \times L$ matrix $\mathbf{Z}$ of instruments. Recapping, $\mathbf{X}$ is a $n \times K$ matrix of regressors. L is denoted as the no. of instruments and K is the no. of regressors. IV is used when there are endogenity presence in the regressors. Similar to the OLS weak exogeneity assumption, there is also a condition associated for the instruments

$$E(\mathbf{z}_i \epsilon_i) = E(\mathbf{z}_i (y_i - \mathbf{x}_i \beta)) = E(\mathbf{g}_i) = 0$$

Note $\mathbf{z}_i$ can interpreted as the vector the instruments for the i observation. In a similar fashion, we can use the estimated residuals $\hat{\epsilon}_i$ and an estimator of β, $\hat{\beta}_0$, to derive the corresponding sample moment for the above term

$$\overline{\mathbf{g}}\left(\hat{\beta}_0\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_i\left(\hat{\beta}_0\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i \hat{\epsilon}_i = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i\left(y_i - \mathbf{x}_i \hat{\beta}_0\right)$$

We can rewrite the above term is matrix form

$$\overline{\mathbf{g}}\left(\hat{\beta}_0\right) = \frac{1}{n}\mathbf{Z}'\hat{\epsilon} = \frac{1}{n}\mathbf{Z}'\left(\mathbf{y} - \mathbf{X}\hat{\beta}_0\right) = \frac{1}{n}\mathbf{Z}'\mathbf{y} - \frac{1}{n}\mathbf{Z}'\mathbf{X}\hat{\beta}_0$$

Next, if we assume $L = K$, that is the no. of instruments exactly equal the no. of regressors, then we can use MM to derive the IV estimator. First, we set the set population moment equal to sample moment

$$\overline{\mathbf{g}}\left(\hat{\beta}_0\right) = E(\mathbf{g}_i)$$

Note, from the population moment condition, we have

$$\begin{aligned}
\overline{\mathbf{g}}\left(\hat{\beta}_0\right) &= 0 \\
\frac{1}{n}\mathbf{Z}'\mathbf{y} - \frac{1}{n}\mathbf{Z}'\mathbf{X}\hat{\beta}_0 &= 0 \\
\frac{1}{n}\mathbf{Z}'\mathbf{y} &= \frac{1}{n}\mathbf{Z}'\mathbf{X}\hat{\beta}_0 \\
\mathbf{Z}'\mathbf{y} &= \mathbf{Z}'\mathbf{X}\hat{\beta}_0
\end{aligned}$$

Pre-multiply both sides by $(\mathbf{Z}'\mathbf{X})^{-1}$,

$$(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{X}\hat{\beta}_0$$

Since $(\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{Z}'\mathbf{X}) = \mathbf{I}$

$$\hat{\beta}_0 = \hat{\beta}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$$

which is the IV estimator.

5 Question 5

The sampling error of the GMM estimator is, in data matrix form, $\left(\widehat{\beta}_{GMM} - \beta\right) = (X'ZW_nZ'X)^{-1} X'ZW_nZ'\varepsilon$. Rewrite this in sample moment form and use it to show that the GMM estimator $\widehat{\beta}_{GMM}$ is consistent.

Some Basic:

When we have the no. of instruments greater than the no. of regressor $L > K$, we have an overidentified model and MM cannot work. We have to use Generalised method of moments (GMM). In the GMM approach, we use this weighting matrix $\mathbf{W}_n$ and rewrite the moment conditions in quadratic form. Thus, we derive the GMM estimator by

$$\hat{\beta}_{GMM} = \underset{\hat{\beta}_0}{\operatorname{argmin}} \left\{ n\overline{\mathbf{g}}\left(\hat{\beta}_0\right)' \mathbf{W}_n \overline{\mathbf{g}}\left(\hat{\beta}_0\right) \right\}$$
$$\hat{\beta}_{GMM} = \underset{\hat{\beta}_0}{\operatorname{argmin}} n \left(\frac{1}{n}\mathbf{Z}'\mathbf{y} - \frac{1}{n}\mathbf{Z}'\mathbf{X}\hat{\beta}_0\right)' \mathbf{W}_n \left(\frac{1}{n}\mathbf{Z}'\mathbf{y} - \frac{1}{n}\mathbf{Z}'\mathbf{X}\hat{\beta}_0\right)$$

This is similar to minimising the SSR of the linear regression model to derive the OLS estimator. Taking the first-order condition respect to $\hat{\beta}_0$ and setting it equal to 0 , will yield

$$\hat{\beta}_{GMM} = (\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{y}$$
$$\hat{\beta}_{GMM} = (\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'(\mathbf{X}\beta + \epsilon)$$

we get the sampling error of

$$\hat{\beta}_{GMM} - \beta = (\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\epsilon$$

Answer:

Next, we want to rewrite the sampling error $\hat{\beta}_{GMM} - \beta$ into a sample moment form

$$\hat{\beta}_{GMM} - \beta = \left(\frac{1}{n}\mathbf{X}'\mathbf{Z}\mathbf{W}_n\frac{1}{n}\mathbf{Z}'\mathbf{X}\right)^{-1} \frac{1}{n}\mathbf{X}'\mathbf{Z}\mathbf{W}_n\frac{1}{n}\mathbf{Z}'\epsilon$$

If we define

$$\mathbf{S}_{ZX} = \frac{1}{n}\mathbf{Z}'\mathbf{X}$$
$$\overline{\mathbf{g}} = \frac{1}{n}\mathbf{Z}'\epsilon$$

Then,

$$\hat{\beta}_{GMM} - \beta = \left(\mathbf{S}'_{ZX}\mathbf{W}_n\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}'_{ZX}\mathbf{W}_n\overline{\mathbf{g}}$$

Similar to the OLS proof of consistency, we want to show

$$p \lim_{n\longrightarrow\infty} \hat{\beta}_{GMM} = \beta$$

The first step of this proof is to apply Hayashi Lemma 2.3(a) (or Slutsky's theorem) that is

$$p \lim_{n\longrightarrow\infty}\left(\hat{\beta}_{GMM} - \beta\right) = p \lim_{n\longrightarrow\infty}\left\{\left(\mathbf{S}'_{ZX}\mathbf{W}_n\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}'_{ZX}\mathbf{W}_n\overline{\mathbf{g}}\right\}$$
$$p \lim_{n\longrightarrow\infty}\left(\hat{\beta}_{GMM} - \beta\right) = p \lim_{n\longrightarrow\infty}\left(\left(\mathbf{S}'_{ZX}\mathbf{W}_n\mathbf{S}_{ZX}\right)^{-1}\right) \times p\underset{\substack{ZX \\ n\longrightarrow\infty}}{\lim}\left(\mathbf{S}'_{ZX}\right) \times p \lim_{n\longrightarrow\infty}\left(\mathbf{W}_n\right) \times p \lim_{n\longrightarrow\infty}\left(\overline{\mathbf{g}}\right)$$

Next, we apply the strong law of large numbers (SLLN) where the sample mean converges to the population mean and the continuous mapping theorem,

$$\mathbf{S}'_{ZX} \underset{a.s}{\longrightarrow} \Sigma'_{ZX}, \Rightarrow \mathbf{S}'_{ZX} \underset{p}{\longrightarrow} \Sigma'_{ZX}$$
$$\overline{\mathbf{g}}_n \underset{a.s}{\longrightarrow} E\left(\overline{\mathbf{g}}_i\right), \Rightarrow \overline{\mathbf{g}}_n \underset{p}{\longrightarrow} E\left(\mathbf{g}_i\right)$$
$$\mathbf{W}_n \underset{a.s}{\longrightarrow} \mathbf{W}, \Rightarrow \mathbf{W}_n \underset{p}{\longrightarrow} \mathbf{W}$$

Note we also have a weak exogeneity assumption, that is $E\left(\mathbf{g}_i\right) = 0$. Therefore,

$$p \lim_{n\longrightarrow\infty}\left(\hat{\beta}_{GMM} - \beta\right) = \left(\Sigma'_{ZX}\mathbf{W}\Sigma_{ZX}\right)^{-1} \times \Sigma'_{ZX} \times \mathbf{W} \times 0$$
$$p_{n\longrightarrow\infty}\left(\hat{\beta}_{GMM} - \beta\right) = 0, \Leftrightarrow, \hat{\beta}_{GMM} \underset{p}{\longrightarrow} \beta$$

Therefore, the GMM estimator is consistent.

6 Question 6

The sampling error of the GMM estimator is, in data matrix form, $\left(\widehat{\beta}_{GMM} - \beta\right) = \left(\mathrm{X'ZW_nZXX}\right)^{-1}\mathrm{X'ZW_nZ'}\varepsilon$. Rewrite this in sample moment form and use it to show that the asymptotic variance of the GMM estimator is $\left(\Sigma'_{zx}\mathrm{W}\Sigma_{zx}\right)^{-1}\left(\Sigma'_{zx}\mathrm{WSW}\Sigma_{zx}\right)\left(\Sigma'_{zx}\mathrm{W}\Sigma_{zx}\right)^{-1}$. Include in your answer an explanation of what Σ_{zx}, W and S are.

Answer:

We want to prove that

$$\sqrt{n}(\hat{\beta} - \beta) \underset{d}{\longrightarrow} N\left(0, \left(\Sigma'_{ZX}\mathbf{W}\Sigma_{ZX}\right)^{-1}\left(\Sigma'_{ZX}\mathbf{WSW}\Sigma_{ZX}\right)\left(\Sigma'_{ZX}\mathbf{W}\Sigma_{ZX}\right)^{-1}\right)$$

The first step of this proof is we want to use the sampling error

$$\hat{\beta}_{GMM} - \beta = \left(\mathbf{S}'_{ZX}\mathbf{W}_n\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}'_{ZX}\mathbf{W}_n\overline{\mathbf{g}}$$

Next, we multiply both sides with $\sqrt{n}$

$$\sqrt{n}\left(\hat{\beta}_{GMM} - \beta\right) = \sqrt{n}\left(\mathbf{S}'_{ZX}\mathbf{W}_n\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}'_{ZX}\mathbf{W}_n\mathbf{g}_n$$

Note, multiplying $\sqrt{n}$ on both sides ensure that we do not converge to a degenerate distribution, and it converges to a well-defined distribution. Next, we apply the Slutsky theorem (or Hayashi Lemma 2.4) where $\mathbf{x}_n \underset{d}{\longrightarrow} \mathbf{x}, \mathbf{x} \sim N(0, \Sigma)$, and $\mathbf{A}_n \underset{p}{\longrightarrow} \mathbf{A}$ then $\mathbf{A}_n\mathbf{x}_n \underset{d}{\longrightarrow} N\left(0, \mathbf{A}\Sigma\mathbf{A}'\right)$. In this proof, we can assume $\mathbf{A}_n = \left(\mathbf{S}'_{ZX}\mathbf{W}_n\mathbf{S}_{ZX}\right)^{-1}\mathbf{S}'_{ZX}\mathbf{W}_n$ and $\mathbf{x}_n = \sqrt{n}\overline{\mathbf{g}}_n$. Note, from the proof in consistency we already know that

$$\mathbf{S}'_{ZX} \underset{a.s}{\longrightarrow} \Sigma_{ZX}, \Rightarrow \mathbf{S}'_{ZX} \underset{p}{\longrightarrow} \Sigma'_{ZX}$$

$$\mathbf{W}_n \underset{a.s}{\longrightarrow} \mathbf{W}, \Rightarrow \mathbf{W}_n \underset{p}{\longrightarrow} \mathbf{W}$$

Therefore $\mathbf{A} = \left(\Sigma'_{ZX}\mathbf{W}\Sigma_{ZX}\right)^{-1}\Sigma'_{ZX}\mathbf{W}$ and $\mathbf{A}' = \mathbf{W}\Sigma_{ZX}\left(\Sigma'_{ZX}\mathbf{W}\Sigma_{ZX}\right)^{-1}$. Similar to the OLS proof, for $\sqrt{n}\overline{\mathbf{g}}_n$, we apply assumption 35 and using the central limit theorem we get

$$\sqrt{n}\overline{\mathbf{g}}_n \underset{d}{\longrightarrow} N(0, \mathbf{S})$$

where $\mathbf{S} = E\left[\mathbf{g}_i\mathbf{g}'_i\right]$ is the asymptotic variance of $\overline{\mathbf{g}}_n$. Thus, applying the Slutsky theorem, we get

$$\sqrt{n}\left(\hat{\beta}_{GMM} - \beta\right) \underset{d}{\longrightarrow} N\left(0, \left(\Sigma'_{ZX}\mathbf{W}\Sigma_{ZX}\right)^{-1}\Sigma'_{ZX}\mathbf{W}\mathbf{S}\mathbf{W}\Sigma_{ZX}\left(\Sigma'_{ZX}\mathbf{W}\Sigma_{ZX}\right)^{-1}\right)$$

Thus, we have proved that the GMM estimator is asymptotic normal.

7 Question 7

Show that the conditional moment restriction $E\left(\varepsilon_i \mid x_i\right) = 0$ implies that any function $f(.)$ of x_i is orthogonal to ε_i

Answer:

Under the OLS assumption 1.2, we assume 'Strict exogeneity' which implies a zero conditional mean

$$E\left(\epsilon_i \mid \mathbf{X}\right) = 0$$

and the implication of this assumption are

$$E(\epsilon_i) = 0, \forall i = 1, \ldots n$$
$$E(\mathbf{x}_j \epsilon_i) = 0$$

That is the unconditional mean of error is zero and all the regressors are orthogonal to all the errors. Note weak exogeneity is

$$E(\mathbf{x}_i \epsilon_i) = 0$$

assumes that $\mathbf{x}_i$ is just orthogonal to its comtemporaneous error ϵ_i. This is an example of an unconditional moment restriction. Next, we can make a st ronger assumption that

$$E(\epsilon_i \mid \mathbf{x}_i) = 0, \forall i = 1, \ldots n$$

and this a conditional moment restriction. This conditional moment restriction also implies

$$E(f(\mathbf{x}_i) \epsilon_i) = 0$$

that is not only $\mathbf{x}_i$ is orthogonal to ϵ_i but also any function f of $\mathbf{x}_i$ is orthogonal to ϵ_i too. We can show this by writing the above term using the law of total expectation $E(E(X \mid Y)) = E(X)$,

$$E(f(\mathbf{x}_i) \epsilon_i) = E(E(f(\mathbf{x}_i) \epsilon_i) \mid \mathbf{x}_i)$$

Thus, since $\mathbf{x}_i$ is non-random, we can move the $f(\mathbf{x}_i)$ outside of the expectation operator and

$$E(f(\mathbf{x}_i) \epsilon_i) = E(f(\mathbf{x}_i) E(\epsilon_i \mid \mathbf{x}_i))$$

and using our conditional moment restriction $E(\epsilon_i \mid \mathbf{x}_i) = 0$:

$$E(f(\mathbf{x}_i) \epsilon_i) = E(f(\mathbf{x}_i) 0)$$

$$E(f(\mathbf{x}_i) \epsilon_i) = 0$$

8 Question 8

Consider the simplest possible linear model $y_i = \beta x_i + \varepsilon_i$ where we have a single regressor x_i and all variables are zero-mean. The OLS estimator for this model is $\hat{\beta}_{OLS} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$. Derive an expression for the inconsistency in the OLS estimator $p\lim\left(\hat{\beta}_{OLS} - \beta\right)$ in the case that weak exogeneity fails, i.e., $E(x_i \varepsilon_i) \neq 0$. Interpret this expression and explain what determines the sign, i.e., when is it positive or negative.

Answer:

Consider a linear regression model with no constant

$$y_i = x_i\beta + \epsilon_i$$

and the OLS estimator is

$$\hat{\beta} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

We want to get the above equation in sample error form so we subsititute (22) into (23)

$$\hat{\beta} = \frac{\sum_i x_i \left(x_i\beta + \epsilon_i\right)}{\sum_i x_i^2}$$
$$\hat{\beta} = \frac{\sum_i x_i^2\beta}{\sum_i x_i^2} + \frac{\sum_i x_i\epsilon_i}{\sum_i x_i^2}$$

$$\hat{\beta} = \beta + \frac{\sum_i x_i\epsilon_i}{\sum_i x_i^2}$$
$$\hat{\beta} - \beta = \frac{\sum_i x_i\epsilon_i}{\sum_i x_i^2}$$

Next, we want to get the above term in sample moment form (similar to section 1.2 or the OLS proof of consistency)

$$\hat{\beta} - \beta = \frac{\frac{1}{n}\sum_i x_i\epsilon_i}{\frac{1}{n}\sum_i x_i^2}$$

We want to find

$$p\lim_{n\longrightarrow\infty}(\hat{\beta} - \beta) = p\lim_{n\longrightarrow\infty}\left(\frac{\frac{1}{n}\sum_i x_i\epsilon_i}{\frac{1}{n}\sum_i x_i^2}\right)$$

Firstly, we can apply Hayashi Lemma 2.3(a) (or Slutsky's theorem) that is

$$p\lim_{n\longrightarrow\infty}(\hat{\beta} - \beta) = \frac{p\lim_{n\longrightarrow\infty}\left(\frac{1}{n}\sum_i x_i\epsilon_i\right)}{p\lim_{n\longrightarrow\infty}\left(\frac{1}{n}\sum_i x_i^2\right)}$$

Next, we apply the strong law of large numbers (SLLN) where the sample mean converges to the population mean and the continuous mapping theorem (note we assumes that both $\mathbf{y}$ and $\mathbf{X}$ are jointly stationary and ergodic)

$$\left(\frac{1}{n}\sum_i x_i^2\right)^{-1} \underset{a.s}{\longrightarrow} E\left(x_i^2\right)^{-1}, \Rightarrow \left(\frac{1}{n}\sum_i x_i^2\right)^{-1} \underset{p}{\longrightarrow} E\left(x_i^2\right)^{-1}$$

$$\frac{1}{n}\sum_i x_i\epsilon_i \underset{a_{is}}{\longrightarrow} E\left(x_i\epsilon_i\right), \Rightarrow \frac{1}{n}\sum_i x_i\epsilon_i \underset{p}{\longrightarrow} E\left(x_i\epsilon_i\right)$$

Note remember "Almost sure convergence" implies "Convergence in probability" which is what the above two terms is showing. Thus,

$$p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) = \frac{E(x_i \epsilon_i)}{E(x_i^2)}$$

However, here we assumed that $E(x_i \epsilon_i) \neq 0$ which does not imply $p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) = 0$. Note the terms in (24) are used to the calculate the covariance ($\operatorname{cov}(x_i \epsilon_i) = E(x_i \epsilon_i) - E(x_i) E(\epsilon_i)$) and the variance $\left(\operatorname{var}(x_i) = E\left(x_i^2\right) - E(x_i)^2\right)$, therefore (24) can be loosely written as

$$p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) = \frac{\operatorname{cov}(x_i \epsilon_i)}{\operatorname{var}(x_i)}$$

Since the denominator is always positive, the sign of the inconsistency in the OLS estimator is given by the numerator. If the regressor is positively correlated with the error, the OLS estimator is biased upwards; if the regressor is negatively correlated with the error, the OLS estimator is biased downwards.

9 Question 9

Consider the simplest possible linear model $y_i = \beta x_i + \varepsilon_i$ where we have a single regressor x_i and all variables are zero-mean. Say we also have a single excluded instrument z_i that is also zero-mean. The IV estimator for this model is $\hat{\beta}_{iv} = \frac{\sum_i z_i y_i}{\sum_i z_i x_i}$. Derive an expression for the inconsistency in the IV estimator $p \lim \left(\hat{\beta}_{IV} - \beta\right)$ in the case that weak exogeneity fails, i.e., $E(z_i \varepsilon_i) \neq 0$. Interpret this expression and explain what determines the sign, i.e., when is it positive or negative.

Answer:

Consider a linear regression model with no constant

$$y_i = x_i \beta + \epsilon_i$$

and the IV estimator is

$$\hat{\beta}_{iv} = \frac{\sum_i z_i y_i}{\sum_i z_i x_i}$$

We want to get the above equation in sample error form so we subsititute (26) into (27)

$$\hat{\beta}_{iv} = \frac{\sum_i z_i (x_i \beta + \epsilon_i)}{\sum_i z_i x_i}$$
$$\hat{\beta}_{iv} = \frac{\sum_i z_i x_i \beta}{\sum_i z_i x_i} + \frac{\sum_i z_i \epsilon_i}{\sum_i z_i x_i}$$
$$\hat{\beta}_{iv} = \beta + \frac{\sum_i z_i \epsilon_i}{\sum_i z_i x_i}$$
$$\hat{\beta}_{iv} - \beta = \frac{\sum_i z_i \epsilon_i}{\sum_i z_i x_i}$$

Next, we want to get the above term in sample moment form (similar to section 1.2 or the OLS proof of consistency)

$$\hat{\beta}_{iv} - \beta = \frac{\frac{1}{n}\sum_i z_i\epsilon_i}{\frac{1}{n}\sum_i z_i x_i}$$

We want to find

$$p\lim_{n\longrightarrow\infty}\left(\hat{\beta}_{iv} - \beta\right) = p\lim_{n\longrightarrow\infty}\left(\frac{\frac{1}{n}\sum_i z_i\epsilon_i}{\frac{1}{n}\sum_i z_i x_i}\right)$$

Firstly, we can apply Hayashi Lemma 2.3(a) (or Slutsky's theorem) that is

$$p\lim_{n\longrightarrow\infty}\left(\hat{\beta_{iv}} - \beta\right) = \frac{p\lim_{n\longrightarrow\infty}\left(\frac{1}{n}\sum_i z_i\epsilon_i\right)}{p\lim_{n\longrightarrow\infty}\left(\frac{1}{n}\sum_i z_i x_i\right)}$$

Next, we apply the strong law of large numbers (SLLN) where the sample mean converges to the population mean and the continuous mapping theorem (note we assumes that both $\mathbf{y}, \mathbf{X}$ and $\mathbf{Z}$ are jointly stationary and ergodic)

$$\left(\frac{1}{n}\sum_i z_i x_i\right)^{-1} \underset{a.s}{\longrightarrow} E\left(z_i x_i\right)^{-1}, \Rightarrow \left(\frac{1}{n}\sum_i z_i x_i\right)^{-1} \underset{p}{\longrightarrow} E\left(z_i x_i\right)^{-1}$$
$$\frac{1}{n}\sum_i z_i\epsilon_i \underset{a.s}{\longrightarrow} E\left(z_i\epsilon_i\right), \Rightarrow \frac{1}{n}\sum_i z_i\epsilon_i \underset{p}{\longrightarrow} E\left(z_i\epsilon_i\right)$$

Note remember "Almost sure convergence" implies "Convergence in probability" which is what the above two terms is showing. Thus,

$$p\lim_{n\longrightarrow\infty}\left(\hat{\beta}_{iv} - \beta\right) = \frac{E\left(z_i\epsilon_i\right)}{E\left(z_i x_i\right)}$$

However, here we assumed that $E\left(z_i\epsilon_i\right) \neq 0$ which does not imply $p\lim_{n\rightarrow\infty}\left(\hat{\beta}_{iv} - \beta\right) = 0$. Note the terms in (28) are used to the calculate the covariance ($\operatorname{cov}\left(z_i\epsilon_i\right) = E\left(z_i\epsilon_i\right) - E\left(z_i\right)E\left(\epsilon_i\right)$ and $\operatorname{cov}\left(z_i x_i\right) = E\left(z_i x_i\right) - E\left(z_i\right)E\left(x_i\right)$), therefore (28) can be loosely written as

$$p\lim_{n\longrightarrow\infty}\left(\hat{\beta_{iv}} - \beta\right) = \frac{\operatorname{cov}\left(z_i\epsilon_i\right)}{\operatorname{cov}\left(z_i x_i\right)}$$

Either way, it is the covariance of the instrument and the error divided by the covariance of the instrument and the regressor. Now the sign depends on both the numerator and denominator. If the instrument is positively correlated with the regressor, the denominator is positive, and the sign of the inconsistency in the IV estimator is given by sign of the correlation bet ween the instrument and the error (positive $\rightarrow$ upward bias, negative $\rightarrow$ downward bias). If the instrument is negatively correlated with the regressor, the denonminator is negative and the sign of the inconsistency in the IV estimator is the opposite of the sign of the correlation bet ween the inst rument and the error.

10 Question 10

To obtain a consistent estimate of the asymptotic variance of the OLS estimator, we need consistent estimates of Σ_{xx} and S. Explain how you would approach this task if (a) you have cross-section data, you suspect there is a heteroskedasticity problem, but you are unwilling to estimate the form of heteroskedasticity parametrically; (b) you have time-series data, you suspect there is a serial correlation problem, but you are unwilling to estimate the form of serial correlation parametrically. How would your answers to (a) and (b) change if you were willing to use a parametric approach?

Answer Guide:
(a) The natural approach is to use a heteroskedastic-consistent covariance estimator based on $\widehat{\mathbf{S}}_{HC}$. Should provide the formulas, e.g., slide 42 in Lecture Notes 6 plus the sandwich formula for the asymptotic variance of the OLS estimator. Cross-section means OK to assume independence.

(b) The natural approach is to use a heteroskedastic- and autocorrelation-consistent covariance estimator based on $\widehat{S}_{HAC}$. Should provide the formulas, e.g., slide 27 in Lecture Notes 7 plus the sandwich formula for the asymptotic variance of the OLS estimator. Alternatively, could assume that the lag at which serial correlation disappears is known (the question doesn't rule this out) and so use the special case on slide 28 in Lecture Notes 7. Another alternative - assume that only serial correlation is a concern (the question doesn't rule this out) and use a version that assumes conditional homoskedasticity as on slide 36 of Lecture Notes 7 Could expand discussion by talking about different kernels etc.

(a)+ parametric approach means using GLS or FGLS (can also call this WLS). If GLS, should state that you need to know the form of the skedastic function; otherwise it needs to be estimated, as in FGLS. Could discuss ways of estimating the skedastic function flexibly; we covered one method in a Stata lab and assignment but there are others.

(b)+ parametric approach again means GLS or FGLS, or optionally could mention ML. This was not covered in detail in the lectures so you wouldn't be expected to go into great detail here. Could discuss in general terms what is involved here, maybe using the AR(1) example discussed on slides 43-46 of Lecture Notes 6.