Introductory Econometrics

Lecture 3: The Properties of the OLS

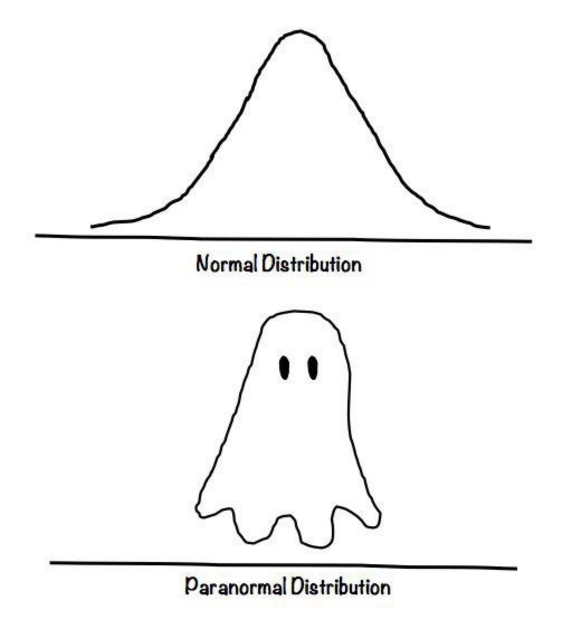
Alex Alekseev

October 31, 2022

University of Regensburg, Department of Economics

Previously on *Introductory Econometrics*...

- Population regression model
- Conditional expectation function
- Linear CEF
- Sample regression model
- Ordinary Least Squares (OLS) estimator


Normal Distribution
Paranormal Distribution

Algebraic Properties of OLS

Sketch of proofs

- Use the definition
- Make an appropriate substitution
- Split the sum
- **Magic happens**
- Get the result

Mean of residuals

- The mean of the residuals is zero **by construction**
- The first **normal** equation says that

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0,$$

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

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

- The mean, as well as the sum of the residuals, is always zero.

Residuals and predictor

- The sample covariance between the predictor and the residuals is zero **by construction**
- The second **normal** equation says that

$$
\begin{aligned}
\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0, \\
\sum_{i=1}^{n} x_i \hat{u}_i = 0 \\
\frac{1}{n}\sum_{i=1}^{n} x_i \hat{u}_i = 0
\end{aligned}
$$

- The covariance (hence, correlation) between the residuals and predictor is always zero

Residuals and fitted values

- The sample covariance between the fitted values and residuals is zero **by construction**.

$$
\begin{aligned}
\widehat{Cov}(\hat{y}_i, \hat{u}_i) &= \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i\hat{u}_i - \left(\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i\hat{u}_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\hat{u}_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_0\hat{u}_i + \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_1 x_i\hat{u}_i \\
&= \frac{1}{n}\hat{\beta}_0\sum_{i=1}^{n}\hat{u}_i + \frac{1}{n}\hat{\beta}_1\sum_{i=1}^{n}x_i\hat{u}_i \\
&= 0
\end{aligned}
$$

Mean of predictor and mean of outcome

- If we plug in the mean of the predictor in the equation for the regression line, we get the mean of the outcome
- In other words, the point $(\bar{x}, \bar{y})$ lies on the regression line.
- Recall the formula for $\hat{\beta}_0$ and re-arrange the terms

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \bar{y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x}\end{aligned}$$

Goodness-of-fit

Sums of squares

- The **total** sum of squares (SST) is

$$SST \equiv \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

- The **explained** sum of squares (SSE) is

$$SSE \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2.$$

- The **residual** sum of squares (SSR) is

$$SSR \equiv \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \hat{u}_2^2.$$

- The following identity holds for these sums:

$$SST = SSE + SSR$$

Proof

- Start with the definition of the SST

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

- Add and subtract $\hat{y}_i$:

$$SST = \sum_{i=1}^{n} (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2.$$

- Notice that $y_i - \hat{y}_i = \hat{u}_i$
- Now compute the square

$$SST = \sum_{i=1}^{n} (\hat{u}_i^2 + (\hat{y}_i - \bar{y})^2 + 2\hat{u}_i(\hat{y}_i - \bar{y})).$$

Proof

- Split the sum into three parts

$$SST = \sum_{i=1}^{n} \hat{u}_i^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} 2\hat{u}_i(\hat{y}_i - \bar{y})).$$

- Notice that the last term can be written as

$$\sum_{i=1}^{n} 2\hat{u}_i(\hat{y}_i - \bar{y})) = 2\left[\sum_{i=1}^{n} \hat{u}_i\hat{y}_i - \bar{y}\sum_{i=1}^{n} \hat{u}_i\right],$$

  which equals to zero, since both $\sum_{i=1}^{n} \hat{u}_i\hat{y}_i = 0$ and $\sum_{i=1}^{n} \hat{u}_i = 0$.

- Therefore, we get that

$$\begin{aligned} SST &= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} \hat{u}_i^2 \\ &= SSE + SSR \end{aligned}$$

RMSE

- The root mean squared error (RMSE) is computed as

$$RMSE \equiv \sqrt{\frac{SSR}{n}}$$

- It shows by how much, on average, the observed values deviate from the fitted values
- It is a measure of the **lack of fit**
- If the fitted values are close to the observed values then RMSE will be small, and we can the model fits the data well
- If the fitted and observed values are far from each other, then the RMSE will be large, and the model does not fit the data well

R-squared

- The **R-squared** is defined as

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST}.$$

- It shows the share of the total variation in the outcome explained by the predictor
- R-squared is bounded between 0 and 1
- Low values of R-squared imply a poor (linear) fit, while high values imply a good (linear) fit

R-squared

- The **R-squared** is defined as

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST}.$$

- It shows the share of the total variation in the outcome explained by the predictor
- R-squared is bounded between 0 and 1
- Low values of R-squared imply a poor (linear) fit, while high values imply a good (linear) fit
- R-squared is related to two correlation coefficients:
  - R-squared is equal to the square of the correlation between the observed and fitted values
  - R-squared is equal to the square of the correlation between the outcome and the predictor (true only for the simple regression)

R-squared

- The **R-squared** is defined as

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST}.$$

- It shows the share of the total variation in the outcome explained by the predictor
- R-squared is bounded between 0 and 1
- Low values of R-squared imply a poor (linear) fit, while high values imply a good (linear) fit

Correlation

Recall that the sample correlation between two random variables X and Y equals

$$\widehat{Corr}(X, Y) = \frac{\widehat{Cov}(X, Y)}{\widehat{sd}(X)\widehat{sd}(Y)},$$

where sd is the standard deviation.

R-squared and the observed and fitted values

- The first identity says that

$$R^2 = (\widehat{Corr}(Y, \hat{Y}))^2$$

- First, consider the sample covariance between the observed and fitted values

$$\widehat{Cov}(Y, \hat{Y}) = \frac{1}{n}\sum_{i=1}^{n} y_i \hat{y}_i - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} \hat{y}_i\right)$$

- Recall that $y_i = \hat{y}_i + \hat{u}_i$ and plug this into the expression for the covariance.

$$\widehat{Cov}(Y, \hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i + \hat{u}_i)\hat{y}_i - \left(\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i + \hat{u}_i)\right)\left(\frac{1}{n}\sum_{i=1}^{n} \hat{y}_i\right).$$

Proof

- Now split the sums as follows

$$\widehat{Cov}(Y,\hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i^2 + \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i\hat{y}_i - \left(\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i + \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i\right)$$

- Recall that $\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i\hat{y}_i = 0$ and $\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i = 0$. The expression becomes

$$\widehat{Cov}(Y,\hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i\right)^2$$

- This expression is the sample variance of $\hat{Y}$
- On the other hand, this variance also equals

$$\widehat{Var}(\hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2,$$

since the sample mean of Y is the same as the sample mean of $\hat{Y}$

Proof

- Therefore, we get that

$$\widehat{Cov}(Y, \hat{Y}) = \frac{1}{n} SSE.$$

- Turning to the standard deviations, we have

$$\widehat{sd}(Y) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2} = \sqrt{\frac{1}{n} SST}$$

$$\widehat{sd}(\hat{Y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2} = \sqrt{\frac{1}{n} SSE}$$

- Then the sample correlation between Y and $\hat{Y}$ is

$$\widehat{Corr}(Y, \hat{Y}) = \frac{\widehat{Cov}(Y, \hat{Y})}{\widehat{sd}(Y)\widehat{sd}(\hat{Y})} = \frac{\frac{1}{n} SSE}{\frac{1}{n}\sqrt{SSE \cdot SST}} = \sqrt{\frac{SSE}{SST}}.$$

- Taking the squares on both sides completes the proof.

R-squared and the outcome and predictor

- The second identity says that

$$R^2 = (\widehat{Corr}(X, Y))^2$$

- To prove it, it helps to first derive a few auxiliary identities
- First, let's show that

$$\widehat{Cov}(X, Y) = \widehat{Cov}(X, \hat{Y})$$

Proof

- Using the definition of the sample covariance, we get

$$\widehat{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right).$$

- Substitute $y_i = \hat{y}_i + \hat{u}_i$.

$$\widehat{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n} x_i(\hat{y}_i + \hat{u}_i) - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i + \hat{u}_i)\right).$$

- Now split the sums as follows

$$\widehat{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n} x_i\hat{y}_i + \frac{1}{n}\sum_{i=1}^{n} x_i\hat{u}_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i + \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i\right)$$

Proof

- Recall that $\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i = 0$ and that $\frac{1}{n}\sum_{i=1}^{n}x_i\hat{u}_i = 0$
- Then

$$\widehat{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}x_i\hat{y}_i - \left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i\right) = \widehat{Cov}(X,\hat{Y}).$$

Proof

- Second, let's show that

$$\widehat{Cov}(Y, \hat{Y}) = \hat{\beta}_1 \widehat{Cov}(X, Y)$$

- Using the definition of the sample covariance, we get

$$\widehat{Cov}(Y, \hat{Y}) = \frac{1}{n}\sum_{i=1}^{n} y_i \hat{y}_i - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} \hat{y}_i\right).$$

- Substitute $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ to get

$$\widehat{Cov}(Y, \hat{Y}) = \frac{1}{n}\sum_{i=1}^{n} y_i(\hat{\beta}_0 + \hat{\beta}_1 x_i) - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right).$$

- Split the sum as follows

$$\widehat{Cov}(Y, \hat{Y}) = \frac{1}{n}\sum_{i=1}^{n} y_i \hat{\beta}_0 + \frac{1}{n}\sum_{i=1}^{n} y_i \hat{\beta}_1 x_i - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_0 + \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_1 x_i\right).$$

Proof

- Simplify to get

$$\widehat{Cov}(Y,\hat{Y}) = \hat{\beta}_0 \frac{1}{n}\sum_{i=1}^{n} y_i + \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} y_i x_i - \hat{\beta}_0 \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{\beta}_1 \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)$$

- Canceling the terms and taking $\hat{\beta}_1$ out of the brackets, we get

$$\widehat{Cov}(Y,\hat{Y}) = \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} y_i x_i - \hat{\beta}_1 \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \hat{\beta}_1 \widehat{Cov}(X,Y)$$

Proof

- Third, let's show that

$$\widehat{Var}(\hat{Y}) = \hat{\beta}_1^2 \widehat{Var}(X)$$

- Using the definition of the sample variance, we get

$$\widehat{Var}(\hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2.$$

- Recall that $\bar{y} = \bar{\hat{y}}_i$
- Substitute $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ to get

$$\widehat{Var}(\hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})^2.$$

- Recall that $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$, which implies that $\hat{\beta}_0 - \bar{y} = -\hat{\beta}_1 \bar{x}$

$$\widehat{Var}(\hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_1 x_i - \hat{\beta}_1 \bar{x})^2 = \hat{\beta}_1^2 \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \hat{\beta}_1^2 \widehat{Var}(X).$$

Proof

- We can now show that

$$\widehat{Corr}(X, \hat{Y}) = 1.$$

- Using the definition of the sample correlation, we get

$$\widehat{Corr}(X, \hat{Y}) = \frac{\widehat{Cov}(X, \hat{Y})}{\widehat{sd}(X)\widehat{sd}(\hat{Y})}.$$

- Using the fact that

$$\hat{\beta}_1 = \frac{\widehat{Cov}(X, Y)}{\widehat{Var}(X)}$$

and that $\widehat{Cov}(X, Y) = \widehat{Cov}(X, \hat{Y})$, we get

$$\widehat{Cov}(X, \hat{Y}) = \hat{\beta}_1 \widehat{Var}(X).$$

Proof

- Plugging this into the formula for the correlation, we get

$$\widehat{Corr}(X, \hat{Y}) = \frac{\hat{\beta}_1 \widehat{Var}(X)}{\widehat{sd}(X)\widehat{sd}(\hat{Y})}.$$

- Now use the fact that $\widehat{Var}(\hat{Y}) = \hat{\beta}_1^2 \widehat{Var}(X)$:

$$\widehat{Corr}(X, \hat{Y}) = \frac{\hat{\beta}_1 \widehat{Var}(X)}{\widehat{sd}(X)\hat{\beta}_1 \widehat{sd}(X)} = 1.$$

Finally...

$$
\begin{aligned}
\widehat{Corr}(X,Y) &= \frac{\widehat{Cov}(X,Y)}{\widehat{sd}(X)\widehat{sd}(Y)} \\
&= \frac{\widehat{Cov}(Y,\hat{Y})}{\widehat{sd}(X)\widehat{sd}(Y)}\frac{1}{\hat{\beta}_1} \\
&= \frac{\widehat{Cov}(Y,\hat{Y})}{\widehat{sd}(X)\widehat{sd}(Y)}\frac{\widehat{Var}(X)}{\widehat{Cov}(X,Y)} \\
&= \frac{\widehat{Cov}(Y,\hat{Y})}{\widehat{sd}(Y)}\frac{\widehat{sd}(X)}{\widehat{Cov}(X,Y)} \\
&= \frac{\widehat{Cov}(Y,\hat{Y})}{\widehat{sd}(Y)\widehat{sd}(\hat{Y})}\frac{\widehat{sd}(X)\widehat{sd}(\hat{Y})}{\widehat{Cov}(X,\hat{Y})} \\
&= \frac{\widehat{Cov}(Y,\hat{Y})}{\widehat{sd}(Y)\widehat{sd}(\hat{Y})}\frac{1}{\widehat{Corr}(X,\hat{Y})} \\
&= \frac{\widehat{Cov}(Y,\hat{Y})}{\widehat{sd}(Y)\widehat{sd}(\hat{Y})} = \sqrt{R^2}
\end{aligned}
$$

Statistical properties of the OLS

Example

https://www.econometrics-with-r.org/SmallSampleDIstReg.html

OLS assumptions

- Let's recall the assumptions we made so far

Assumption: Linear CEF

The conditional expectation function is linear

$$\mathbb{E}[Y \mid X] = \beta_0 + \beta_1 X$$

This assumption implies that the population regression model is

$$Y = \beta_0 + \beta_1 X + u,$$

where $\mathbb{E}[u \mid X] = 0$.

OLS assumptions

- Let's recall the assumptions we made so far

Assumption: Random Sampling

Our sample $(x_i, y_i)_{i=1}^n$ is a random sample from the population, i.e., the observations are pairwise independent and identically distributed.

This assumption means that we can write for each observation i that

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

OLS assumptions

- Let's recall the assumptions we made so far

Assumption: Sample Variation in the Predictor

The sample variance of X is non-zero.

This assumption implies, among other things, that we can compute the OLS estimates.

Unbiasedness

- The estimator is called **unbiased** if it's expectation gives the true value
- In our case, this means that

$$\mathbb{E}[\beta_i] = \beta_i, i = 0, 1.$$

- If one keeps repeatedly drawing new samples and estimating parameters, then the average of all obtained estimates will eventually get close to the population parameters
- Unbiasedness is a property of the sample distribution of the OLS estimators
- It **does not imply** that the population parameters are perfectly estimated for any given sample.

Proof

- To prove unbiasedness, we first introduce some new terms

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \sum_{i=1}^n w_i y_i,$$

where

$$w_i \equiv \frac{x_i - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

- The w_i terms have the following properties:

$$\sum_{i=1}^n w_i = 0$$

$$\sum_{i=1}^n w_i x_i = 1$$

$$\sum_{i=1}^n w_i^2 = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Proof

- Now consider the expectation of $\hat{\beta}_1$ conditional on $x_1, \dots, x_n$:

$$\mathbb{E}[\hat{\beta}_1 \mid x_1, \dots, x_n] = \mathbb{E}\left[\sum_{i=1}^{n} w_i y_i \mid x_1, \dots, x_n\right].$$

- Substituting $y_i = \beta_0 + \beta_1 x_i + u_i$, we get

$$\mathbb{E}[\hat{\beta}_1 \mid x_1, \dots, x_n] = \mathbb{E}\left[\sum_{i=1}^{n} w_i(\beta_0 + \beta_1 x_i + u_i) \mid x_1, \dots, x_n\right].$$

- Now split the sum as follows

$$\mathbb{E}[\hat{\beta}_1 \mid x_1, \dots, x_n] = \mathbb{E}\left[\beta_0 \sum_{i=1}^{n} w_i + \beta_1 \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} w_i u_i \mid x_1, \dots, x_n\right].$$

Proof

- Using the properties of w_i, this expression reduces to

$$\begin{aligned}
\mathbb{E}[\hat{\beta}_1 \mid x_1, \dots, x_n] &= \mathbb{E}\left[\beta_1 + \sum_{i=1}^{n} w_i u_i \mid x_1, \dots, x_n\right] \\
&= \beta_1 + \sum_{i=1}^{n} \mathbb{E}[w_i u_i \mid x_1, \dots, x_n] \\
&= \beta_1 + \sum_{i=1}^{n} w_i \mathbb{E}[u_i \mid x_1, \dots, x_n] \\
&= \beta_1.
\end{aligned}$$

Variance

- First, we derive the variance of the OLS estimator under the assumption of **homoskedasticity**.

Assumption: Homoskedasticity

The variance of each error term u_i is identical and equal to σ^2

Variance

- From the definition

$$Var(\hat{\beta}_1 \mid x_1, \dots, x_n) = Var(\sum_{i=1}^{n} w_i y_i \mid x_1, \dots, x_n).$$

- Since our sample is random, the variance of the sum is the sum of the variances:

$$\begin{aligned} Var(\hat{\beta}_1 \mid x_1, \dots, x_n) &= \sum_{i=1}^{n} Var(w_i y_i \mid x_1, \dots, x_n) \\ &= \sum_{i=1}^{n} w_i^2 Var(\beta_0 + \beta_1 x_i + u_i \mid x_1, \dots, x_n) \\ &= \sum_{i=1}^{n} w_i^2 Var(u_i \mid x_1, \dots, x_n) \\ &= \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}. \end{aligned}$$

Variance

- The variance of $\hat{\beta}_0$ is given by

$$Var(\hat{\beta}_0 \mid x_1, \ldots, x_n) = \frac{\sigma^2 \frac{1}{n} \sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

- Consider again the formula for $Var(\hat{\beta}_1 \mid x_1, \ldots, x_n)$
- Notice that $\sum_{i=1}^n (x_i - \bar{x})^2 = n\widehat{Var}(X)$
- Therefore, the variance of $\hat{\beta}_1$ increases in σ^2 and decreases in the number of observations and the variance of X.

Error variance

- The formula for the variance of the OLS estimator involves a term σ^2 that we do not observe
- Hence, we need to replace it with a corresponding sample moment
- Recall that
$$\sigma^2 = Var(u) = \mathbb{E}[u^2].$$
- The estimate of σ^2 is given by
$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2 = \frac{SSR}{n-2}.$$
- This estimate is an unbiased estimate of σ^2

Error variance

- The formula for the variance of the OLS estimator involves a term σ^2 that we do not observe
- Hence, we need to replace it with a corresponding sample moment
- Recall that

$$\sigma^2 = Var(u) = \mathbb{E}[u^2].$$

- The estimate of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{u}_i^2 = \frac{SSR}{n-2}.$$

- This estimate is an unbiased estimate of σ^2

Note

The $n-2$ in the denominator accounts for the fact that we lose two *degrees of freedom* when estimating the residuals. These two degrees of freedom are the moment restrictions we impose (the normal equations).

Error variance

- The square root of $\hat{\sigma}^2$, $\hat{\sigma}$, is called the **standard error** of the regression (or sometimes the root mean squared error).
- Using this estimate, we can find the sample variance and standard deviation of the OLS estimator

$$Var(\hat{\beta}_1 \mid x_1, \ldots, x_n) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

- The square root of this variance, the standard deviation of $\hat{\beta}_1$, is called the **standard error** of $\hat{\beta}_1$.

Robust standard errors

- If we do not assume homoskedasticity, we can arrive at a different estimate for the variance
- **Heteroskedasticity** means that each error term has its own variance σ_i^2
- Recall our derivations

$$
\begin{aligned}
Var(\hat{\beta}_1 \mid x_1, \ldots, x_n) &= \sum_{i=1}^{n} w_i^2 Var(u_i \mid x_1, \ldots, x_n) \\
&= \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sigma_i^2}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}
\end{aligned}
$$

- Replacing σ_i^2 with its estimate $\hat{u}_i^2$ allows us to estimate this variance and hence the standard errors in the sample
- Standard errors derived this way are called **robust**.

Next Time on *Introductory Econometrics*...

Multiple regression