P218 Econometrics I

TA Session 3

Gabriel Simões Gaspar

London Business School

Fall 2022

Roadmap

Question 3

Part (a)

We are given that for $j \in \{1, 2\}$ independent samples we have that:

$$\bar{Y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ji}$$

$$s_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (Y_{ji} - \bar{Y}_j)^2$$

The usual test under the assumption of normality to test $H_0 : \sigma^2 = \sigma_2^2$ is to look at

$$s_1^2 / s_2^2 \sim F(n_1 - 1, n_2 - 2)$$

Part (a)

We want to prove that:

$$T = \left(\frac{n_1 n_2}{n_1 + n_2}\right)^{1/2} (\log s_1^2 - \log s_2^2) \xrightarrow{d} \mathcal{N}(0, 2)$$

Note that

$$\log s_1^2 - \log s_2^2 = \log\left(\frac{s_1^2}{s_2^2}\right) \sim \log\left(F(n_1 - 1, n_2 - 2)\right)$$

Remember that an F distribution is just the ratio of two χ^2 distributions divided by their degrees of freedom, so

$$\log\left(F(n_1 - 1, n_2 - 2)\right) \sim \log\left(\frac{\chi^2(n_1 - 1)}{n_1 - 1}\right) - \log\left(\frac{\chi^2(n_2 - 1)}{n_2 - 1}\right)$$

Part (a)

We know also by the CLT that

$$\frac{\chi^2(n_1 - 1)}{n_1 - 1} \approx 1 + \frac{1}{\sqrt{n_1}}\mathcal{N}(0, 2)$$

But we can use a Taylor expansion to show that

$$\log(1 + x) \approx x$$

So we will have that

$$\log\left(\frac{\chi^2(n_1 - 1)}{n_1 - 1}\right) \approx \frac{1}{\sqrt{n_1}}\mathcal{N}(0, 2)$$

Part (a)

Going back to the original expression we will have that

$$
\begin{aligned}
\log\left(\frac{\chi^2(n_1-1)}{n_1-1}\right) - \log\left(\frac{\chi^2(n_2-1)}{n_2-1}\right) &\approx \frac{1}{\sqrt{n_1}}\mathcal{N}_1(0,2) + \frac{1}{\sqrt{n_1}}\mathcal{N}_2(0,2) \\
&= \mathcal{N}\left(0, 2 \times \frac{n_1+n_2}{n_1 n_2}\right) \\
&= \left(\frac{n_1 n_2}{n_1+n_2}\right)^{1/2} \mathcal{N}(0,2)
\end{aligned}
$$

Part (a)

Going back to the start and putting everything together we have

$$
\begin{aligned}
T &= \left(\frac{n_1 n_2}{n_1 + n_2}\right)^{1/2} (\log s_1^2 - \log s_2^2) \\
&= \left(\frac{n_1 n_2}{n_1 + n_2}\right)^{1/2} \log\Bigg(F(n_1 - 1, n_2 - 2)\Bigg) \\
&= \left(\frac{n_1 n_2}{n_1 + n_2}\right)^{1/2} \mathcal{N}\left(0, 2 \times \frac{n_1 + n_2}{n_1 n_2}\right) \\
&\xrightarrow{d} \mathcal{N}(0, 2)
\end{aligned}
$$

Part (b)

Given our formula for the variance estimator, we can manipulate the variance estimator as follows:

$$
\begin{aligned}
s_j^2(n_j-1) &= \sum_{i=1}^{n_j}(Y_{ij}-\mu_j-\bar{Y}_j+\mu_j)^2 \\
&= \sum_{i=1}^{n_j}([Y_{ij}-\mu_j]-[\bar{Y}_j-\mu_j])^2 \\
&= \sum_{i=1}^{n_j}(Y_{ij}-\mu_j)^2 - 2\sum_{i=1}^{n_j}([Y_{ij}-\mu_j][\bar{Y}_j-\mu_j]) + \sum_{i=1}^{n_j}(\bar{Y}_j-\mu_j)^2 \\
&= \sum_{i=1}^{n_j}(Y_{ij}-\mu_j)^2 - n_j(\bar{Y}_j-\mu_j)^2
\end{aligned}
$$

where the last equality holds since $\sum_{i=1}^{n_j}(Y_{ij}-\mu_j)=0$.

Part (b)

Given the manipulation above, we now turn our attention to finding the distribution of $\sqrt{n_j}(s_j^2 - \sigma_j^2)$. We can then to the following:

$$\sqrt{n_j}(s_j^2 - \sigma_j^2) = \frac{\sqrt{n_j}}{n_j - 1}\sum_{i=1}^{n_j}(Y_{ij} - \mu_j)^2 - \frac{n_j}{n_j - 1}\sqrt{n_j}(\bar{Y}_j - \mu_j)^2 - \sqrt{n_j}\sigma^2$$

$$= \frac{n_j}{n_j - 1}\left(\sqrt{n_j}\left[\frac{1}{n_j}\sum_{i=1}^{n_j}(Y_{ij} - \mu_j)^2 - \sigma_j^2\right]\right)$$

$$- \frac{n_j}{n_j - 1}\sqrt{n_j}(\bar{Y}_j - \mu_j)^2 - \frac{\sqrt{n_j}}{n_j - 1}\sigma_j^2$$

But note that as $n \to \infty$ we will have that $\frac{n_j}{n_j-1} \to 1$ and that $\frac{\sqrt{n_j}}{n_j-1}\sigma_j^2 \to 0$. We can then use Stlutsky's lemma to see that:

$$\sqrt{n_j}(\bar{Y}_j - \mu_j)^2 = \sqrt{n_j}(\bar{Y}_j - \mu_j)(\bar{Y}_j - \mu_j) \xrightarrow{p} 0$$

Part (b)

Focusing on the first term of our expression of interest, we will have that given the central limit theorem:

$$\frac{n_j}{n_j - 1}\left(\sqrt{n_j}\left[\frac{1}{n_j}\sum_{i=1}^{n_j}(Y_{ij} - \mu_j)^2 - \sigma_j^2\right]\right) \xrightarrow{d} \mathcal{N}(0, \sigma_j^4(\kappa_j - 1))$$

Where κ_j is the kurtosis from the distribution $G_j(y)$. And so:

$$\sqrt{n_j}(s_j^2 - \sigma_j^2) \xrightarrow{d} \mathcal{N}(0, \sigma_j^4(\kappa_j - 1))$$

Part (b)

Another element that we need to keep in mind for this question is the Delta method. Given a function $g(\cdot) = \log(\cdot)$, we will have that the following result is true by the Delta method:

$$\sqrt{n_j}(\bar{Y}_j - \mu_j) \xrightarrow{d} \mathcal{N}(0, \sigma_j^2) \implies \sqrt{n_j}(\log(\bar{Y}_j) - \log(\mu_j)) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma_j^2}{\mu_j}\right)$$

This means that, for our variables of interest, the following holds:

$$\sqrt{n_j}(\log(s_j^2) - \log(\sigma_j^2)) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma_j^4}{\sigma_j^4}(\kappa_j - 1)\right) = \mathcal{N}\left(0, \kappa_j - 1\right)$$

This then implies that:

$$\log(s_j^2) \xrightarrow{d} \mathcal{N}\left(\log(\sigma_j^2), \frac{\kappa_j - 1}{n_j}\right)$$

Part (b)

Given the above, we then know that the distribution for our two samples $j \in \{1, 2\}$ is then:

$$\log(s_1^2) - \log(s_2^2) \xrightarrow{d} \mathcal{N}\left(\log\left(\frac{\sigma_1^2}{\sigma_2^2}\right), \frac{n_2(\kappa_1 - 1) - n_1(\kappa_2 - 1)}{n_1 n_2}\right)$$

But under the null hypothesis that $\sigma_1^2 = \sigma_2^2$ and that $\kappa_1 = \kappa_2 = \kappa$, this simplifies to:

$$\log(s_1^2) - \log(s_2^2) \xrightarrow{d} \mathcal{N}\left(0, \frac{(n_1 + n_2)(\kappa - 1)}{n_1 n_2}\right)$$

Then our expression for T becomes simply:

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\left[\log\left(s_1^2\right) - \log\left(s_2^2\right)\right] \xrightarrow{d} \mathcal{N}(0, \kappa - 1)$$

Part (c)

We can then see that the comparison of s_1^2/s_2^2 to an F distribution is asymptotically equivalent to comparing an $\mathcal{N}(0, \kappa - 1)$ random variable to a $\mathcal{N}(0, 2)$ distribution. When $\kappa = 5$ the asymptotic size of the test based on such a comparison is just

$$\Pr(\mathcal{N}(0,4) \geq \sqrt{2} \times 1.64\Bigg) = \Pr(\mathcal{N}(0,1) \geq 1.1597) = 0.1231$$

IV Overview

Omitted Variable Bias

Imagine you want to estimate the following long regression

$$\mathbb{E}[y_i \mid 1, fam, educ, init] = \alpha_0 + \alpha_1 \times fam_i + \alpha_2 \times educ_i + \alpha_3 \times init_i$$

but you don't have data on $init_i$. You can try to run the least-squares regression on the variables that you do have and the estimators will converge in probability to

$$\mathbb{E}[y_i \mid 1, fam, educ, init] = \beta_0 + \beta_1 \times fam_i + \beta_2 \times educ_i$$

What is the relationship between the estimators in the two regressions above?

Omitted Variable Bias

Consider the auxiliary regression

$$\mathbb{E}[init_i \mid 1, fam_i, educ_i] = \gamma_0 + \gamma_1 \times fam_i + \gamma_2 \times educ_i$$

The omitted variable formula gives us that

$$\beta_j = \alpha_j + \alpha_3 \times \gamma_j, \text{ for } j = 0, 1, 2$$

How can we prove that? Running the short regression is a problem if $\alpha_3 \neq 0$ or if $\gamma_j \neq 0$, so we need to figure out a way around this.

Instrument

Assume that we have another variable sub_i which is the subsidy that individual i receives. This variable observes the following two exclusion restrictions:

$$\mathbb{E}[y_i \mid 1, fam, sub, educ, init] = \alpha_0 + \alpha_1 \times fam_i + \alpha_2 \times educ_i + \alpha_3 \times init_i$$
$$\mathbb{E}[init_i \mid 1, fam, sub] = \lambda_0 + \lambda_1 \times fam_i$$

The first exclusion restriction says that sub_i does not help to predict y_i if added to the long regression. Why? The second tells us that sub_i does not help to predict $init_i$ in a linear predictor with fam_i. Where am I going with this? What is the endogenous variable here?

Instrument

Let:

$$\begin{aligned} init_i &= \lambda_0 + \lambda_1 \times fam_i + \varepsilon_i \\ y_i &= \alpha_0 + \alpha_1 \times fam_i + \alpha_2 \times educ_i + \alpha_3 \times init_i + u_i \end{aligned}$$

Combine both equations to get

$$\begin{aligned} y_i &= \alpha_0 + \alpha_1 \times fam_i + \alpha_2 \times educ_i + \alpha_3 \times (\lambda_0 + \lambda_1 \times fam_i + \varepsilon_i) + u_i \\ &= \underbrace{(\alpha_0 + \alpha_3 \times \lambda_0)}_{\equiv \delta_0} + \underbrace{(\alpha_1 + \alpha_3 \times \lambda_1)}_{\equiv \delta_1} \times fam_i + \alpha_2 \times educ_i + \underbrace{(u_i + \alpha_3 \times \varepsilon_i)}_{\equiv v_i} \end{aligned}$$

Note then that we will have

$$\begin{aligned} \mathbb{E}[fam_i \cdot v_i] &= 0 \\ \mathbb{E}[sub_i \cdot v_i] &= 0 \end{aligned}$$

Instrument

If we define

$$R_i \equiv \begin{bmatrix} 1 & fam_i & educ_i \end{bmatrix}$$

$$W_i \equiv \begin{bmatrix} 1 \\ fam_i \\ sub_i \end{bmatrix}$$

$$\gamma \equiv \begin{bmatrix} \delta_0 \\ \delta_1 \\ \alpha_2 \end{bmatrix}$$

The exclusion restrictions imply

$$y_i = R_i\gamma + v_i, \qquad \mathbb{E}[W_i \cdot v_i] = 0$$

We can obtain a consistent estimator for γ, which will give us a consistent estimator for δ_0, δ_1, α_2. Is this really true?

Just-Identified Instrument

We can play around with our formula for y_i as follows

$$W_i y_i = (W_i R_i)\gamma + W_i V_i \implies \gamma = (\mathbb{E}[W_i R_i])^{-1}\mathbb{E}[W_i Y_i]$$

What condition do we need to impose on $\mathbb{E}[W_i R_i]$?

Replace the population expectations by sample averages to get a consistent estimator for γ:

$$\hat{\gamma} = \left(\frac{1}{n}\sum_{i=1}^{n} W_i R_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{i=1}^{n} W_i Y_i\right) = S_{WR}^{-1} S_{WY}$$

Just-Identified Instrument

Suppose that $fam_i = 0$ and that it doesn't matter. So our population equations become

$$y_i = \delta_0 + \alpha_2 \times educ_i + v_i, \qquad \mathbb{E}[v_i] = 0, \qquad \mathbb{E}[sub_i \cdot v_i] = 0$$

This means then that $\text{Cov}(sub_i, v_i) = 0$ and that

$$\text{Cov}(sub_i, y_i) = \text{Cov}(sub_i, educ_i) \times \alpha_2 \implies \alpha_2 = \frac{\text{Cov}(sub_i, y_i)}{\text{Cov}(sub_i, educ_i)}$$

This leads us to our second condition for the instrument: $\text{Cov}(sub_i, educ_i) \neq 0$. Again, we can replace the population covariances with their sample counterparts to get a consistent estimator:

$$\hat{\alpha}_2 = \frac{\widehat{\text{Cov}}(sub_i, y_i)}{\widehat{\text{Cov}}(sub_i, educ_i)}$$

Over-Identified Instrument

If you got the answers to the previous part right you know that we have considered that $\text{rank}(R) = \text{rank}(W)$. But what something goes wrong if we assume that $\text{rank}(R) \neq \text{rank}(W)$. What is going on?

To get around this issue, we can estimate γ using a minimum-distance estimator:

$$\begin{aligned}\hat{\gamma} &= \arg\min_a (S_{WY} - S_{WR}a)'\hat{\Omega}(S_{WY} - S_{WR}a) \\ &= \left(S_{WR}'\hat{\Omega}S_{WR}\right)^{-1} S_{WR}'\hat{\Omega}S_{WY}\end{aligned}$$

What requirements do we need for the matrix $\hat{\Omega}$? More on this in a second.

2SLS

Let's put it all together. We have that:

$$y_i = \delta_0 + \delta_1 \times fam_i + \alpha_2 \times educ_i + v_i, \qquad \mathbb{E}[W_i \cdot v_i] = 0$$

We can use the population equation above to write the best linear predictor of y_i given W_i as:

$$\mathbb{E}[y_i \mid W_i] = \delta_0 + \delta_1 \times fam_i + \alpha_2 \times \mathbb{E}[educ_i \mid W_i]$$

Define

$$educ_i^* = \mathbb{E}[educ_i \mid W_i] = W_i' \tau$$

Then the best linear predictor of Y_i given fam_i and $educ_i$ identifies δ_0, δ_1 and α_2:

$$\mathbb{E}[y_i \mid W_i] = \delta_0 + \delta_1 \times fam_i + \alpha_2 \times educ_i^*$$

IV & GMM

Setup

Consider the canonical setup for IV

$$Y_i = X_i\beta + W_i\gamma_1 + \varepsilon_i$$
$$X_i = Z_i\pi + W_i\gamma_2 + u_i$$

where X_i is our endogenous regressor and Z_i is our instrument. As always, we need the following assumptions in this setup:

- Relevance: $\pi \neq 0$
- Exclusion: $\mathbb{E}[\varepsilon_i Z_i \mid W_i] = 0$

I can also write the exclusion restriction as $\mathbb{E}[\varepsilon_i Z_i^*] = 0$. What would Z_i^* be in this case? And what does this mean?

A System of Equations

Alternatively, we can write the canonical setup as system of K equations

$$\mathbb{E}[(Y_i - X_i\beta + W_i\gamma_1)Z_i] = 0$$
$$\mathbb{E}[(Y_i - X_i\beta + W_i\gamma_1)W_i] = 0$$

The first equation is simply the exclusion restriction. What does the second equation tell us?

Let $\tilde{X}_i \equiv \begin{bmatrix} X_i & W_i \end{bmatrix}$ and $\tilde{Z}_i \equiv \begin{bmatrix} Z_i & W_i \end{bmatrix}$, so we can write the system of linear equations simply as

$$\mathbf{g}(\beta, \gamma) = \mathbb{E}[(Y_i - \tilde{X}_i\tilde{\beta})\tilde{Z}_i]$$

GMM

Given this system of K moments, for a positive-definite $K \times K$ weight matrix Ω our linear GMM estimator is simply

$$\hat{\beta}_{GMM} = \arg\min_{\tilde{\beta}} \mathbf{g}(\beta, \gamma)' \Omega \mathbf{g}(\beta, \gamma)$$

This is equal to

$$\hat{\beta}_{GMM} = \frac{\tilde{X}' \tilde{Z} \Omega \tilde{Z}' Y}{\tilde{X}' \tilde{Z} \Omega \tilde{Z}' \tilde{X}}$$

Plug in $\tilde{Y}$ above and we have that

$$\hat{\beta}_{GMM} - \tilde{\beta} = \frac{\tilde{X}' \tilde{Z} \Omega \tilde{Z}' \varepsilon}{\tilde{X}' \tilde{Z} \Omega \tilde{Z}' \tilde{X}} \to 0$$

What assumptions do we need for this to be true?

GMM vs IV

We can compare the GMM and the usual 2SLS estimator.

$$\hat{\beta}_{GMM} = \frac{\tilde{X}'\tilde{Z}\Omega\tilde{Z}'Y}{\tilde{X}'\tilde{Z}\Omega\tilde{Z}'\tilde{X}}$$
$$\hat{\beta}_{2SLS} = \frac{\tilde{X}'\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'Y}{\tilde{X}'\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\tilde{X}}$$

So the 2SLS estimator is just a special case of the GMM estimator when we use $\Omega = (\tilde{Z}'\tilde{Z})^{-1}$! Of course, we might be able to get a better weight matrix using fancy techniques as you saw in lecture. But if we have homoskedasticity, it turns out that $\Omega = (\tilde{Z}'\tilde{Z})^{-1}$ is optimal.

Exclusion Restriction

Exclusion Restriction

We saw that we need relevance and exclusion in order to compute our IV estimator. The exclusion restriction says that

$$\mathbb{E}[\varepsilon_i Z_i \mid W_i] = 0$$

Can we test this? What if we use $\hat{\varepsilon}_i$ as a replacement for ε_i?

Exclusion Restriction

The exclusion restriction is something that is untestable! So if you are using IV for one of your papers, you should motivate why this r estriction should hold.

Let's think of a case where your instrument is virtually (near) random (e.g. a lottery, weather, etc.) and the relevance restriction is satisfied. Does that mean that we can take a deep breath and assume that the exclusion restriction holds?

Three examples

Let's discuss three examples:

1. The Vietnam War lottery: let y_i be death, $viet_i$ is whether a person served in the Vietnam war and our instrument z_i is their lottery number.
2. Rain as an instrument: let y_i be conflict, $incom_i$ is income and z_i is the amount of rain in location i.
3. Legal origin: let y_i be some development indicator, $institut_i$ is the presence of some institution and z_i is the legal origin of a given country.

A Framework

How can we think about the exclusion restriction then when reading a paper or doing our own research? Check out Angrist, Imbens and Rubin (1996) for proof:

$$\frac{\mathbb{E}[Y_i(1, D_i(1) - Y_i(0, D_i(0)]}{\mathbb{E}[D_i(1) - D_i(0)]} = \mathbb{E}[Y_i(1, D_i(1) - Y_i(0, D_i(0) \mid i \text{ complier}] + \mathbb{E}[H_i \mid i \text{ non-complier}] \times \frac{\Pr(i \text{ non-complier})}{\Pr(i \text{ complier})}$$

and $H_i = Y_i(1, d) - Y_i(0, d)$ where $d = 1$ for always-taker and $d = 0$ for never-taker.

Extra Topics

Shift-Share IV

Assume that we are interested in understanding the following regression

$$y_\ell = \beta_0 + \beta_1 x_\ell + \varepsilon_\ell$$

such that $\mathbb{E}[\varepsilon_\ell x_\ell] \neq 0$. We have that y_ℓ is the growth in employment, x_ℓ is the growth in import exposure to China. We need an instrument for location-level exposure to trade with China. To find an IV, let's look at two accounting identities

$$x_\ell = \sum_{k=1}^{K} z_{\ell k} g_{\ell k}$$

$$g_{\ell k} = g_k + \tilde{g}_{\ell k}$$

where $z_{\ell k}$ is the share of industry k in location ℓ, $g_{\ell k}$ is the growth of imports (from China) from industry k in location ℓ, g_k is the national growth of industry k and $\tilde{g}_{\ell k}$ is some idiosyncratic location-industry growth.

Shift-Share IV

The shift-share (Bartik) instrument is given by

$$B_\ell = \sum_{k=1}^{K} z_{\ell k} g_k$$

This gives us the following 2SLS structure

$$\begin{aligned} y_\ell &= \beta_0 + \beta_1 x_\ell + \varepsilon_\ell \\ x_\ell &= \pi_0 + \pi_1 B_\ell + u_\ell \\ B_\ell &= \sum_{k=1}^{K} z_{\ell k} g_k \\ g_{\ell k} &= g_k + \tilde{g}_{\ell k} \end{aligned}$$

We, of course, need that $\pi_1 \neq 0$ and $\mathbb{E}[\varepsilon_\ell B_\ell] = 0$

Judge, Examiner, Leniency IV

Setting: randomized examiner (e.g. judges, loan officers, administrator, etc.) deciding some outcome (e.g. sentence, loan, cash programme, etc.). Examiners vary in how they decide outcome (some judges are nicer than others).

Example: if you are arrested you can either go to jail, pay bail or be released with no bail (in both cases you have to come back for trial). Judge is randomly assigned to judge your case. Can use judge assignment as an instrument to explore effect of bail (or being held in prison) on the probability that you plead guilty. This takes care of endogeneity that some people are held in jail or are given bail because of confounding factors.