1 Linear regression model

We have a linear regression model in matrix form

$$\mathbf{y} = \mathbf{X}\beta + \epsilon, \epsilon \sim N(0, \sigma^2 \mathbf{I}_n), \tag{1}$$

and from the lecture we know the OLS estimator for both β and σ^2 is

$$\hat{\beta} \quad = \quad (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{2}$$

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})}{n-k}. \tag{3}$$

1.1 Conditional moment restriction

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

$$E(\epsilon_i|\mathbf{X}) = 0,$$

and the implication of this assumption are

$$E(\epsilon_i) = 0, \forall i = 1, ...n,$$

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

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 stronger assumption that

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

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 $E(E(X|Y)) = E(X)$,

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

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

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

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

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

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

1.2 Proof of consistency

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

$$\hat{\beta} - \beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon, \tag{4}$$

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

$$p \lim_{n \longrightarrow \infty} \hat{\beta} = \beta, \tag{5}$$

We first starting writing the (4) in a sample moment form

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

From the lecture notes we can define

$$(\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1} = \mathbf{S}_{xx}^{-1},$$

$$\frac{1}{n}\mathbf{X}'\epsilon = \bar{\mathbf{g}}_n,$$

Thus,

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

and want to prove

$$p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) = p \lim_{n \longrightarrow \infty} (\mathbf{S}_{xx}^{-1}\bar{\mathbf{g}}) = 0, \tag{6}$$

The first step want to take in the proof, is apply Hayashi Lemma 2.3(a) (or Slutsky's theorem) to (6) , for example if we two scalar sequences $\{x_n\} \underset{p}{\longrightarrow} \delta$ and $\{y_n\} \underset{p}{\longrightarrow} \gamma$ then $x_n y_n \underset{p}{\longrightarrow} \delta\gamma$. Then

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

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

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

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

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

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

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 $E(\mathbf{g}_i) = 0$.
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.

1.3 Proof of the OLS estimator is asymptotic Normality

We want to prove that

$$\sqrt{n}(\hat{\beta} - \beta) \underset{d}{\longrightarrow} N(0, \Sigma_{xx}^{-1} \mathbf{S} \Sigma_{xx}^{-1}), \tag{7}$$

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

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

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

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

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 $\mathbf{A}_n \mathbf{x}_n \underset{d}{\longrightarrow} N(0, \mathbf{A} \Sigma \mathbf{A}')$. In this proof, we can assume $\mathbf{A}_n = \mathbf{S}_{xx}^{-1}$ and $\mathbf{x}_n = \sqrt{n} \bar{\mathbf{g}}_n$. Note, from the proof in consistency we already know that

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

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

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

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

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

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

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

If we assume conidtional homoskedasticity and independence, then

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

where $\mathbf{S} = E[\mathbf{g}_i\mathbf{g}_i'] = E[\epsilon_i^2\mathbf{x}_i\mathbf{x}_i'] = E[\epsilon_i^2]E[\mathbf{x}_i\mathbf{x}_i'] = E[E[\epsilon_i^2\mathbf{x}_i\mathbf{x}_i'|\mathbf{x}_i]] = E[\mathbf{x}_i\mathbf{x}_i'E[\epsilon_i^2|\mathbf{x}_i]] = \sigma^2 E[\mathbf{x}_i\mathbf{x}_i'] = \sigma^2\Sigma_{xx}$. Thus,

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

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

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

Note conidtional homoskedasticity and independence implies

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

2 Method of moments

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

$$E(\mathbf{x}_i\epsilon_i) = E(\mathbf{x}_i(y_i - \mathbf{x}_i\beta)) = E(\mathbf{g}_i) = 0, \tag{8}$$

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

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

We can rewrite the above term is matrix form

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

Now for the MM, want to set

$$\bar{\mathbf{g}}(\hat{\beta}_0) = E(\mathbf{g}_i),$$

Given the population moment condition we have

$$\bar{\mathbf{g}}(\hat{\beta}_0) = 0,$$

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

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

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

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.

2.1 With instrumental variables

With instrumental variables (IV), we have some linear regression function form of (1) 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

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

We can rewrite the above term is matrix form

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

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

$$\bar{\mathbf{g}}(\hat{\beta}_0) = E(\mathbf{g}_i),$$

Note, from the population moment condition, we have

$$\bar{\mathbf{g}}(\hat{\beta}_0) = 0,$$

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

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

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

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.

3 GMM

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

$$\hat{\beta}_{GMM} = \underset{\hat{\beta}_0}{\operatorname{argmin}}\{n\bar{\mathbf{g}}(\hat{\beta}_0)'\mathbf{W}_n\bar{\mathbf{g}}(\hat{\beta}_0)\}, \tag{9}$$

$$\hat{\beta}_{GMM} = \underset{\hat{\beta}_0}{\operatorname{argmin}} n(\frac{1}{n}\mathbf{Z}'\mathbf{y} - \frac{1}{n}\mathbf{Z}'\mathbf{X}\hat{\beta}_0)'\mathbf{W}_n(\frac{1}{n}\mathbf{Z}'\mathbf{y} - \frac{1}{n}\mathbf{Z}'\mathbf{X}\hat{\beta}_0), \tag{10}$$

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

$$\hat{\beta}_{GMM} = (\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{y}, \tag{11}$$

$$\hat{\beta}_{GMM} = (\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'(\mathbf{X}\beta + \epsilon), \tag{12}$$

and if we substitute (1) into the above term, we get the sampling error of

$$\hat{\beta}_{GMM} - \beta = (\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\epsilon, \tag{13}$$

3.1 Proof of consistency

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

$$\hat{\beta}_{GMM} - \beta = (\frac{1}{n}\mathbf{X}'\mathbf{Z}\mathbf{W}_n\frac{1}{n}\mathbf{Z}'\mathbf{X})^{-1}\frac{1}{n}\mathbf{X}'\mathbf{Z}\mathbf{W}_n\frac{1}{n}\mathbf{Z}'\epsilon, \tag{14}$$

If we define

$$\mathbf{S}_{ZX} = \frac{1}{n}\mathbf{Z}'\mathbf{X},$$

$$\bar{\mathbf{g}} = \frac{1}{n}\mathbf{Z}'\epsilon,$$

Then,

$$\hat{\beta}_{GMM} - \beta = (\mathbf{S}'_{ZX}\mathbf{W}_n\mathbf{S}_{ZX})^{-1}\mathbf{S}'_{ZX}\mathbf{W}_n\bar{\mathbf{g}}, \tag{15}$$

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

$$p\lim_{n\longrightarrow\infty}(\hat{\beta}_{GMM} - \beta) = p\lim_{n\longrightarrow\infty}\{(\mathbf{S}'_{ZX}\mathbf{W}_n\mathbf{S}_{ZX})^{-1}\mathbf{S}'_{ZX}\mathbf{W}_n\bar{\mathbf{g}}\}, \tag{16}$$

$$p\lim_{n\longrightarrow\infty}(\hat{\beta}_{GMM} - \beta) = p\lim_{n\longrightarrow\infty}((\mathbf{S}'_{ZX}\mathbf{W}_n\mathbf{S}_{ZX})^{-1}) \times p\lim_{n\longrightarrow\infty}(\mathbf{S}'_{ZX}) \times p\lim_{n\longrightarrow\infty}(\mathbf{W}_n) \times p\lim_{n\longrightarrow\infty}(\bar{\mathbf{g}}), \tag{17}$$

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

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

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

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

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

$$p\lim_{n\longrightarrow\infty}(\hat{\beta}_{GMM} - \beta) = (\Sigma'_{ZX}\mathbf{W}\Sigma_{ZX})^{-1} \times \Sigma'_{ZX} \times \mathbf{W} \times 0, \tag{18}$$

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

Therefore, the GMM estimator is consistent.

3.2 Proof the GMM estimator asymptotic Normality

We want to prove that

$$\sqrt{n}(\hat{\beta} - \beta) \underset{d}{\longrightarrow} N(0, (\Sigma'_{ZX}\mathbf{W}\Sigma_{ZX})^{-1}(\Sigma'_{ZX}\mathbf{W}\mathbf{S}\mathbf{W}\Sigma_{ZX})(\Sigma'_{ZX}\mathbf{W}\Sigma_{ZX})^{-1}), \tag{19}$$

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

$$\hat{\beta}_{GMM} - \beta = (\mathbf{S}_{ZX}^{'}\mathbf{W}_n\mathbf{S}_{ZX})^{-1}\mathbf{S}_{ZX}^{'}\mathbf{W}_n\bar{\mathbf{g}}, \tag{20}$$

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

$$\sqrt{n}(\hat{\beta}_{GMM} - \beta) = \sqrt{n}(\mathbf{S}_{ZX}^{'}\mathbf{W}_n\mathbf{S}_{ZX})^{-1}\mathbf{S}_{ZX}^{'}\mathbf{W}_n\bar{\mathbf{g}}_n, \tag{21}$$

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 $\mathbf{A}_n\mathbf{x}_n \underset{d}{\longrightarrow} N(0, \mathbf{A}\Sigma\mathbf{A}')$. In this proof, we can assume $\mathbf{A}_n = (\mathbf{S}_{ZX}^{'}\mathbf{W}_n\mathbf{S}_{ZX})^{-1}\mathbf{S}_{ZX}^{'}\mathbf{W}_n$ and $\mathbf{x}_n = \sqrt{n}\bar{\mathbf{g}}_n$. 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 $\mathbf{A} = (\Sigma_{ZX}^{'}\mathbf{W}\Sigma_{ZX})^{-1}\Sigma_{ZX}^{'}\mathbf{W}$ and $\mathbf{A'} = \mathbf{W}\Sigma_{ZX}(\Sigma_{ZX}^{'}\mathbf{W}\Sigma_{ZX})^{-1}$. Similar to the OLS proof, for $\sqrt{n}\bar{\mathbf{g}}_n$, we apply assumption 35 and using the central limit theorem we get

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

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

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

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

4 Other

4.1 OLS

Consider a linear regression model with no constant

$$y_i = x_i\beta + \epsilon_i, \tag{22}$$

and the OLS estimator is

$$\hat{\beta} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}, \tag{23}$$

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

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

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

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

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

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

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

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

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

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)

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

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

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

$$p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) = \frac{E(x_i \epsilon_i)}{E(x_i^2)}, \tag{24}$$

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 ($cov(x_i \epsilon_i) = E(x_i \epsilon_i) - E(x_i)E(\epsilon_i)$) and the variance($var(x_i) = E(x_i^2) - E(x_i)^2$), therefore (24) can be loosely written as

$$p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) = \frac{cov(x_i \epsilon_i)}{var(x_i)}, \tag{25}$$

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.

4.2 IV

Consider a linear regression model with no constant

$$y_i = x_i\beta + \epsilon_i, \tag{26}$$

and the IV estimator is

$$\hat{\beta}_{iv} = \frac{\sum_i z_i y_i}{\sum_i z_i x_i}, \tag{27}$$

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

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

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

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

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)

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

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

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

$$p \lim_{n \longrightarrow \infty} (\hat{\beta}_{iv} - \beta) = \frac{E(z_i \epsilon_i)}{E(z_i x_i)}, \tag{28}$$

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

$$p \lim_{n \longrightarrow \infty} (\hat{\beta_{iv}} - \beta) = \frac{cov(z_i \epsilon_i)}{cov(z_i x_i)}, \tag{29}$$

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 between the instrument and the error (positive →upward bias, negative→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 between the instrument and the error.

5 Heteroskedascity and autocorrelation

Assume we have time-series data and we believe there is both heteroskedascity and serial correlation presence in the data. A non-parametric approach, that is we does not assume any specific functional form regarding the heteroskedascity and the serial correlation, is to use the heteroskedastic and autocorrelation-consistent covariance (HAC) estimator $\bar{\mathbf{S}}_{\mathbf{HAC}}$ based on lecture 7. This HAC estimator is a consistent estimator of $\mathbf{S}$, which means we are able to construct the asymptotic variance of $\hat{\beta}$ or $\widehat{\text{AVAR}}(\hat{\beta})$ that is robust to heteroskedasticity and autocorrelation.
For the parameteric approach, we could assume an ARMA representation that explicitly takes into consideration both heteroskedascity and serial correlation, for instance ARMA(2,2)

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \theta_2 \epsilon_{t-2} + \theta_1 \epsilon_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma_t^2), \tag{30}$$

The serial correlation comes via $\theta_2 \epsilon_{t-2} + \theta_1 \epsilon_{t-1}$ and the heteroskedascity comes via σ_t^2 (that is the variance varies across time periods). There are two ways to estimate (30), either the Frequentist or the Bayesian approach. In the frequentist approach, a standard way to estimate an ARMA model is through maximum likelihood (ML). ML is a method of estimating the parameters of a probability distribution by maximising a likelihood function, so that under the assumed statistical model the observed data is most probable. However, there can be issues with ML, for instance a model's likelihood function could contain various local maximums, which implies that the starting value of the optimisation of the ML is very important. It is for this reason, majority of the ARMA models in the macroeconometrics literature are now estimated via a Bayesian approach as it is more flexible compared to standard frequentist approach. For example, a good Bayesian paper on ARMA models "Stochastic volatility models with ARMA innovations: An application to G7 inflation" is by Zhang et al. (2020) that is published in the International Journal of Forecasting.