Advanced Econometric Methods EMET3011/8014

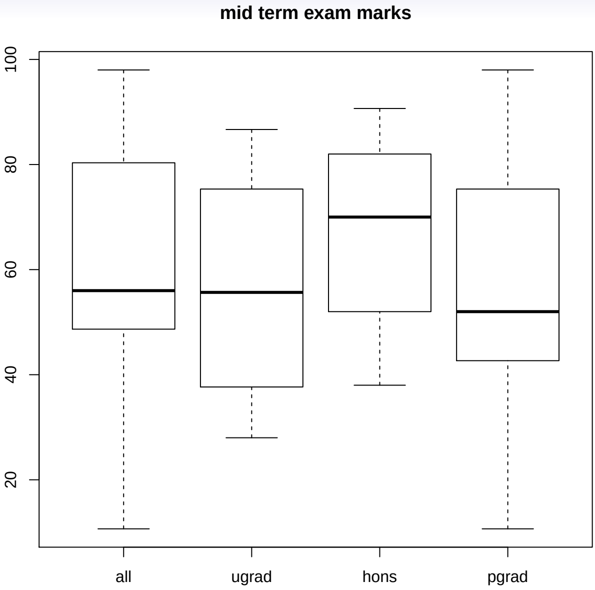
Lecture 10

John Stachurski

Semester 1, 2011

Announcements/Reminders

- Please get a fresh copy of the course notes PDF
- Midterm exam marks will be posted on home page today



mid term exam marks
100
80
60
40
20
all
ugrad
hons
pgrad

Today's Lecture

- Goodness of Fit
- The Standard OLS Model
- Gaussian OLS
- Endogeneity Bais

Goodness of Fit

Given $\mathbf{X}$ and $\mathbf{y}$, the **coefficient of determination** or R-squared is

$$R^2 := \frac{\mathrm{ESS}}{\mathrm{TSS}} = \frac{\|\mathbf{P}\mathbf{y}\|^2}{\|\mathbf{y}\|^2}$$

Exercise:

- Show that $0 \leq R^2 \leq 1$ always holds (Hint: Use OPT II)
- Show that if $R^2 = 1$, then $\mathbf{y} \in \mathrm{rng}(\mathbf{X})$

R^2 often viewed as a one-number summary of the success of a regression model

This is not really a good idea. . .

One problem: R squared increases montonically with new regressors

Fact. If $\mathbf{X}_a \subset \mathbf{X}_b$, in the sense that every column of $\mathbf{X}_a$ is also a column of $\mathbf{X}_b$, then $R_a^2 \leq R_b^2$

Proof: If $\mathbf{X}_a \subset \mathbf{X}_b$, then $\operatorname{rng}(\mathbf{X}_a) \subset \operatorname{rng}(\mathbf{X}_b)$

$$\therefore \quad \mathbf{P}_a \mathbf{P}_b \mathbf{y} = \mathbf{P}_a \mathbf{y}$$

$$\therefore \quad \frac{R_a^2}{R_b^2} = \left(\frac{\|\mathbf{P}_a \mathbf{y}\|}{\|\mathbf{P}_b \mathbf{y}\|} \right)^2 = \left(\frac{\|\mathbf{P}_a \mathbf{P}_b \mathbf{y}\|}{\|\mathbf{P}_b \mathbf{y}\|} \right)^2 \leq 1$$

(Why does the last inequality hold?)

Intuitition:

- We are increasing the column space of $\mathbf{X}$
- This always brings $\mathbf{Py}$ (weakly) closer to $\mathbf{y}$
- Which increases R^2

Often we can make R^2 arbitrarily close to one by putting in more and more regressors

Can be done by

- Adding new variables
- Transformations of existing variables
- Some combination of the two

Let's illustrate by simulation...

Let x_n and y_n to be independent draws from $U[0,1]$

1. Draw 25 pairs (y_n, x_n)
2. Fit polynomial of degree K to data for $K = 1, 2, \ldots$

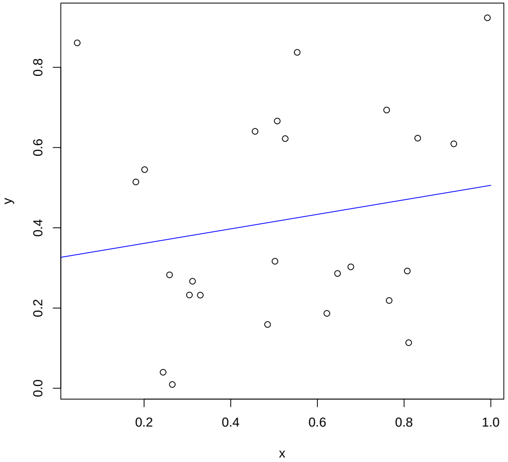


Figure: $K = 1$

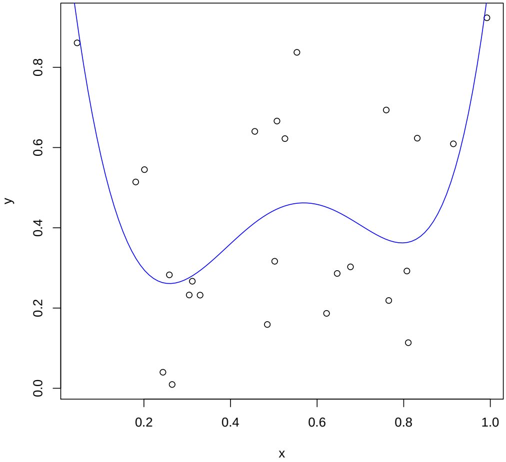


Figure: $K = 5$

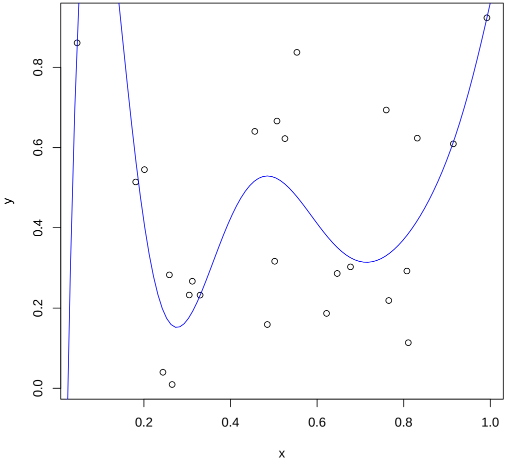


Figure: $K = 8$

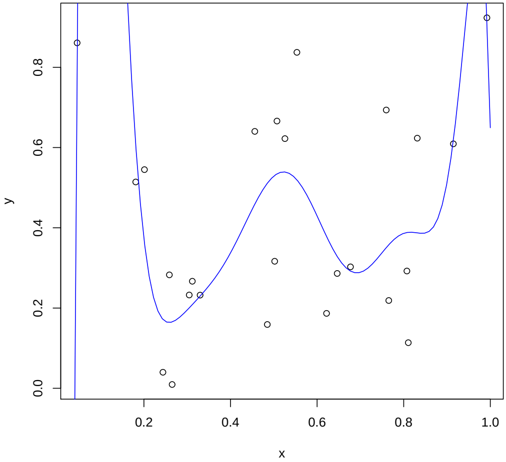


Figure: $K = 10$

Despite the fact that x and y are completely unrelated,

- $R^2 = 0.87$ at $K = 10$
- $R^2 = 0.95$ at $K = 25$

Here $R^2 \to 1$ because empirical risk is going to zero:

$$
\begin{aligned}
R^2 &= \frac{\|\mathbf{P}\mathbf{y}\|^2}{\|\mathbf{y}\|^2} \\
&= \frac{\|\mathbf{y}\|^2 - \|\mathbf{M}\mathbf{y}\|^2}{\|\mathbf{y}\|^2} \qquad (\because \text{TSS} = \text{ESS} + \text{SSR}) \\
&= 1 - \|\mathbf{y}\|^{-2}\|\mathbf{M}\mathbf{y}\|^2 \\
&= 1 - \|\mathbf{y}\|^{-2}\sum_{n=1}^{N}(y_n - \hat{\boldsymbol{\beta}}'\mathbf{x}_n)^2
\end{aligned}
$$

But low empirical risk is not the same thing as low risk

Recall our earlier simulation experiment with model

$$x \sim U[-1,1] \quad \text{and} \quad y = \cos(\pi x) + u \quad \text{where} \quad u \sim N(0,1)$$

Data fitted by polynomials of increasing degree

As degree increases

- empirical risk goes to zero, but
- risk <u>explodes</u>

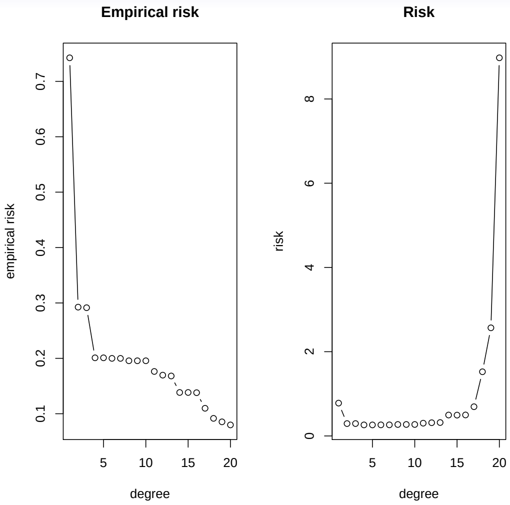


Figure: Risk and empirical risk as complexity increases

Moral: Good in-sample fit is not the main story

Inductive/statistical learning is about generalization

Generalization means generalizing beyond the sample

Effective generalization means we can predict out-of-sample

Model assessment must be based on this criterion

The Classical OLS Assumptions

For the remainder of the lecture we

- Adopt the classical OLS assumptions
- Describe performance of linear LSQ under these assumptions

The assumptions are very strong

But form the bread and butter of standard econometrics

The Linear Model Assumption. The input-output pairs all satisfy

$$y_n = \boldsymbol{\beta}' \mathbf{x}_n + u_n$$

where

- $\boldsymbol{\beta}$ is an unknown $K \times 1$ vector of parameters
- $u_1, \ldots, u_N$ are unobservable random variables

Letting $\mathbf{u} := (u_1, u_2, \ldots, u_N)'$, the N equations become

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{1}$$

Exercise: Show that (1) implies $\mathbf{My} = \mathbf{Mu}$ and $\textsc{ssr} = \mathbf{u}'\mathbf{Mu}$

First Moment Assumption. $\mathbb{E}[\mathbf{u} \,|\, \mathbf{X}] = \mathbf{0}$

Fact. Given this assumption, we have

1. $\mathbb{E}[\mathbf{u}] = \mathbf{0}$
2. $\mathbb{E}[u_m \,|\, x_{nk}] = 0$ for any m, n, k
3. $\mathbb{E}[u_m x_{nk}] = 0$ for any m, n, k
4. $\mathrm{cov}[u_m, x_{nk}] = 0$ for any m, n, k

Proof of part 1:

$$\mathbb{E}[\mathbf{u}] = \mathbb{E}[\mathbb{E}[\mathbf{u} \,|\, \mathbf{X}]] = \mathbb{E}[\mathbf{0}] = \mathbf{0}$$

Exercise: Verify remaining facts (solutions in course notes)

Second Moment Assumption $\mathbb{E}\left[\mathbf{u}\mathbf{u}' \,|\, \mathbf{X}\right] = \sigma^2 \mathbf{I}$

Here σ is an unknown positive constant

We now have

$$\mathrm{var}[\mathbf{u} \,|\, \mathbf{X}] := \mathbb{E}\left[\mathbf{u}\mathbf{u}' \,|\, \mathbf{X}\right] - \mathbb{E}\left[\mathbf{u} \,|\, \mathbf{X}\right]\mathbb{E}\left[\mathbf{u}' \,|\, \mathbf{X}\right] = \sigma^2 \mathbf{I}$$

Fact. Under our assumptions,

1. $\mathrm{var}[\mathbf{u}] = \mathbb{E}\left[\mathbf{u}\mathbf{u}'\right] = \sigma^2 \mathbf{I}$
2. Shocks are **homoskedastic**: $\mathbb{E}\left[u_i^2 \,|\, \mathbf{X}\right] = \mathbb{E}\left[u_j^2 \,|\, \mathbf{X}\right] = \sigma^2$
3. Distinct shocks are uncorrelated: $\mathbb{E}\left[u_i u_j \,|\, \mathbf{X}\right] = 0$ when $i \neq j$

Exercise: Check these facts

To repeat, our assumptions are

1. $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$
2. $\mathbb{E}\left[\mathbf{u} \,|\, \mathbf{X}\right] = \mathbf{0}$
3. $\mathbb{E}\left[\mathbf{u}\mathbf{u}' \,|\, \mathbf{X}\right] = \sigma^2 \mathbf{I}$

Unless otherwise stated, these assumptions hold throughout the lecture

The OLS Estimators

Standard estimator of $\boldsymbol{\beta}$ and σ^2 are

$$\hat{\boldsymbol{\beta}} := (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad \text{and} \quad \hat{\sigma}^2 := \frac{\mathrm{SSR}}{N-K}$$

A useful expression for $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

A useful expression for $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{\mathbf{u}'\mathbf{M}\mathbf{u}}{N-K}$$

Bias

Under the OLS assumptions, both $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are unbiased

Thm. $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \mathbb{E}[\hat{\boldsymbol{\beta}} \,|\, \mathbf{X}] = \boldsymbol{\beta}$

Proof: Using $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$ we obtain

$$\begin{aligned}
\mathbb{E}[\hat{\boldsymbol{\beta}} \,|\, \mathbf{X}] &= \mathbb{E}[\boldsymbol{\beta} \,|\, \mathbf{X}] + \mathbb{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \,|\, \mathbf{X}] \\
&= \boldsymbol{\beta} + \mathbb{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \,|\, \mathbf{X}] \\
&= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\mathbf{u} \,|\, \mathbf{X}] \\
&= \boldsymbol{\beta}
\end{aligned}$$

$$\therefore \quad \mathbb{E}[\hat{\boldsymbol{\beta}}] = \mathbb{E}[\mathbb{E}[\hat{\boldsymbol{\beta}} \,|\, \mathbf{X}]] = \mathbb{E}[\boldsymbol{\beta}] = \boldsymbol{\beta}$$

Thm. $\mathbb{E}\left[\hat{\sigma}^2\right] = \mathbb{E}\left[\hat{\sigma}^2 \mid \mathbf{X}\right] = \sigma^2$

Proof: Note first that $\text{trace}(\mathbf{M}) = N - K$, because

$$\begin{aligned}\text{trace}(\mathbf{M}) &= \text{trace}(\mathbf{I}_N - \mathbf{P}) \\ &= \text{trace}(\mathbf{I}_N) - \text{trace}(\mathbf{P}) = N - \text{trace}(\mathbf{P})\end{aligned}$$

and

$$\begin{aligned}\text{trace}(\mathbf{P}) &= \text{trace}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \\ &= \text{trace}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}] = \text{trace}[\mathbf{I}_K] = K\end{aligned}$$

(By which rule about trace?)

Letting $m_{ij}(\mathbf{X})$ be the i, j-th element of $\mathbf{M}$, we have

$$\mathbb{E}\left[\mathbf{u}'\mathbf{M}\mathbf{u} \mid \mathbf{X}\right] = \mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} u_i u_j m_{ij}(\mathbf{X}) \mid \mathbf{X}\right]$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} m_{ij}(\mathbf{X})\mathbb{E}\left[u_i u_j \mid \mathbf{X}\right] = \sum_{n=1}^{N} m_{nn}(\mathbf{X})\sigma^2$$

$$\therefore \quad \mathbb{E}\left[\mathrm{SSR} \mid \mathbf{X}\right] = \mathbb{E}\left[\mathbf{u}'\mathbf{M}\mathbf{u} \mid \mathbf{X}\right] = \mathrm{trace}(\mathbf{M})\sigma^2$$

$$\therefore \quad \mathbb{E}\left[\hat{\sigma}^2 \mid \mathbf{X}\right] := \mathbb{E}\left[\frac{\mathrm{SSR}}{N-K} \mid \mathbf{X}\right] = \frac{\mathrm{trace}(\mathbf{M})\sigma^2}{N-K} = \sigma^2$$

$$\therefore \quad \mathbb{E}\left[\hat{\sigma}^2\right] = \mathbb{E}\left[\mathbb{E}\left[\hat{\sigma}^2 \mid \mathbf{X}\right]\right] = \sigma^2$$

Variance of $\hat{\boldsymbol{\beta}}$

Thm. $\operatorname{var}[\hat{\boldsymbol{\beta}} \,|\, \mathbf{X}] = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$

Proof: If $\mathbf{A} := (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, then $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{A}\mathbf{u}$, and

$$\operatorname{var}[\hat{\boldsymbol{\beta}} \,|\, \mathbf{X}] = \operatorname{var}[\boldsymbol{\beta} + \mathbf{A}\mathbf{u} \,|\, \mathbf{X}] = \operatorname{var}[\mathbf{A}\mathbf{u} \,|\, \mathbf{X}]$$

Since $\mathbf{A}$ is a function of $\mathbf{X}$, we have

$$\operatorname{var}[\mathbf{A}\mathbf{u} \,|\, \mathbf{X}] = \mathbf{A} \operatorname{var}[\mathbf{u} \,|\, \mathbf{X}] \mathbf{A}' = \mathbf{A}(\sigma^2 \mathbf{I})\mathbf{A}'$$

But $\mathbf{A}(\sigma^2 \mathbf{I})\mathbf{A}' = \sigma^2 \mathbf{A}\mathbf{A}' = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$

$$\therefore \quad \operatorname{var}[\hat{\boldsymbol{\beta}} \,|\, \mathbf{X}] = \operatorname{var}[\mathbf{A}\mathbf{u} \,|\, \mathbf{X}] = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

The Gauss-Markov Theorem

We have shown that $\hat{\boldsymbol{\beta}}$ is unbiased

Next step: Show that $\hat{\boldsymbol{\beta}}$ has low variance

"Low" means relative to other linear unbiased estimators

- Linearity of estimator $\mathbf{b}$ means that $\mathbf{b} = \mathbf{C}\mathbf{y}$ for some $\mathbf{C}$
- Here $\mathbf{C}$ may depend on $\mathbf{X}$

Key result: The Gauss-Markov theorem

- Tells us that $\hat{\boldsymbol{\beta}}$ is BLUE

Theorem (Gauss-Markov) If $\mathbf{b}$ is linear and unbiased for $\boldsymbol{\beta}$, then $\text{var}[\mathbf{b} \,|\, \mathbf{X}] - \text{var}[\hat{\boldsymbol{\beta}} \,|\, \mathbf{X}]$ is nonnegative definite

Exercise: $\text{var}[\mathbf{b} \,|\, \mathbf{X}] - \text{var}[\hat{\boldsymbol{\beta}} \,|\, \mathbf{X}]$ nonneg def implies that

- $\text{var}[\ell(\mathbf{b}) \,|\, \mathbf{X}] \geq \text{var}[\ell(\hat{\beta}) \,|\, \mathbf{X}]$ for any linear $\ell \colon \mathbb{R}^K \to \mathbb{R}$
- and hence $\text{var}[b_k \,|\, \mathbf{X}] \geq \text{var}[\hat{\beta}_k \,|\, \mathbf{X}]$ for all k

Clarification:

- Unbiasedness means $\mathbb{E}\left[\mathbf{b} \,|\, \mathbf{X}\right] = \boldsymbol{\beta}$ for <u>any</u> given $\boldsymbol{\beta} \in \mathbb{R}^K$

Proof: If $\mathbf{b} = \mathbf{C}\mathbf{y}$, $\mathbf{A} := (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{D} := \mathbf{C} - \mathbf{A}$, then

$$\mathbf{b} = \mathbf{C}\mathbf{y} = \mathbf{D}\mathbf{y} + \mathbf{A}\mathbf{y} = \mathbf{D}(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) + \hat{\boldsymbol{\beta}} = \mathbf{D}\mathbf{X}\boldsymbol{\beta} + \mathbf{D}\mathbf{u} + \mathbf{A}\mathbf{u} + \boldsymbol{\beta}$$

Using unbiasedness, we obtain

$$\boldsymbol{\beta} = \mathbb{E}\left[\mathbf{b} \,|\, \mathbf{X}\right] = \mathbf{D}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}$$

$$\therefore \quad \mathbf{D}\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$$

$$\therefore \quad \mathbf{b} = \mathbf{D}\mathbf{u} + \mathbf{A}\mathbf{u} + \boldsymbol{\beta}$$

In fact we have shown that

$$\mathbf{D}\mathbf{X}\boldsymbol{\beta} = \mathbf{0} \quad \text{for } \underline{\text{any}} \quad \boldsymbol{\beta} \in \mathbb{R}^K$$

$$\therefore \quad \mathbf{D}\mathbf{X} = \mathbf{0} \quad (\because \mathbf{z}'\boldsymbol{\beta} = 0, \ \forall \boldsymbol{\beta} \implies \mathbf{z} = \mathbf{0})$$

Using $\mathbf{b} = \mathbf{Du} + \mathbf{Au} + \boldsymbol{\beta}$ we obtain

$$\begin{aligned}
\operatorname{var}[\mathbf{b} \mid \mathbf{X}] &= \operatorname{var}[(\mathbf{D} + \mathbf{A})\mathbf{u} \mid \mathbf{X}] \\
&= (\mathbf{D} + \mathbf{A}) \operatorname{var}[\mathbf{u} \mid \mathbf{X}] (\mathbf{D} + \mathbf{A})' \\
&= \sigma^2 (\mathbf{D} + \mathbf{A})(\mathbf{D}' + \mathbf{A}') \\
&= \sigma^2 (\mathbf{DD}' + \mathbf{DA}' + \mathbf{AD}' + \mathbf{AA}')
\end{aligned}$$

On one hand, $\mathbf{DA}' = \mathbf{DX}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0}$

On the other hand, $\mathbf{AA}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}$

$$\therefore \quad \operatorname{var}[\mathbf{b} \mid \mathbf{X}] = \sigma^2[\mathbf{D}'\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}] = \sigma^2 \mathbf{D}'\mathbf{D} + \operatorname{var}[\hat{\boldsymbol{\beta}} \mid \mathbf{X}]$$

The proof is done (why?)

Gaussian OLS

To perform inference (tests, etc.) we need to determine the distribution of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$

This can be done either by

1. Finite Sample Methods: Imposing more structure on the distribution of $\mathbf{u}$, such as normality
2. Large Sample Methods: Based on CLT

First we treat the former

Note that the large sample method assumptions are much weaker

Gaussian OLS

Assumption N: $\mathbf{X}$ and $\mathbf{u}$ are independent, and $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$

Assumption N implies both $\mathbb{E}[\mathbf{u} \mid \mathbf{X}] = \mathbf{0}$ and $\mathbb{E}[\mathbf{u}\mathbf{u}' \mid \mathbf{X}] = \sigma^2 \mathbf{I}$

Theorem: Under assumption N,

$$\hat{\boldsymbol{\beta}} \text{ given } \mathbf{X} \text{ is } \mathcal{N}(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$$

Proof: This follows from assumption N and $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$

(Supply details)

We have

$$\hat{\boldsymbol{\beta}} \text{ given } \mathbf{X} \text{ is } \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

$$\therefore \quad \mathbf{e}_k'\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\mathbf{e}_k'\boldsymbol{\beta}, \sigma^2\mathbf{e}_k'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{e}_k)$$

$$\therefore \quad \hat{\beta}_k \sim \mathcal{N}(\beta_k, \sigma^2\mathbf{e}_k'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{e}_k)$$

$$\therefore \quad z_k := \frac{\hat{\beta}_k - \beta_k}{\sigma\sqrt{\mathbf{e}_k'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{e}_k}} \sim \mathcal{N}(0,1)$$

The t-test

Let $\beta_k^0 \in \mathbb{R}$, and consider null hypothesis

$$H_0 \colon \beta_k = \beta_k^0$$

Recall that standard deviation of $\hat{\beta}_k$ is $\sigma\sqrt{\mathbf{e}_k'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{e}_k}$

Standard error of $\hat{\beta}_k$ is the estimated standard deviation

$$\mathrm{se}(\hat{\beta}_k) := \sqrt{\hat{\sigma}^2\mathbf{e}_k'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{e}_k}$$

The **t-statistic** for $\hat{\beta}_k$ is

$$t_k := \frac{\hat{\beta}_k - \beta_k^0}{\mathrm{se}(\hat{\beta}_k)}$$

Theorem: If assumption N and H_0 hold, then, given $\mathbf{X}$,

$$t_k := \frac{\hat{\beta}_k - \beta_k^0}{\mathrm{se}(\hat{\beta}_k)} \sim \text{Student-t with } N-K \text{ d.f.}$$

Proof: We need to show that

$$t_k = Z\sqrt{\frac{N-K}{Q}}$$

where

- $Z \sim \mathcal{N}(0,1)$
- $Q \sim \chi^2(N-K)$
- z_k and Q are independent

Under null hypothesis we have

$$t_k := \frac{\hat{\beta}_k - \beta_k^0}{\operatorname{se}(\hat{\beta}_k)} = \frac{\hat{\beta}_k - \beta_k}{\operatorname{se}(\hat{\beta}_k)} = \frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 \mathbf{e}_k' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{e}_k (\hat{\sigma}^2 / \sigma^2)}}$$

Since $\hat{\sigma}^2 = \text{SSR}/(N-K)$, we then have

$$t_k = z_k \sqrt{\frac{N-K}{\text{SSR}/\sigma^2}} =: z_k \sqrt{\frac{N-K}{Q}} \qquad (Q := \text{SSR}/\sigma^2)$$

Remains to show that

(a) $Q \sim \chi^2(N-K)$
(b) z_k and Q are independent

Regarding (a), we need to show that if assumption N holds, then

$$Q := \frac{\text{SSR}}{\sigma^2} \sim \chi^2(N-K)$$

Proof: Since $\text{SSR} = \mathbf{u}'\mathbf{M}\mathbf{u}$ we have

$$Q = \frac{\mathbf{u}'\mathbf{M}\mathbf{u}}{\sigma^2} = (\sigma^{-1}\mathbf{u})'\mathbf{M}(\sigma^{-1}\mathbf{u})$$

Exercise: Show that r.h.s. is $\chi^2(N-K)$

Proof that z_k and Q are independent

Fact: If $\mathbf{a}$ and $\mathbf{b}$ are independent random vectors and f and g are two functions, then $f(\mathbf{a})$ and $g(\mathbf{b})$ are independent

- $z_k =$ function of $\hat{\boldsymbol{\beta}}$
- $Q =$ function of $\mathbf{Mu}$

Suffices to show that $\hat{\boldsymbol{\beta}}$ and $\mathbf{Mu}$ are independent

Since both normally distributed given $\mathbf{X}$, suffices to show their covariance is zero

$$
\begin{aligned}
\operatorname{cov}[\hat{\boldsymbol{\beta}}, \mathbf{Mu} \mid \mathbf{X}] &= \mathbb{E}\left[\hat{\boldsymbol{\beta}}\,(\mathbf{Mu})' \mid \mathbf{X}\right] - \mathbb{E}\left[\hat{\boldsymbol{\beta}} \mid \mathbf{X}\right]\mathbb{E}\left[\mathbf{Mu} \mid \mathbf{X}\right]' \\
&= \mathbb{E}\left[\hat{\boldsymbol{\beta}}\,(\mathbf{Mu})' \mid \mathbf{X}\right] \\
&= \mathbb{E}\left[\boldsymbol{\beta}(\mathbf{Mu})' + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\,(\mathbf{Mu})' \mid \mathbf{X}\right] \\
&= \mathbb{E}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\,(\mathbf{Mu})' \mid \mathbf{X}\right] \\
&= \mathbb{E}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{M} \mid \mathbf{X}\right] \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{M} \\
&= \mathbf{0}
\end{aligned}
$$

The t test: Reject H_0 if $|t_k| > c$

For desired size α, choose c to solve

$$\mathbb{P}\{|t_k| > c\} = \alpha$$

As discussed in course notes, solution is $c_\alpha = F^{-1}(1 - \alpha/2)$

- F is the Student-t cdf with $N - K$ degrees of freedom

Most common implementation of t-test is $H_0 : \beta_k = 0$, implying t-statistic

$$t_k = \frac{\hat{\beta}_k}{\mathrm{se}(\hat{\beta}_k)}$$ (sometimes called the **Z-score**)

The F-test

The t-test used to test hypotheses about individual regressors

For multiple regressors, most common test is F-test

Special case: $H_0 =$ some subset of coefficients are zero

Let $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$

- $\boldsymbol{\beta}_1$ is $K_1 \times 1$
- $\boldsymbol{\beta}_2$ is $K_2 \times 1$

Let $\mathbf{P}_1$ and $\mathbf{M}_1$ be projection, annihilator associated with $\mathbf{X}_1$

Our hypothesis: $H_0\colon \boldsymbol{\beta}_2 = \mathbf{0}$

The unrestricted regression model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}$$

Under the null $\boldsymbol{\beta}_2 = \mathbf{0}$, this becomes

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{u}$$

Test statistic for our null hypothesis is

$$F := \frac{(\text{RSSR} - \text{USSR})/K_2}{\text{USSR}/(N-K)}$$

where $\text{USSR} := \|\mathbf{M}\mathbf{y}\|^2$ and $\text{RSSR} := \|\mathbf{M}_1\mathbf{y}\|^2$

Theorem: If assumption N holds and H_0 true, then

$$F = \frac{(\text{RSSR} - \text{USSR})/K_2}{\text{USSR}/(N-K)} \sim F(K_2, N-K)$$

(Here and in proof I omit the "conditional on $\mathbf{X}$")

Proof: Let $Q_1 := (\text{RSSR} - \text{USSR})/\sigma^2$ and let $Q_2 := \text{USSR}/\sigma^2$, so that

$$F = \frac{Q_1/K_2}{Q_2/(N-K)}$$

Suffices to show that, under the null hypothesis,

(a) Q_1 is chi-squared with K_2 degrees of freedom
(b) Q_2 is chi-squared with $N-K$ degrees of freedom
(c) Q_1 and Q_2 are independent

Note that (b) was proved earlier on

Part (a) claims that $Q_1 := (\text{RSSR} - \text{USSR})/\sigma^2 \sim \chi^2(K_2)$

Proof: Under H_0,

- $\text{USSR} = \|\mathbf{My}\|^2 = \|\mathbf{M}(\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{u})\|^2 = \|\mathbf{Mu}\|^2 = \mathbf{u}'\mathbf{Mu}$, and
- $\text{RSSR} = \|\mathbf{M}_1\mathbf{y}\|^2 = \|\mathbf{M}_1(\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{u})\|^2 = \|\mathbf{M}_1\mathbf{u}\|^2 = \mathbf{u}'\mathbf{M}_1\mathbf{u}$

$$\therefore \quad \text{RSSR} - \text{USSR} = \mathbf{u}'\mathbf{M}_1\mathbf{u} - \mathbf{u}'\mathbf{Mu} = \mathbf{u}'(\mathbf{M}_1 - \mathbf{M})\mathbf{u}$$

$$\therefore \quad Q_1 = \frac{\mathbf{u}'(\mathbf{I} - \mathbf{P}_1 - \mathbf{I} + \mathbf{P})\mathbf{u}}{\sigma^2} = (\sigma^{-1}\mathbf{u})'(\mathbf{P} - \mathbf{P}_1)(\sigma^{-1}\mathbf{u})$$

Exercise: Show that $(\mathbf{P} - \mathbf{P}_1)$ is symmetric and idempotent

Can then show (exercise, see earlier proof) that

$$\begin{aligned} \operatorname{rank}(\mathbf{P} - \mathbf{P}_1) &= \operatorname{trace}(\mathbf{P} - \mathbf{P}_1) \\ &= \operatorname{trace}(\mathbf{P}) - \operatorname{trace}(\mathbf{P}_1) = K - K_1 = K_2 \end{aligned}$$

Remains to show that Q_1 and Q_2 are independent

Q_1 is a function of $(\mathbf{P} - \mathbf{P}_1)\mathbf{u}$, while Q_2 is a function of $\mathbf{Mu}$

Since both normal, suffices to show that covariance is zero:

$$\begin{aligned}
\operatorname{cov}[(\mathbf{P} - \mathbf{P}_1)\mathbf{u}, \mathbf{Mu} \,|\, \mathbf{X}] &= \mathbb{E}\left[(\mathbf{P} - \mathbf{P}_1)\mathbf{u}(\mathbf{Mu})' \,|\, \mathbf{X}\right] \\
&= \mathbb{E}\left[(\mathbf{P} - \mathbf{P}_1)\mathbf{uu}'\mathbf{M} \,|\, \mathbf{X}\right] \\
&= (\mathbf{P} - \mathbf{P}_1)\mathbb{E}\left[\mathbf{uu}' \,|\, \mathbf{X}\right]\mathbf{M} \\
&= \sigma^2(\mathbf{P} - \mathbf{P}_1)\mathbf{M} = \sigma^2(\mathbf{P} - \mathbf{P}_1)(\mathbf{I} - \mathbf{P})
\end{aligned}$$

Covariance is zero, because

$$(\mathbf{P} - \mathbf{P}_1)(\mathbf{I} - \mathbf{P}) = \mathbf{P} - \mathbf{P}^2 - \mathbf{P}_1 + \mathbf{P}_1\mathbf{P} = \mathbf{P} - \mathbf{P} - \mathbf{P}_1 + \mathbf{P}_1 = \mathbf{0}$$

Most common implementation of the F test:

$$\mathbf{y} = \mathbf{1}\beta_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u} \quad \text{and null is} \quad \boldsymbol{\beta}_2 = \mathbf{0}$$

Exercise: Show that in this case we have

$$\mathrm{RSSR} = \sum_{n=1}^{N} (y_n - \bar{y})^2$$

Hint: Show first that

$$\mathbf{M}_1 = \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}'$$

Bias: Failure of Exogeneity

Recall:

- $\mathbb{E}[\mathbf{u} \mid \mathbf{X}] = \mathbf{0}$ called the "exogeneity" assumption
- Exogeneity + linear model yields $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$

Conversely, if exogeneity fails, then OLS estimator can be biased

When might exogeneity fail?

Let's look at an example

Suppose our data generated according to AR(1) model

$$y_0 = 0 \quad \text{and} \quad y_n = \beta y_{n-1} + u_n \quad \text{for} \quad n = 1, \ldots, N \tag{2}$$

Assume: $\beta \neq 0$ and $\{u_n\}$ is IID with $\mathbb{E}[u_n] = 0$ and $\text{var}[u_n] = \sigma^2$

Letting

$$\mathbf{y} := \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \quad \text{and} \quad \mathbf{x} := \begin{pmatrix} y_0 \\ \vdots \\ y_{N-1} \end{pmatrix}$$

we can express our model as

$$\mathbf{y} = \beta \mathbf{x} + \mathbf{u}$$

The OLS estimator of β is

$$\hat{\beta} := (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}$$

Is it unbiased?

Exogeneity assumption requires $\mathbb{E}\left[\mathbf{u} \mid \mathbf{x}\right] = \mathbf{0}$

One implication of exogeneity is

$$\mathbb{E}\left[u_m x_{n+1}\right] = 0 \quad \text{for any} \quad n, m$$

Therefore

$$\text{exogeneity} \implies \mathbb{E}\left[u_m y_n\right] = 0 \quad \text{for any} \quad n, m$$

Claim: $\mathbb{E}[u_m y_n] = 0$ fails in our model whenever $n \geq m$

Intuition:

$$y_t = \beta y_{t-1} + u_t \quad \text{for} \quad t = 1, \ldots, N$$

Current shock u_t affects current state y_t

Current state affects future state

This means current shock and future state are correlated

This means $\mathrm{cov}[u_m, y_n] = \mathbb{E}[u_m y_n]$ is nonzero when $n \geq m$

Proof that $\mathbb{E}[u_m y_n] = 0$ fails in our model whenever $n \geq m$:

It's an exercise to show that y_n can be expressed as

$$y_n = \beta^{n-1} u_1 + \beta^{n-2} u_2 + \cdots \beta^0 u_n = \sum_{j=0}^{n-1} \beta^j u_{n-j}$$

$$\therefore \quad \mathbb{E}[y_n u_m] = \sum_{j=0}^{n-1} \beta^j \mathbb{E}[u_{n-j} u_m] = \beta^{n-m} \sigma^2 \quad \text{whenever } n \geq m$$

$\therefore$ exogeneity assumption fails

As a result, $\hat{\beta}$ may be biased

It turns out that $\hat{\beta}$ is biased downwards when $\beta \in (0,1)$

We can illustrate this by simulation:

1. fix $\beta \in (0,1)$ and take $\{u_n\} \stackrel{\text{IID}}{\sim} \mathcal{N}(0,1)$
2. generate data M times
3. compute OLS estimate $\hat{\beta}$ on each occasion
4. take sample mean $\frac{1}{M}\sum_{m=1}^{M} \hat{\beta}_m \approx \mathbb{E}[\hat{\beta}]$

We find that

$$\frac{1}{M}\sum_{m=1}^{M} \hat{\beta}_m < \beta$$

```
N <- 20
y <- numeric(N)
y_zero <- 0
beta <- 0.9
num_reps <- 10000
betahat_obs <- numeric(num_reps)

for (j in 1:num_reps) {
    u <- rnorm(N)
    y[1] <- beta * y_zero + u[1]
    for (t in 1:(N-1)) {
        y[t+1] <- beta * y[t] + u[t+1]
    }
    x <- c(y_zero, y[-N])  # Lagged y
    betahat_obs[j]  <- sum(x * y) / sum(x^2)
}
print(mean(betahat_obs))
```

Histogram:

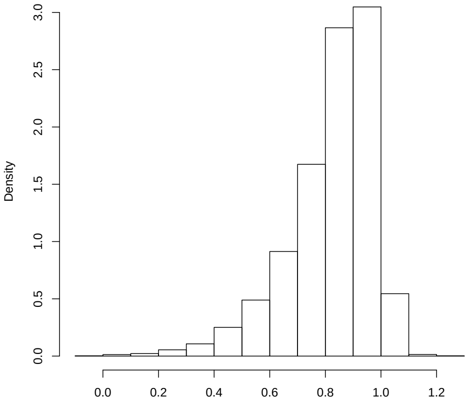


Figure: Observations of $\hat{\beta}$

In the simulation:

- $\beta = 0.9$
- resulting sample mean was 0.82

Asymptotic 95% confidence interval for $\mathbb{E}\left[\hat{\beta}\right]$ was (0.818, 0.824)

Note:

- Althought $\hat{\beta}$ is biased under our assumptions
- it is in fact consistent...

We'll talk about this next week