Econ 210 - Simple Linear Regression[1]

Sidharth Sah

September 8, 2023

[1] Thanks to Azeem Shaikh and Max Tabord-Meehan for useful material

Setting

- Let X, Y, and U be rv's such that:

$$Y = \beta_0 + \beta_1 X + U \tag{1}$$

- Will call Y the *regressand* or *dependant variable*, X the *regressor* or *independant variable*, and U the *error term*
- β_0, β_1 are the parameters - β_0 is the *intercept* and β_1 is the *slope parameter*

Interpretations of SLR

- We can interpret equation (1) in three distinct ways:
  - 1. Linear conditional expectation
  - 2. Best linear predictor/best linear approximation to conditional expectation
  - 3. Causal model

Linear Conditional Expectation Interpretation

- Under the *linear conditional expectation* interpretation, we suppose that:

$$E[Y|X] = \beta_0 + \beta_1 X$$

- **Define** $U = Y - E[Y|X]$
- Then, by construction,

$$Y = \beta_0 + \beta_1 X + U$$

- β_0 and β_1 do not have a causal interpretation - they describe features of the joint distribution of X and Y - specifically the conditional expectation

U Under Linear Conditional Expectation Interp

- By construction,

$$E[U|X] = E[Y - E[Y|X]|X] = E[Y|X] - E[Y|X] = 0$$

- This further implies:

$$\begin{aligned} E[U] &= 0 \\ Cov(X, U) = E[XU] &= 0 \end{aligned}$$

Best Linear Approximation to Conditional Expectation Interpretation

- Very often, the conditional expectation of Y given X will not be exactly linear
- However, can choose β_0, β_1 in order to form "best" linear approximation to any conditional expectation
- That is, we now say that β_0, β_1 are the values of b_0, b_1 that satisfy the following:

$$\min_{b_0, b_1} E[(E[Y|X] - (b_0 + b_1 X))^2] \tag{2}$$

- Thus, our parameters minimize the "distance" between the conditional expectation and $\beta_0 + \beta_1 X$

Best Linear Predictor Interpretation

- Instead of defining our parameters to minimize the "distance" between the *conditional expectation* of Y and $\beta_0 + \beta_1 X$, we can also just minimize the "distance" between Y itself and $\beta_0 + \beta_1 X$
- That is, we now say that β_0, β_1 are the values of b_0, b_1 that satisfy the following:

$$\min_{b_0, b_1} E[(Y - (b_0 + b_1 X))^2] \tag{3}$$

Equivalence of Best Linear Approximation to Conditional Expectation and Best Linear Predictor

- Turns out the solutions to (2) and (3) are identical:

$$
\begin{aligned}
E[(Y - b_0 - b_1 X)^2] &= E[(\underbrace{(Y - E[Y|X])}_{=V} \\
&\quad + (E[Y|X] - b_0 - b_1 X))^2] \\
&= E[V^2] + 2E[VE[Y|X]] \\
&\quad - 2b_0 E[V] - 2b_1 E[VX] \\
&\quad + E[(E[Y|X] - b_0 - b_1 X)^2] \\
&= E[V^2] + 2E[VE[Y|X]] \\
&\quad + E[(E[Y|X] - b_0 - b_1 X)^2] \quad \text{(LIE)}
\end{aligned}
$$

Equivalence of Best Linear Approximation to Conditional Expectation and Best Linear Predictor

- Imagine picking $b_0 = \beta_0$ and $b_1 = \beta_1$ to minimize $E[(Y - b_0 - b_1X)^2]$
- We see above that this same choice will minimize $E[V^2] + 2E[VE[Y|X]] + E[(E[Y|X] - b_0 - b_1X)^2]$
- However, the first two terms do not depend on b_0, b_1 at all, so this is then further equivalent to minimizing $E[(E[Y|X] - b_0 - b_1X)^2]$
- Thus, β_0, β_1 will minimize both problems

Best Linear Approximation to Conditional Expectation/Best Linear Predictor Interpretation

- **Define** $U = Y - \beta_0 - \beta_1 X$, so, by construction:

$$Y = \beta_0 + \beta_1 + U$$

- Again, β_0, β_1 do not have a causal interpretation - they describe a feature of the joint distribution of X and Y - either the best linear approximation to conditional expectation or the best linear predictor

U Under Best Linear Approximation to Conditional Expectation/Best Linear Predictor Interp

- Taking first-order conditions of minimization problem (3) suggests:

$$E[Y - b_0 - b_1 X] = 0$$
$$E[X(Y - b_0 - b_1 X)] = 0$$

- Plug in the definition of U we are using for this interpretation to see

$$E[U] = E[XU] = 0$$

- We have made almost no assumptions, unlike the 1st interpretation, but we can no longer say that $E[U|X] = 0$

Causal Model Interpretation

- Under the *causal model* interpretation, we suppose that:

$$Y = g(X, U)$$

where

$$X = \text{an observed determinant of } Y$$
$$U = \text{unobserved determinants of } Y$$

- Now we have a model saying that Y is causally determined by X and U
- Could have Y as wages, X as years of schooling, and U as all other determinants of wage (socio-economic background, intelligence, determination, etc)

Causal Model Interpretation

- The causal effect of X on Y, holding U constant, is given by:

$$\frac{dY}{dX} = \frac{dg(X,U)}{dX}$$

- If we assume that

$$g(X,U) = \beta_0 + \beta_1 X + U$$

then,

$$Y = \beta_0 + \beta_1 X + U$$

by assumption. Moreover, $\frac{dg(X,U)}{dX} = \beta_1$, so our slope parameter has a causal interpretation

U Under Causal Model Interpretation

- Given the linear form of $g(X, U)$, we can always normalize such that $E[U] = 0$:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X + U \\ &= \underbrace{\beta_0 + E[U]}_{=\beta_0'} + \beta_1 X + \underbrace{U - E[U]}_{=U'} \\ &= \beta_0' + \beta_1 X + U' \end{aligned}$$

- Then $E[U'] = 0$ and we have the same β_1 of interest
- But, we *can't* say much about $E[U|X]$ or $E[XU]$. Claiming that either of those is 0 is a *substantive* assumption about the world, not an artifact of how we define U, like in interp. 1 or 2

Requirements to Calculate β_0 and β_1

- We now turn to discussion of how to calculate β_0 and β_1 from moments of X and Y, where X and Y satisfy equation (1). For this, we will maintain the following assumptions:
  (a) $E[U] = 0$
  (b) $E[XU] = 0$
  (c) $0 < \text{Var}[X] < \infty$
- We are agnostic about how we arrive at these assumptions - (a) is "free" under any interpretation, (b) is "free" under interpretation 2

Calculating β_0 and β_1

- Given that $U = Y - \beta_0 - \beta_1 X$ and (a)

$$\begin{aligned} E[Y - \beta_0 - \beta_1 X] &= 0 \\ \Rightarrow \beta_0 &= E[Y] - \beta_1 E[X] \end{aligned}$$

- We can plug this into (b) to see:

$$\begin{aligned} E[X(Y - \beta_0 - \beta_1 X)] &= 0 \\ E[X((Y - E[Y]) - \beta_1(X - E[X]))] &= 0 \\ \Rightarrow \underbrace{E[X(Y - E[Y])]}_{=Cov(X,Y)} &= \beta_1 \underbrace{E[X(X - E[X])]}_{=Var(X)} \end{aligned}$$

Calculating β_0 and β_1

- This implies:

$$\beta_1 = \frac{Cov(X, Y)}{Var(X)}$$

  which is defined, thanks to (c)

- Further,

$$\begin{aligned} \beta_0 &= E[Y] - \beta_1 E[X] \\ &= E[Y] - \frac{Cov(X, Y)}{Var(X)} E[X] \end{aligned}$$

Calculating β_0 and β_1 for Binary X

- These coefficients simplify particularly nicely when X is binary (i.e. $X \in \{0, 1\}$). Letting $p = P(X = 1)$, we observe:

$$
\begin{aligned}
E[XY] &= pE[Y|X = 1] \\
E[X]E[Y] &= p^2 E[Y|X = 1] + p(1 - p)E[Y|X = 0] \\
Var(X) &= p(1 - p)
\end{aligned}
$$

- Plugging into our expression for β_1 yields:

$$\beta_1 = \frac{Cov(Y, X)}{Var(X)} = E[Y|X = 1] - E[Y|X = 0]$$

Calculating β_0 and β_1 for Binary X

- We can similarly show that:

$$\beta_0 = E[Y|X = 0]$$

- This comports with the interpretation of the regression equation as representing the linear conditional expectation of Y given X. Given binary X, the conditional expectation of Y has to be binary, and take on exactly the form:

$$E[Y|X] = E[Y|X = 0] + (E[Y|X = 1] - E[Y|X = 0])X$$

SLR For Experiments

- This might be particularly useful for an experiment in which we randomize participants into some "treatment", represented by binary X:

$$X = \begin{cases} 1 \text{ if treated} \\ 0 \text{ if note treated} \end{cases}$$

- We can think of an outcome, Y in terms of *potential outcomes*, where Y_1 represents a person's outcome if they're treated and Y_0 represents a person's outcome if they're not treated

SLR For Experiments

- The *treatment effect* for any one person is given by $Y_1 - Y_0$, and we might be particularly interested in the *Average Treatment Effect (ATE)* $E[Y_1 - Y_0]$
- However (due to the fundamental problem of causal inference), we only every observe one potential outcome per person:

$$Y = Y_1 X + Y_0(1 - X)$$

- Can deal with this by assigning X at random. Then, might assume that potential outcomes are independent of X:

$$(Y_0, Y_1) \perp X$$

SLR For Experiments

- Then,

$$E[Y|X=1] - E[Y|X=0] = E[Y_1|X=1] - E[Y_0|X=0]$$
$$= E[Y_1 - Y_0]$$

- Thus, if we regress

$$Y = \beta_0 + \beta_1 X + U$$

we will recover our ATE with the β_1 parameter

Requirements to Estimate β_0 and β_1

- We now discuss how to estimate β_0 and β_1 from finite samples. For this, we will continue to maintain the assumptions:
  (a) $E[U] = 0$
  (b) $E[XU] = 0$
  (c) $0 < \text{Var}[X] < \infty$
- We also add that $(X_1, Y_1), \dots, (X_n, Y_n)$ are iid $\sim (X, Y)$, where X and Y satisfy equation (1)

Estimators Used to Estimate Components of β_0 and β_1

- We have shown that β_1 and β_0 can be calculated using $E[X]$, $E[Y]$, $Var(X)$, and $Cov(X, Y)$
- For the first three, we already know to estimate using $\overline{X}_n$, $\overline{Y}_n$, and $\hat{\sigma}^2_{X,n}$
- We'll use the *sample covariance*:

$$\hat{\sigma}_{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \overline{X}_n \overline{Y}_n$$

as an estimator for the covariance. This is a consistent estimator (see PSet 2)

Estimating β_0 and β_1

- Natural estimators of β_0 and β_1 are:

$$\hat{\beta}_1 = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}^2_{X,n}}$$

$$\hat{\beta}_0 = \overline{Y}_n - \hat{\beta}_1 \overline{X}_n$$

- These are called the Ordinary Least Squares (OLS) estimators of β_0 and β_1

Ordinary Least Squares

- Remember that β_0 and β_1 solve:

$$\min_{b_0, b_1} E[(Y - (b_0 + b_1 X))^2]$$

- We can show that $\hat{\beta}_0$ and $\hat{\beta}_1$ satisfy:

$$\min_{b_0, b_1} \frac{1}{n} \sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_i))^2]$$

- They, thus, also satisfy the FOC's:

$$\frac{1}{n} \sum_{i=1}^{n} Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i = 0$$

$$\frac{1}{n} \sum_{i=1}^{n} X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

Residuals

- We call:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

  the fitted or predicted values. The amounts that these are "off" by are called the Residuals:

$$\hat{U}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

- By implication of the previous FOC's,

$$\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i \hat{U}_i = 0$$

R^2

- The $\underline{R^2}$ of a regression is a measure of how well the estimated regression parameters fit the data:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS} \qquad (*)$$

where

$$TSS = \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2 = n\hat{\sigma}^2_{Y,n}$$

$$ESS = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y}_n)^2$$

$$SSR = \sum_{i=1}^{n} \hat{U}_i^2$$

R^2

- The second equality in $(*)$ follows if $TSS = ESS + SSR$:

$$\begin{aligned} TSS &= \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2 \\ &= \sum_{i=1}^{n} ((Y_i - \hat{Y}_i) + (\hat{Y}_i - \overline{Y}_n))^2 \\ &= SSR + ESS + 2\sum_{i=1}^{n} \hat{U}_i(\hat{Y}_i - \overline{Y}_n) \\ &= SSR + ESS \end{aligned}$$

- The last line follows using the fact that

$$\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i = \frac{1}{n}\sum_{i=1}^{n} X_i \hat{U}_i = 0$$

R^2

- Thus, it is the case that $0 \leq R^2 \leq 1$, with:

$$R^2 = 1 \Leftrightarrow SSR = 0 \Leftrightarrow \hat{Y}_i = Y_i \ \forall i$$
$$R^2 = 0 \Leftrightarrow ESS = 0 \Leftrightarrow \hat{\beta}_1 = 0$$

- Keep in mind that the R^2 is a descriptive measure of goodness-of-fit. High (low) R^2 does not help (hurt) support a causal interpretation of linear regression

Properties of OLS Estimators

- For all following discussion, we continue to maintain the assumptions that:
  (a) $E[U] = 0$
  (b) $E[XU] = 0$
  (c) $0 < \text{Var}[X] < \infty$

  and that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are iid $\sim (X, Y)$, where X and Y satisfy:
  $$Y = \beta_0 + \beta_1 X + U$$
- For each individual property under consideration, we may then add in additional assumptions

Unbiasedness of OLS Estimators

- Along with maintained assumptions, assume that $E[U|X] = 0$. Then, the OLS estimators are unbiased:

$$E[\hat{\beta}_0] = \beta_0$$
$$E[\hat{\beta}_1] = \beta_1$$

- We will now show the second of these statements to be true

Unbiasedness of OLS Estimators

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)Y_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)X_i} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)(\beta_0 + \beta_1 X_i + U_i)}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)X_i} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)(\beta_1 X_i + U_i)}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)X_i} \\
&= \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)U_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)X_i}
\end{aligned}
$$

Unbiasedness of OLS Estimators

$$
\begin{aligned}
E[\hat{\beta}_1|X_1,...,X_n] &= \beta_1 + E[\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)U_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)X_i}|X_1,...,X_n] \\
&= \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)E[U_i|X_1,...,X_n]}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)X_i} \\
&= \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)E[U_i|X_i]}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)X_i} \\
&\qquad ((Y_i, X_i) \perp (Y_j, X_j),\ i \neq j) \\
&= \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)E[U|X]}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)X_i} \\
&\qquad ((X_i, Y_i) \sim (X, Y)) \\
&= \beta_1
\end{aligned}
$$

Unbiasedness of OLS Estimators

- Finally, applying the law of iterated expectations suggests:

$$
\begin{aligned}
E[\hat{\beta}_1] &= E[E[\hat{\beta}_1 | X_1, ..., X_n]] \\
&= E[\beta_1] \\
&= \beta_1
\end{aligned}
$$

- Showing that $\hat{\beta}_0$ is unbiased is straightforward given the unbiasedness of $\hat{\beta}_1$

Consistency of OLS Estimators

- Along with the maintained assumptions, assume that $E[Y^2], E[X^4] < \infty$. Then, the OLS estimators are consistent,

$$\hat{\beta}_0 \xrightarrow{p} \beta_0$$
$$\hat{\beta}_1 \xrightarrow{p} \beta_1$$

- We will show the second result to be true

Consistency of OLS Estimators

- Recall that, under our assumptions, we know that $\hat{\sigma}^2_{X,n} \xrightarrow{p} \sigma^2_X$ and $\hat{\sigma}_{X,Y} \xrightarrow{p} \sigma_{X,Y}$
- Given that $\sigma^2_X > 0$ by assumption, we can apply the CMT to say:

$$\hat{\beta}_1 = \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}^2_{X,n}} \xrightarrow{p} \frac{\sigma_{X,Y}}{\sigma^2_X} = \beta_1$$

- It is then straightforward to show that $\hat{\beta}_0$ is also consistent

OLS if $E[XU] \neq 0$

- We can notice that $\hat{\beta}_1 \xrightarrow{p} \frac{\sigma_{X,Y}}{\sigma_X^2}$ even if $E[XU] \neq 0$ - that assumption only comes into play to show that $\frac{\sigma_{X,Y}}{\sigma_X^2} = \beta_1$
- Assume that $E[XU] \neq 0$ - what would our slope parameter estimator converge to? (By nature of the question, we are considering the causal model interpretation)

OLS if $E[XU] \neq 0$

- To see what we'll converge to, note that

$$\begin{aligned} \sigma_{X,Y} &= Cov(X, \beta_0 + \beta_1 X + U) \\ &= \beta_1 Var(X) + Cov(X, U) \end{aligned}$$

- Then, our estimator will converge in probability to:

$$\hat{\beta}_1 = \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}^2_{X,n}} \xrightarrow{p} \frac{\sigma_{X,Y}}{\sigma^2_X} = \beta_1 + \frac{Cov(X, U)}{Var(X)}$$

- Whether our estimate will converge to something that is above or below the causal β_1 depends on the sign of $Cov(X, U)$

Limiting Distribution of OLS Estimators

- Along with the maintained assumptions, assume that $E[Y^4], E[X^4] < \infty$. Then,

$$\sqrt{n}(\hat{\beta}_0 - \beta_0) \xrightarrow{d} N(0, \sigma_0^2)$$
$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, \sigma_1^2)$$

  where

$$\sigma_0^2 = \frac{Var[(1 - \frac{E[X]}{E[X^2]}X)U]}{E[(1 - \frac{E[X]}{E[X^2]}X)^2]^2}$$
$$\sigma_1^2 = \frac{Var[(X - E[X])U]}{Var(X)^2}$$

- We'll focus on the limiting distribution for $\hat{\beta}_1$ for now

Limiting Distribution of OLS Estimators

- During the unbiasedness proof, we showed:

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)U_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)X_i}$$

$$\Rightarrow \sqrt{n}(\hat{\beta}_1 - \beta_1) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \overline{X}_n)U_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)X_i}$$

- We also know that the denominator is equal to the sample variance and that $\hat{\sigma^2}_{X,n} \xrightarrow{p} \sigma_X^2$

Limiting Distribution of OLS Estimators

- Turn to the numerator:

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\overline{X}_n)U_i &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-E[X]+E[X]-\overline{X}_n)U_i \\
&= \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-E[X])U_i}_{=A} + \\
&\quad \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(E[X]-\overline{X}_n)U_i}_{=B}
\end{aligned}
$$

Limiting Distribution of OLS Estimators

- Note that $E[(X - E[X])U] = Cov(X, U) = 0$
- Apply CLT to A to see,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - E[X])U_i \xrightarrow{d} N(0, Var[(X - E[X])U])$$

- Need to verify that $E[(X - E[X])^2 U^2] < \infty$, which follows from $E[Y^4], E[X^4] < \infty$ and a little extra work that we'll skip

Limiting Distribution of OLS Estimators

- For B, we can see

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(E[X]-\overline{X}_n)U_i = \underbrace{(E[X]-\overline{X}_n)}_{\xrightarrow{p} 0} * \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i}_{\xrightarrow{d} N(0, Var(U))}$$

- Apply Slutsky's Lemma to see

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(E[X]-\overline{X}_n)U_i \xrightarrow{d} 0 * N(0, Var(U)) = 0$$

- $X \xrightarrow{d} c \Leftrightarrow X \xrightarrow{p} c$ for scalar c, so:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(E[X]-\overline{X}_n)U_i \xrightarrow{p} 0$$

Limiting Distribution of OLS Estimators

- Can use Slutsky again to see the numerator converges:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \overline{X}_n)U_i = A + B \xrightarrow{d} N(0, Var[(X - E[X])U])$$

- Use what we know about the denominator and Slutsky again to say:

$$\begin{aligned}\sqrt{n}(\hat{\beta}_1 - \beta_1) &\xrightarrow{d} \frac{1}{Var(X)}N(0, Var[(X - E[X])U]) \\ &\xrightarrow{d} N(0, \frac{Var[(X - E[X])U]}{Var(X)^2})\end{aligned}$$

Inference on OLS Estimators

- In order to make this result useful, we have to be able to estimate the variance of the limiting distribution $\sigma_1^2 = \frac{Var[(X - E[X])U]}{Var(X)^2}$
- How we estimate this will depend on what we're willing to assume about the conditional variance of the error term, U

Homoskedasticity vs Heteroskedasticity

- We say that U is Homoskedastic if $E[U|X] = 0$ and $Var(U|X) = Var(U)$. If not, then we say U is Heteroskedastic
- Under the assumption of homoskedastic U, σ_1^2 reduces to a nice, simple expression:

$$\sigma_1^2 = \frac{Var[(X - E[X])U]}{Var(X)^2}$$

$$= \frac{Var(X)Var(U)}{Var(X)^2}$$

$$= \frac{Var(U)}{Var(X)}$$

Homoskedasticity vs Heteroskedasticity

- However, it's often difficult to motivate the homoskedasticity assumption in real-world contexts. We'll focus instead on estimating σ_1^2 while allowing for heteroskedastic errors
- Under the same assumptions we used to get the limiting distribution, we can show that the *heteroskedasticity-robust variance estimator*:

$$\hat{\sigma}_1^2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \hat{U}_i^2}{(\hat{\sigma}_{X,n}^2)^2}$$

is a consistent estimator for σ_1^2

Inference on OLS Estimators

- Assume further that $\sigma_1^2 > 0$. Then, by Slutsky,

$$\frac{\sqrt{n}}{\hat{\sigma}_1}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, 1)$$

- We'll call

$$SE(\hat{\beta}_1) = \frac{\hat{\sigma}_1}{\sqrt{n}}$$

the Standard Error of the OLS estimator for $\hat{\beta}_1$ - a measure of the "precision" of our estimate

Hypothesis Testing with OLS

- Say we want to test $H_0: \beta_1 = \beta_{1,0}$ against $H_1 \ \beta_1 \neq \beta_{1,0}$. We'll use the test-statistic:

$$T_n = |\frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}|$$

- Using the same arguments as we saw with testing the mean of a distribution, we'll reject, at a given significance level, α, if:

$$T_n > c = \Phi^{-1}(1 - \frac{\alpha}{2})$$

Confidence Intervals for OLS

- Again, using the same arguments as we saw with testing means, we can also construct a confidence interval with a pre-specified probability of including β_1, α as:

$$[\hat{\beta}_1 - \Phi^{-1}(1 - \frac{\alpha}{2})SE(\hat{\beta}_1), \hat{\beta}_1 + \Phi^{-1}(1 - \frac{\alpha}{2})SE(\hat{\beta}_1)]$$