Econometrics I - TA section

Eric Schulman

The University of Texas at Austin

November 20, 2020

8.2 In the model

$$y = X_1\beta_1 + X_2\beta_2 + e$$

show directly from definition (8.3) that the CLS estimate of $\beta = (\beta_1, \beta_2)$, subject to the constraint that $\beta_1 = c$ (where c is some given vector) is the OLS regression of $y - X_1 c$ on X_2 .

- (8.3) $\hat{\beta} = \text{argmin} SSE(\beta)$ subject to $R'\beta = c$
- What are the dimensions of R?

8.2 In the model

$$y = X_1\beta_1 + X_2\beta_2 + e$$

show directly from definition (8.3) that the CLS estimate of $\beta = (\beta_1, \beta_2)$, subject to the constraint that $\beta_1 = c$ (where c is some given vector) is the OLS regression of $y - X_1 c$ on X_2 .

- (8.3) $\hat{\beta} = \text{argmin} SSE(\beta)$ subject to $R'\beta = c$
- What are the dimensions of R? $k \times q$

8.2 In the model

$$y = X_1\beta_1 + X_2\beta_2 + e$$

show directly from definition (8.3) that the CLS estimate of $\beta = (\beta_1, \beta_2)$, subject to the constraint that $\beta_1 = c$ (where c is some given vector) is the OLS regression of $y - X_1 c$ on X_2 .

- (8.3) $\hat{\beta} = \text{argmin} SSE(\beta)$ subject to $R'\beta = c$
- What are the dimensions of R? $k \times q$
- What about c?

8.2 In the model

$$y = X_1\beta_1 + X_2\beta_2 + e$$

show directly from definition (8.3) that the CLS estimate of $\beta = (\beta_1, \beta_2)$, subject to the constraint that $\beta_1 = c$ (where c is some given vector) is the OLS regression of $y - X_1 c$ on X_2 .

- (8.3) $\hat{\beta} = \text{argmin} SSE(\beta)$ subject to $R'\beta = c$
- What are the dimensions of R? $k \times q$
- What about c? $q \times 1$

8.2 In the model

$$y = X_1\beta_1 + X_2\beta_2 + e$$

show directly from definition (8.3) that the CLS estimate of $\beta = (\beta_1, \beta_2)$, subject to the constraint that $\beta_1 = c$ (where c is some given vector) is the OLS regression of $y - X_1 c$ on X_2 .

- (8.3) $\hat{\beta} = \text{argmin} SSE(\beta)$ subject to $R'\beta = c$
- What are the dimensions of R? $k \times q$
- What about c? $q \times 1$
- what is rank(R)?

8.2 In the model

$$y = X_1\beta_1 + X_2\beta_2 + e$$

show directly from definition (8.3) that the CLS estimate of $\beta = (\beta_1, \beta_2)$, subject to the constraint that $\beta_1 = c$ (where c is some given vector) is the OLS regression of $y - X_1 c$ on X_2 .

- (8.3) $\hat{\beta} = \text{argmin} SSE(\beta)$ subject to $R'\beta = c$
- What are the dimensions of R? $k \times q$
- What about c? $q \times 1$
- what is rank(R)? How does it relate to k?

8.2 In the model

$$y = X_1\beta_1 + X_2\beta_2 + e$$

show directly from definition (8.3) that the CLS estimate of $\beta = (\beta_1, \beta_2)$, subject to the constraint that $\beta_1 = c$ (where c is some given vector) is the OLS regression of $y - X_1 c$ on X_2 .

- (8.3) $\hat{\beta} = \operatorname{argmin} SSE(\beta)$ subject to $R'\beta = c$
- What are the dimensions of R? $k \times q$
- What about c? $q \times 1$
- what is rank(R)? How does it relate to k? rank$(R) \leq q \leq k$. There can be redundant constraints. The important thing is the constraints don't contradict each other.

8.2

- What is the objective function?

$$\sum_{i=1}^{n}(y_i - x_{i1}'\beta_1 - x_{i2}'\beta_2)'(y_i - x_{i1}'\beta_1 - x_{i2}'\beta_2)$$

subject to $\beta_1 = c$

- Simplifies to

$$\sum_{i=1}^{n}(y_i - x_{i1}'c - x_{i2}'\beta_2)'(y_i - x_{i1}'c - x_{i2}'\beta_2)$$

- Can redefine $\tilde{y}_i = y_i - x_{i1}'c$

8.3 In the model

$$y = X_1\beta_1 + X_2\beta_2 + e$$

with X_1 and X_2 each $n \times k$, find the CLS estimate of $\beta = (\beta_1, \beta_2)$, subject to the constraint that $\beta_1 = -\beta_2$.

- The objective is

$$\hat{\beta} = \text{argmin} \sum_{i=1}^{n} (y_i - x_{i1}'\beta_1 - x_{i2}'\beta_2)'(y_i - x_{i1}'\beta_1 - x_{i2}'\beta_2)$$

  subject to $\beta_1 = -\beta_2$

- Plugging in the constraint

$$\hat{\beta} = \text{argmin} \sum_{i=1}^{n} (y_i - (x_{i1}' - x_{i2}')\beta_1)'(y_i - (x_{i1}' - x_{i2}')\beta_1)$$

- It's just OLS

$$\hat{\beta}_1 = \left(\sum_{i=1}^{n} (x_{i1} - x_{i2})'(x_{i1} - x_{i2}) \right)^{-1} \left(\sum_{i=1}^{n} (x_{i1} - x_{i2})' y_i \right)$$

Chapter 9

Hypothesis testing. What if we want to infer something about the population using data?

Hypothesis testing. What if we want to infer something about the population using data?

- Step 1 pick a distribution to evaluate the probability of your estimator.

Chapter 9

Hypothesis testing. What if we want to infer something about the population using data?

- Step 1 pick a distribution to evaluate the probability of your estimator.
- Step 2, if your estimator is not very likely. Then you probably picked the wrong distribution. Pick a different one.

Step 1 pick a distribution (e.g. $\beta_0 = 0$) to evaluate your estimator $\hat{\beta}_{obs}$.

- Our goal is to calculate $Pr(\hat{\beta} \leq \hat{\beta}_{obs})$ and see if it's low.
- Remember $\hat{\beta}$ is a random variable. $\hat{\beta}_{obs}$ is data, which is fixed.

Step 1 pick a distribution (e.g. $\beta_0 = 0$) to evaluate your estimator $\hat{\beta}_{obs}$.

- Our goal is to calculate $Pr(\hat{\beta} \leq \hat{\beta}_{obs})$ and see if it's low.
- Remember $\hat{\beta}$ is a random variable. $\hat{\beta}_{obs}$ is data, which is fixed.
- Pick a distribution. Amounts to picking population parameters $\beta_0 = 0$. This is called the "null" hypothesis.

$$\sqrt{n}\frac{\hat{\beta} - \beta_0}{se(\hat{\beta})} \xrightarrow{d} \mathcal{N}(0,1)$$

- As a result, we know the cdf of $\hat{\beta}$ i.e. $\Phi\left(\frac{\hat{\beta} - \beta_0}{se(\hat{\beta})}\right)$

Step 2, if your estimator is not very likely. Then you probably picked the wrong distribution. Pick a different one.

- For example if $Pr(\hat{\beta} \leq \hat{\beta}_{obs})$ is really small i.e. less than say 5 percent.
- If $Pr(\hat{\beta} \leq \hat{\beta}_{obs}) \leq 5$ that means – you had 100 samples generated using β_0, then you would only expect 5 of them to have $\hat{\beta} \leq \hat{\beta}_{obs}$

9.2 You have two independent samples (y_1, X_1) and (y_2, X_2) which satisfy $y_1 = X_1\beta_1 + e_1$ and $y_2 = X_2\beta_2 + e_2$, where $E(x_{1i}e_{1i}) = 0$ and $E(x_{2i}e_{2i}) = 0$, and both X_1 and X_2 have k columns. Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be the OLS estimates of β_1 and β_2. For simplicity, you may assume that both samples have the same number of observations n.

9.2 Find the asymptotic distribution of $\sqrt{n}((\hat{\beta}_2 - \hat{\beta}_1) - (\beta_2 - \beta_1))$ as $n \to \infty$.

9.2 Find the asymptotic distribution of $\sqrt{n}((\hat{\beta}_2 - \hat{\beta}_1) - (\beta_2 - \beta_1))$ as $n \to \infty$.

- Apply Lindeberg-Levy CLT. Both statistics are independent from each other.
  $\mathcal{N}(0, V_{\beta_1} + V_{\beta_2})$
- Also, notice how picking β_1 and β_2 would result in different asymptotic distributions i.e. $\beta_1 = 1$ would be a different distribution for $\hat{\beta}_1$ than $\beta_1 = 3$

9.2

- Find an appropriate test statistic for $H_0 : \beta_2 = \beta_1$.
- Find the asymptotic distribution of this statistic under H_0

9.2 For $k = 1$ use $\hat{\beta}_2/se(\hat{\beta}_2) - \hat{\beta}_1/se(\hat{\beta}_1)$

- From part 1 we know the distribution of

$$\sqrt{n}((\hat{\beta}_2 - \hat{\beta}_1) - (\beta_2 - \beta_1))$$

- Under the null, this expression simplifies to

$$\sqrt{n}((\hat{\beta}_2 - \hat{\beta}_1)$$

- As a result, $\sqrt{n}\dfrac{\hat{\beta}_2 - \hat{\beta}_1}{se(\hat{\beta}_2) + se(\hat{\beta}_1)} \xrightarrow{d} \mathcal{N}(0, 1)$

The more general version of this would involve a Wald statistic.

Rejecting is very important, or not i.e. was your choice of β_0 good enough? What can go wrong?

Rejecting is very important, or not i.e. was your choice of β_0 good enough? What can go wrong?

- False positives - Your null hypothesis was actually pretty good, but you reject β_0 anyway. Just a weird sample. The test is too aggressive.
- False negatives - Your null hypothesis is bad, but you don't reject β_0. Not aggressive enough at rejecting.

False positives a.k.a. Type I error

- Calculate the probability of falsely rejecting a true null. This is called **size**.
- This is what you calculate when you derive p-values. The "null" is true.

False positives a.k.a. Type I error

- Calculate the probability of falsely rejecting a true null. This is called **size**.
- This is what you calculate when you derive p-values. The "null" is true.
- What is a simple way to minimize size?

Chapter 9

False positives a.k.a. Type I error

- Calculate the probability of falsely rejecting a true null. This is called **size**.
- This is what you calculate when you derive p-values. The "null" is true.
- What is a simple way to minimize size? Never reject!

False positives a.k.a. Type I error

- Calculate the probability of falsely rejecting a true null. This is called **size**.
- This is what you calculate when you derive p-values. The "null" is true.
- What is a simple way to minimize size? Never reject!

False negatives a.k.a. Type II error

- What's the probability of accepting the null, when it's wrong? This is called **power**.
- Harder to calculate this one, because there are more options for the probability of $\hat{\beta}$ i.e. any DGP except β_0.
- In most cases, **consistency** of the test is enough i.e. with enough data, you're sure to reject a false null, regardless of what the true parameters are.

False positives a.k.a. Type I error

- Calculate the probability of falsely rejecting a true null. This is called **size**.
- This is what you calculate when you derive p-values. The "null" is true.
- What is a simple way to minimize size? Never reject!

False negatives a.k.a. Type II error

- What's the probability of accepting the null, when it's wrong? This is called **power**.
- Harder to calculate this one, because there are more options for the probability of $\hat{\beta}$ i.e. any DGP except β_0.
- In most cases, **consistency** of the test is enough i.e. with enough data, you're sure to reject a false null, regardless of what the true parameters are.
- What's a simple way to minimize power?

Chapter 9

False positives a.k.a. Type I error

- Calculate the probability of falsely rejecting a true null. This is called **size**.
- This is what you calculate when you derive p-values. The "null" is true.
- What is a simple way to minimize size? Never reject!

False negatives a.k.a. Type II error

- What's the probability of accepting the null, when it's wrong? This is called **power**.
- Harder to calculate this one, because there are more options for the probability of $\hat{\beta}$ i.e. any DGP except β_0.
- In most cases, **consistency** of the test is enough i.e. with enough data, you're sure to reject a false null, regardless of what the true parameters are.
- What's a simple way to minimize power? Always reject!

False positives a.k.a. Type I error

- Calculate the probability of falsely rejecting a true null. This is called **size**.
- This is what you calculate when you derive p-values. The "null" is true.
- What is a simple way to minimize size? Never reject!

False negatives a.k.a. Type II error

- What's the probability of accepting the null, when it's wrong? This is called **power**.
- Harder to calculate this one, because there are more options for the probability of $\hat{\beta}$ i.e. any DGP except β_0.
- In most cases, **consistency** of the test is enough i.e. with enough data, you're sure to reject a false null, regardless of what the true parameters are.
- What's a simple way to minimize power? Always reject!

9.4 Let W be a Wald statistic for $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$, where θ is $q \times 1$. Since $W \to_d \chi^2_q$ under H_0 , someone suggests the test "Reject H_0 if $W < c_1$ or $W > c_2$, where c_1 is the $\alpha/2$ quantile of χ^2_q and c_2 is the $1 - \alpha/2$ quantile of χ^2_q

9.4 Show that the asymptotic size of the test is α.

9.4 Show that the asymptotic size of the test is α.

- Apply theorem 9.3 $W \rightarrow d\chi^2_q$ under H_0
- Size is the probability of a False rejection i.e.

$$Pr(W \geq c_1) + Pr(W \leq c_2)$$

- As a result the assymptotic size is

$$\lim_{n\to\infty} Pr(W \leq c_1) + Pr(W \geq c_2)$$

$$\lim_{n\to\infty} Pr(W \leq c_1) + 1 - Pr(W < c_2)$$

- Since W has a χ^2_q distribution as $n \rightarrow \infty$

$$\lim_{n\to\infty} Pr(W \leq c_1) + 1 - Pr(W > c_2) = \alpha/2 + 1 - (1 - \alpha/2) = \alpha$$

9.4 Is this a good test of H_0 versus H_1? Why or why not?

9.4 Is this a good test of H_0 versus H_1? Why or why not?

- Note quite. The test is consistent Theorem 9.9. For $\theta \neq 0$, then $W \to \infty$

9.4 Is this a good test of H_0 versus H_1? Why or why not?

- Note quite. The test is consistent Theorem 9.9. For $\theta \neq 0$, then $W \to \infty$
- Pick c_1' as the α quantile.
- The test is still consistent and has size α, but local rejection probability is higher for any n

$$Pr_n(W \leq c_1') > Pr_n(W \leq c_1)$$

9.1 Prove that if an additional regressor X_{k+1} is added to X, Theil's adjusted $\bar{R}^2$ increases if and only if $|T_{k+1}| > 1$ where $T_{k+1} = \hat{\beta}_{k+1}/s(\beta_{k+1})$ is the t-ratio for $\hat{\beta}_{k+1}$ and

$$s(\hat{\beta}_{k+1}) = \left(s^2[(X'X)^{-1}]_{k+1,k+1}\right)^{1/2}$$

is the homoskedasticity-formula standard error

9.1 Prove that if an additional regressor X_{k+1} is added to X, Theil's adjusted $\bar{R}^2$ increases if and only if $|T_{k+1}| > 1$ where $T_{k+1} = \hat{\beta}_{k+1}/s(\beta_{k+1})$ is the t-ratio for $\hat{\beta}_{k+1}$ and

$$s(\hat{\beta}_{k+1}) = \left(s^2[(X'X)^{-1}]_{k+1,k+1}\right)^{1/2}$$

is the homoskedasticity-formula standard error

- Theil's $\bar{R}^2$ is

$$\bar{R}^2 = 1 - \frac{(n-1)\sum_{i=1}^n \hat{e}_i^2}{(n-k)\sum_{i=1}^n (y_i - \bar{y})^2}$$

Let SST be $S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$

$$(1 - \bar{R}^2) = \frac{n-1}{n-k}\left(1 - 1 + \frac{\sum_{i=1}^n \hat{e}_i^2}{S_{yy}}\right) = \frac{n-1}{S_{yy}} \frac{\sum_{i=1}^n \hat{e}_i^2}{n-k}$$

9.1

Note that $s^2 = \frac{\sum_{i=1}^{n} \hat{e}_i^2}{n-k}$ for homoskedastic standard errors.

$$(1 - \bar{R}^2) = \frac{n-1}{S_{yy}} s^2$$

$$T_{k+1} = \frac{\hat{\beta}_{k+1}}{\left(s^2[(X'X)^{-1}]_{k+1,k+1}\right)^{1/2}}$$

9.1

Note that $s^2 = \frac{\sum_{i=1}^n \hat{e}_i^2}{n-k}$ for homoskedastic standard errors.

$$(1 - \bar{R}^2) = \frac{n-1}{S_{yy}} s^2$$

$$T_{k+1} = \frac{\hat{\beta}_{k+1}}{\left(s^2[(X'X)^{-1}]_{k+1,k+1}\right)^{1/2}}$$

- $\bar{R}^2$ is inversely related with s^2. First equation.
- s^2 is inversely related with T_{k+1}. Second equation.
- $\bar{R}^2$ is positively related with T_{k+1}.