ECON 61001: *Lecture 2*

Alastair R. Hall

The University of Manchester

Semester 1, 2020-21

Outline of today's lecture

- Statistical properties of OLS
  - mean
  - variance
  - sampling distribution
- Confidence intervals
  - coefficients
  - prediction

Parameter estimation: background

Within Classical statistics paradigm, the desired properties for an estimator, $\hat{\theta}_T$, of an unknown $p \times 1$ parameter vector, θ_0, are:

- Unbiasedness: $E[\hat{\theta}_T] = \theta_0$.
- efficiency: $Var[\hat{\theta}_T]$ is an efficient (unbiased) estimator of θ_0 iff

$$Var[\tilde{\theta}_T] - Var[\hat{\theta}_T] = p.s.d.$$

where $\tilde{\theta}_T$ is any other unbiased estimator of θ_0.

Recall that our model is:

$$y = X\beta_0 + u$$

where

- *CA1*: true model is: $y = X\beta_0 + u$.
- *CA2*: X is *fixed in repeated samples*.
- *CA3*: X is rank k.
- *CA4*: $E[u] = 0$.
- *CA5*: $Var[u] = \sigma_0^2 I_T$.
- *CA6*: $u \sim$ Normal.

OLS estimator:

$$\hat{\beta}_T = (X'X)^{-1}X'y.$$

Want to derive sampling distribution of $\hat{\beta}_T$. Will do this in stages, deriving first:

- mean, $E[\hat{\beta}_T]$.
- variance, $Var[\hat{\beta}_T]$.

To this end, we substitute for y using *CA1* then:

$$\hat{\beta}_T = \beta_0 + (X'X)^{-1}X'u.$$

Mean

$$E[\hat{\beta}_T] = E\left[\beta_0 + (X'X)^{-1}X'u\right].$$

From *CA2*, this expectation can be written as:

$$E[\hat{\beta}_T] = \beta_0 + (X'X)^{-1}X'E[u].$$

So, using *CA4*, we have:

$$E[\hat{\beta}_T] = \beta_0.$$

$\Rightarrow$ $\hat{\beta}_T$ is an unbiased estimator of β_0.

Variance

$$Var[\hat{\beta}_T] \;=\; E\left[\hat{\beta}_T \;-\; E[\hat{\beta}_T]\right]\left[\hat{\beta}_T \;-\; E[\hat{\beta}_T]\right]'.$$

Using $E[\hat{\beta}_T] = \beta_0$, formula for $\hat{\beta}_T$ and $CA2$, it follows that:

$$\begin{aligned} Var[\hat{\beta}_T] &= E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}\right], \\ &= (X'X)^{-1}X'\,E[uu']\,X(X'X)^{-1}. \end{aligned}$$

From $CA4$ and $CA5$, it follows that:

$$Var[\hat{\beta}_T] \;=\; (X'X)^{-1}X'\sigma_0^2 I_T X(X'X)^{-1} \;=\; \sigma_0^2(X'X)^{-1}$$

Gauss-Markov Theorem

Under assumptions CA1 - CA5, OLS is the Best Linear (in y) Unbiased Estimator (BLUE) of β_0 in the sense that

$$Var[\tilde{\beta}_T] - Var[\hat{\beta}_T] = p.s.d$$

where $\tilde{\beta}_T$ is any other linear (in y) unbiased estimator of β_0.

Proof: Let $\tilde{\beta}_T = Dy$ where $D = (X'X)^{-1}X' + C$ for some $k \times T$ matrix of constants C. Note that $E[\tilde{\beta}_T] = \beta_0$ implies $CX = 0$.

Using similar arguments to OLS,

$$Var[\tilde{\beta}_T] = \sigma_0^2 DD' = \sigma_0^2\{(X'X)^{-1} + CC'\},$$

and so,

$$Var[\tilde{\beta}_T] - Var[\hat{\beta}_T] = \sigma_0^2 CC'$$

which is psd by construction.

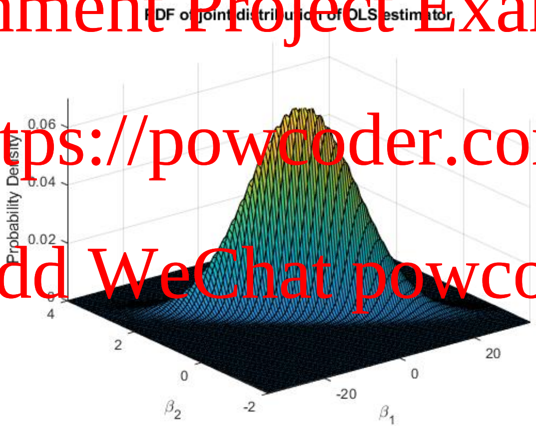
Recall

$$\hat{\beta}_T \; = \; \beta_0 + (X'X)^{-1}X'u$$

so, from *CA2* + *CA6*, $\hat{\beta}_T$ is linear combination of rv's with Normal distribution, and so via Lemma 2.1 (Lecture Notes)

$$\hat{\beta}_T \sim N\left(\beta_0, \sigma_0^2(X'X)^{-1}\right).$$

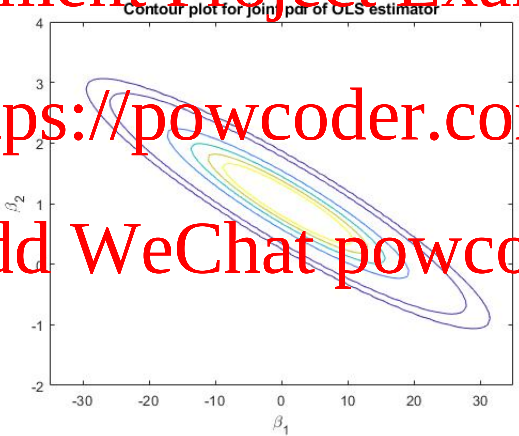
Example from video on *Sampling distributions* (with $T = 5$).



Which in this case looks just like Napoleon's hat!

Sampling distribution: example

Shape better revealed by contour plot in which each ring connects points with same pdf value.



$$Var[\hat{\beta}_T] = \begin{bmatrix} Var[\hat{\beta}_{T,1}] & Cov[\hat{\beta}_{T,1}, \hat{\beta}_{T,2}] \\ Cov[\hat{\beta}_{T,1}, \hat{\beta}_{T,2}] & Var[\hat{\beta}_{T,2}] \end{bmatrix},$$

$$= \begin{bmatrix} 85.02 & -5.29 \\ -5.29 & 0.39 \end{bmatrix}.$$

Recall that under Assumption CA1-CA6:

$$\hat{\beta}_T \sim N\left(\beta_0, \sigma_0^2 (X'X)^{-1}\right).$$

To use this result for inference, we need an estimator of σ_0^2.

OLS estimator is:

$$\hat{\sigma}_T^2 = \frac{e'e}{T-k}$$

We now show $E[\hat{\sigma}_T^2] = \sigma_0^2$.

Consider inference about $\beta_{0,i}$ based on $\hat{\beta}_{T,i}$.

We have

$$\hat{\beta}_{T,i} \sim N\left(\beta_{0,i}, \sigma_0^2 m_{i,i}\right),$$

where $m_{i,i}$ is the i^{th} main diagonal element of $(X'X)^{-1}$, and so

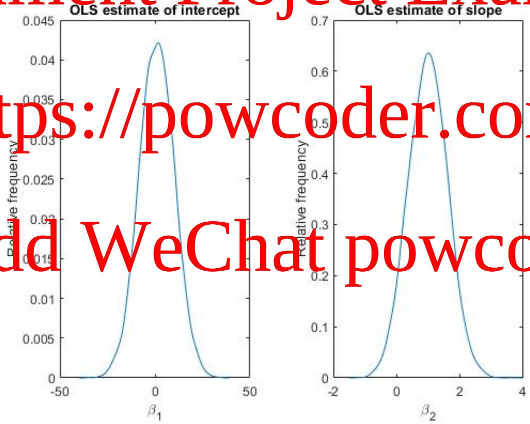
$$\frac{\hat{\beta}_{T,i} - \beta_{0,i}}{\sigma_0 \sqrt{m_{i,i}}} \sim N(0,1).$$

If we replace σ_0 by $\hat{\sigma}_T$ then we

$$\frac{\hat{\beta}_{T,i} - \beta_{0,i}}{\hat{\sigma}_T \sqrt{m_{i,i}}} \sim \text{Student's t distribution with T-k df}$$

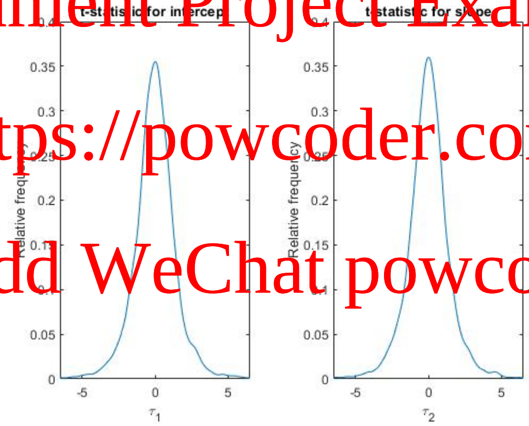
Example from video on *Sampling distributions* (with $T = 5$).

Simulated sampling distributions:

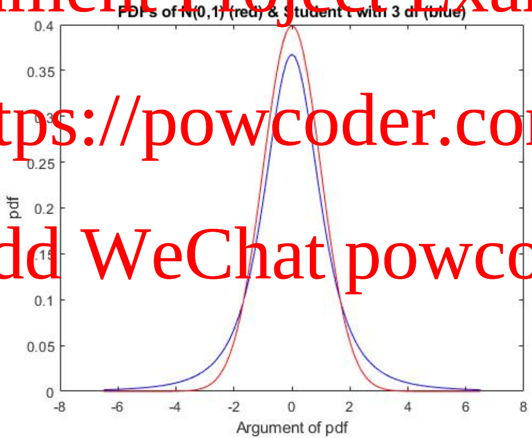


Example

Simulated sampling distribution of t-statistics



Comparison of Student's t distribution with 3 df to standard normal distribution.



Confidence interval for $\beta_{0,i}$

A $100(1-\alpha)\%$ confidence interval for $\beta_{0,i}$ is:

$$\hat{\beta}_{T,i} \pm \tau_{T-k}(1-\alpha/2)\hat{\sigma}_T\sqrt{m_{i,i}}$$

where $\tau_{T-k}(1-\alpha/2)$ is $100(1-\alpha/2)^{th}$ percentile of Student's t distribution with $T-k$ df.

From Lecture 1:

$$\hat{y}_t = \text{controls} - 0.030 * belt_t + 0.0671 * mph_t$$

What do $\hat{\beta}_{belt}$ and $\hat{\beta}_{mph}$ tell us about $\beta_{belt,0}$ and $\beta_{mph,0}$?

- Variability of estimator is: $s.e.(\hat{\beta}_{belt}) = 0.023$
- Leads to 95% confidence interval for $\beta_{belt,0}$: $(-0.076, 0.017)$

- Variability of estimator is: $s.e.(\hat{\beta}_{mph}) = 0.021$
- Leads to 95% confidence interval for $\beta_{mph,0}$: $(0.026, 0.108)$

Suppose we know x_{T+1} but not y_{T+1}. (We assume model satisfies *CA1-CA6* for $t = 1, 2 \ldots T+1$).

Then natural predictor of y_{T+1} is:

$$y^p_{T+1} = x'_{T+1}\hat{\beta}_T$$

and prediction error is:

$$\begin{aligned} e^p_{T+1} &= y_{T+1} - y^p_{T+1} = x'_{T+1}\beta_0 + u_{T+1} - x'_{T+1}\hat{\beta}_T \\ &= u_{T+1} - x'_{T+1}(\hat{\beta}_T - \beta_0) \end{aligned}$$

So

$$e^p_{T+1} \sim N\left(0, \sigma_0^2(1 + x'_{T+1}(X'X)^{-1}x_{T+1})\right)$$

This leads to the $100(1-\alpha)\%$ prediction interval for y_{T+1}:

$$y^{p}_{T+1} \;\pm\; \tau_{T-k}(1-\alpha/2)\,\hat{\sigma}_T\,\sqrt{(1\,+\,x'_{T+1}(X'X)^{-1}x_{T+1})\,)}$$

Example:

- Suppose wish to predict fatalities for Jan 1990 ($weekends = 12$) if $unem = 5$:

- $y^{p}_{T+1} = 0.754122$
- 95% prediction interval for $y_{1990.1}$ is: $(0.629, 0.879)$.

Further reading

- Notes: Sections 2.3-2.6
- See Greene:
  - Mean, Variance - Sections 4.3.1, 4.3.4, C.5.1
  - Gauss-Markov Theorem - Section 4.3.5
  - Confidence Intervals, Section 4.5.1
  - Prediction, Section 4.6.