ECON 6200: Section 2

1/31/25

Interpreting the Linear Model

1) Best Linear Predictor
   - makes sense in very general cases
   - almost always interpreted as a project coefficient under square loss

2) Causal Linear Model
   - underlying "true" model is linear
   - more restrictive, but gives you a causal interpretation

Note: All finite sample theory results are based on the causal interpretation of OLS

Assumptions (Causal Model)

(1) Linearity $\quad Y = X\beta + \varepsilon$

(2) Strong Exogeneity $\quad E(\varepsilon | X) = 0$

↳ conditioning on <u>all</u> elements in X

<u>ex</u>: $E[\varepsilon_i | X_j] = 0, \ \forall i \neq j$

(3) Rank Condition $\quad \text{rank}(X) = K$ a.s.

where $X \in \mathbb{R}^{n \times k}$

↳ ie: $X'X$ nonsingular

↳ Identification assumption

If fails, then there exists a set of observationally equivalent "true" coefficient that form a linear subspace of $\mathbb{R}^k$ and all induce the same $\hat{Y}$

(4) Spherical Errors $\quad E[\varepsilon\varepsilon' | X] = \sigma^2 I_n$

↳ combines homoscedasticity $E[\varepsilon_i^2 | X] = \sigma^2$ and no correlation between errors

$E[\varepsilon_i \varepsilon_j | X] = 0$, for $i \neq j$

Note: Identification (in simple terms) means that a parameter is uniquely determined by the distribution of the observed variables

Finite Sample Theory

Theorem

Under the assumptions (1) - (4), we have:

A) $\hat{\beta}$ is unbiased
↳ $E(\hat{\beta} | X) = \beta$

B) $var(\hat{\beta} | X) = \sigma^2 (X'X)^{-1}$

C) Gauss - Markov Thm
↳ $\hat{\beta}$ is BLUE

A) Prove $\hat{\beta}$ unbiased

Recall

$$\begin{aligned} \hat{\beta} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\beta + \varepsilon) \\ &= \beta + \underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon}_{=\text{estimation error}} . \end{aligned}$$

Thus,

$$E[\hat{\beta} \mid X] = \beta + E[(X'X)^{-1} X'\varepsilon \mid X]$$

$$= \beta + (X'X)^{-1} X' \underbrace{E[\varepsilon \mid X]}_{=0 \text{ by } ②}$$

$$= \beta \quad \checkmark$$

B) Prove $var(\hat{\beta} \mid X) = \sigma^2 (X'X)^{-1}$

$$var(\hat{\beta} \mid X) = var(\underset{\text{constant}}{\beta} + (X'X)^{-1} X'\varepsilon \mid X)$$

$$= var((X'X)^{-1} X'\varepsilon \mid X)$$

$$= (X'X)^{-1} X' \, var(\varepsilon \mid X) \, ((X'X)^{-1} X')'$$

$$= (X'X)^{-1} X' \sigma^2 I_n X (X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1} \cancel{X'X} \cancel{(X'X)^{-1}}$$

$$= \sigma^2 (X'X)^{-1} \quad \checkmark$$

$var(AY) = A \, var(Y) A'$

② $E[\varepsilon \mid X] = 0$

④ $E[\varepsilon\varepsilon' \mid X] = \sigma^2 I_n$

$Var(Y) = E[Y^2] - E[Y]^2$

c) Gauss Markov Thm

- OLS estimator $\hat{\beta}$ is "BLUE"
  ↳ BLUE = Best Linear Unbiased Estimator
- Variance is the metric we use to assess "Best"
- Obviously, there exists better estimators with a smaller variance, but they may not <u>linear</u> or <u>unbiased</u>

Theorem 4.4 Gauss-Markov
In the homoskedastic linear regression model (Assumption 4.3) with i.i.d. sampling (Assumption 4.1), if $\widetilde{\beta}$ is a linear unbiased estimator of β then

$$\operatorname{var}\left[\widetilde{\beta} \mid \boldsymbol{X}\right] \geq \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.$$

Let $\tilde{\beta}$ be a linear unbiased estimator that is <u>not</u> the OLS estimator. Let us represent $\tilde{\beta}$ as

$$\tilde{\beta} = \beta_{OLS} + DY$$

since $\tilde{\beta}$ is linear in Y

$$= CY$$

where $D = C - (X'X)^{-1}X'$

Note: Intuitively, $\tilde{\beta} = \beta_{OLS} + \text{noise}$

$$\Rightarrow C = (X'X)^{-1}X' + D$$

[WTS: $\tilde{\beta}$ unbiased]

$$
\begin{aligned}
E[\tilde{\beta} \mid X] &= E[CY \mid X] \\
&= E[((X'X)^{-1}X' + D)Y \mid X] \\
&= E[(X'X)^{-1}X'Y \mid X] + E[DY \mid X] \qquad \text{Distribute} \\
&= \beta_{OLS} + E[D(X\beta_{OLS} + \varepsilon) \mid X] \\
&= \beta_{OLS} + E[DX\beta_{OLS} \mid X] + D\underbrace{E[\varepsilon \mid X]}_{=0 \text{ by } 2}
\end{aligned}
$$

Note: D is a func of X

In order for $\tilde{\beta}$ to be unbiased, we must have

$$E[DX\beta_{OLS} \mid X] = 0 \quad \text{for any } \beta$$

$\Rightarrow$ Since β is a constant and D is a function of X, we must have $DX = 0$

[WTS: $var(\tilde{\beta} \mid X) \geq \sigma^2 (X'X)^{-1}$

$$var(\tilde{\beta} \mid X) = var\left(\left((X'X)^{-1}X' + D\right) Y \mid X\right)$$

$$= var\left(\left((X'X)^{-1}X' + D\right)(X\beta + \varepsilon) \mid X\right)$$

everything here is either a scalar or function of X

$$= var\left[\underbrace{\left((X'X)^{-1}X' + D\right)X\beta}_{\text{treat this as a constant when conditioning on } X} + \left((X'X)^{-1}X' + D\right)\varepsilon \mid X\right]$$

$$= var\left[\underbrace{\left((X'X)^{-1}X' + D\right)}_{\text{func. of } X \Rightarrow \text{treat as constant}} \varepsilon \mid X\right]$$

$$= \left((X'X)^{-1}X' + D\right) var(\varepsilon \mid X) \left((X'X)^{-1}X' + D\right)'$$

$$= \underbrace{\sigma^2}_{\text{by } 4} \left[\left((X'X)^{-1}X' + D\right)\left((X'X)^{-1}X' + D\right)'\right]$$

Recall $(A+B)' = A' + B'$

$$= \sigma^2 \left[\left((X'X)^{-1}X' + D\right)\left(X(X'X)^{-1} + D'\right)\right]$$

$$= \sigma^2 \left[(X'X)^{-1}\cancel{X'X(X'X)^{-1}} + \cancel{(X'X)^{-1}}\underbrace{X'D'}_{(DX)'=0} + \underbrace{DX}_{=0}\cancel{(X'X)^{-1}} + DD'\right]$$

$$= \sigma^2 \left[(X'X)^{-1} + \underbrace{DD'}_{\text{PSD (all entries} \geq 0)}\right]$$

$$\geq \sigma^2 (X'X)^{-1} = var(\hat{\beta} \mid X)$$

Note: This proof assumes homoscedasticity (ie: 2 $E[\varepsilon \mid X] = 0$)

Standard Errors

* What are standard errors?
  - estimators of standard deviations
  - quantifies the uncertainty or variability in an estimator

Ex: $SE(\hat{\beta})$ = uncertainty or variability in the estimate of the regression coefficient $\hat{\beta}$

If I repeatedly sampled from the population and re-estimated the regression model, how does my $\hat{\beta}$ change?

Standard Errors vs Standard Deviation

Standard Errors	Standard Deviation
- variability between multiple samples from the same population	- variability inside a sample
- how precise is our estimate?	- how spread out is my data?

⊛ Estimating Variance

To estimate the variance, in class we first considered:

- plug-in estimator

$$\hat{\sigma}^2 \equiv \frac{1}{n}\hat{\varepsilon}'\hat{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2.$$

However, $\hat{\sigma}^2$ is biased

$$E[\hat{\sigma}^2 | X] = \frac{n-k}{n}\sigma^2 \neq \sigma^2$$

So now take this into account and create:

- df - adjusted estimator

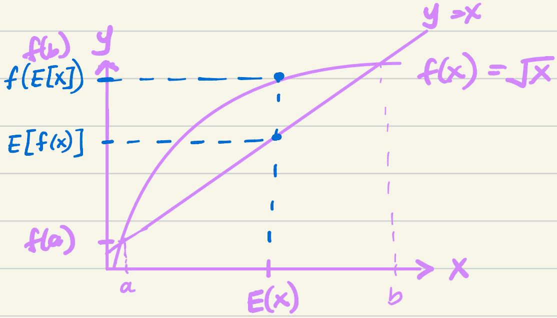
$$s^2 \equiv \frac{1}{n-K}\sum_{i=1}^{n}\hat{\varepsilon}_i^2.$$

where s^2 is unbiased $E[s^2 | X] = \sigma^2$

[Question: If we want to estimate the standard deviation, is $\sqrt{s^2}$ unbiased?]

Ans: No! Expectation is a linear operator. By Jensen's Inequality, we can see that $E[\sqrt{S^2}] \le \underbrace{\sqrt{E[S^2]}}_{\text{standard deviation}}$

Note: Since $\sqrt{(\cdot)}$ is a concave function, by Jensen's Inequality



Previously, we made no distributional assumptions on ε.
However, if we impose a normality assumption on ε, we can derive an <u>exact finite sample distribution</u> and perform hypothesis testing

Let

$$\varepsilon | X \sim N(0, \sigma^2 I_n)$$

then we have

$$(\hat{\beta} - \beta) | X \sim N(0, \sigma^2 (X'X)^{-1})$$

<u>Proof</u>

Previously, we showed that $\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$

$$E[\hat{\beta} - \beta | X] = E[\hat{\beta} | X] - \beta = 0$$

$$\begin{aligned} Var(\hat{\beta} - \beta | X) &= Var(\hat{\beta} | X) \\ &= \sigma^2 (X'X)^{-1} \end{aligned}$$

$$\Rightarrow \hat{\beta} - \beta \sim N(0, \sigma^2 (X'X)^{-1})$$

Problem: MLE vs OLS Under Normality Assumption

$$Y = X\beta + \varepsilon$$

By OLS, our objective function is

$$\hat{\beta}_{OLS} = \arg\min_{b} (y - Xb)'(y - Xb)$$

$$\Rightarrow \hat{\beta}_{OLS} = (X'X)^{-1}(XY)$$

assuming $(X'X)^{-1}$ invertible.

Given data $Y_1, \dots, Y_n \mid X$, show that

$$\hat{\beta}_{MLE} = \hat{\beta}_{OLS} \quad \text{when} \quad \varepsilon \sim N(0, \sigma^2 I_n)$$