Linear Regression with One Regressor

Osman DOĞAN

Outline: Linear Regression with One Regressor

- Linear regression with one regressor:
  1. The linear regression with one regressor (SLR),
  2. Estimation: the ordinary least squares (OLS) estimator,
  3. Measures of fit,
  4. The least squares assumptions,
  5. The sampling distribution of the OLS estimator.
- Readings:
  1. Stock and Watson (2020, Chapter 4),
  2. Hanck et al. (2021, Chapter 4).

The SLR model

- Suppose you are interested in the relationship between a random variable Y (outcome variable) and another random variable X (explanatory variable).
- The true process that generates Y is unknown to the researcher.
- What the researcher can at least try is to model an aspect of the true process that generates Y.
- For example, the researcher could inquire on average what value of Y would be observed given a specific value of X.
- This is simply the conditional mean of Y given X, i.e., $\mathrm{E}(Y|X)$.

The SLR model

- Next, we will assume a relationship between $\mathrm{E}(Y|X)$ and X.
- Specifically, we will assume that $\mathrm{E}(Y|X)$ is a linear function of X such that

$$\mathrm{E}(Y|X) = \beta_0 + \beta_1 X. \tag{1}$$

- The equation (1) is called the population regression line or the population regression function:
  1. β_0 is called the intercept parameter and β_1 is called the slope parameter.
  2. β_0 gives average Y when X is zero, and β_1 gives the average effect of a unit change in X on Y.

The SLR model

- For example, we may be interested in the effect of class size reduction on test scores:

$$\mathrm{E}(TestScore|ClassSize) = \beta_0 + \beta_1 \times ClassSize. \tag{2}$$

- This equation tells you what the test score will be, on average, for districts with class sizes of a certain value.
- It does not tell you what specifically the test score will be in any one district.
- Districts with the same class sizes can nevertheless differ in many ways and in general will have different values of test scores.
- If we use this equation to make a prediction for a given district, we know that prediction will not be exactly right: The prediction will have an error.

The SLR model

- The difference between Y and $\mathrm{E}(Y|X)$ is called the prediction error, denoted by u.
- Then, from $Y - \mathrm{E}(Y|X) = u$, we obtain

$$Y = \beta_0 + \beta_1 X + u. \tag{3}$$

- Suppose we draw a random sample of n observations on (Y, X) from the population.
- Then, we can write (3) for the ith observation as

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \ldots, n, \tag{4}$$

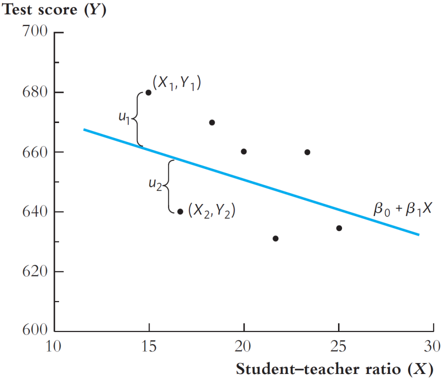
1. Y_i is the outcome (or dependent) variable,
2. X_i is the explanatory (or independent) variable,
3. β_0 and β_1 are the coefficients (or parameters) of the model,
4. u_i is the error (or disturbance) term.

The SLR model

- Figure 4.1 shows the scatter plot of Y and X, and the line plot of $\mathrm{E}(Y_i|X_i) = \beta_0 + \beta_1 X_i$. Since $Y_i - \mathrm{E}(Y_i|X_i) = u_i$, the vertical distance from the point and the line is the error term u_i.

FIGURE 4.1 Scatterplot of Test Score vs. Student–Teacher Ratio (Hypothetical Data)

The scatterplot shows hypothetical observations for seven school districts. The population regression line is $\beta_0 + \beta_1 X$. The vertical distance from the i^{th} point to the population regression line is $Y_i - (\beta_0 + \beta_1 X_i)$, which is the population error term u_i for the i^{th} observation.



Estimating the coefficients of the SLR model

- Given a sample of data set on (Y, X), how can we estimate β_0 and β_1?
- Throughout this semester, we will consider the CA school sample data contained in the `caschool.xlsx` file.
- Y is the average test score (in the district) denoted by `testscr` and X is the average student-teacher ratio (in the district) denoted by `str`.
- The following code can be used to obtain descriptive statistics of data.

```
library(stargazer)
library(readxl)

CAschool = read_excel("caschool.xlsx", col_names = TRUE, skip = 0)
stargazer(data.frame(CAschool), type = "text",
          keep = c("str", "testscr"), align = TRUE)


==============================================
Statistic  N    Mean    St. Dev.   Min      Max
----------------------------------------------
testscr   420 654.157  19.053  605.550 706.750
str       420 19.640   1.892   14.000  25.800
----------------------------------------------
```

Estimating the coefficients of the SLR model

- A scatterplot of these 420 observations on test scores and student-teacher ratios is given in Figure 1.
- The sample correlation is -0.23, indicating a weak negative relationship between the two variables.
- Although larger classes in this sample tend to have lower test scores, there are other determinants of test scores that keep the observations from falling perfectly along a straight line.

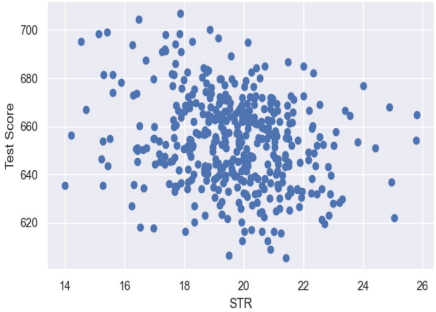


Figure 1: Scatter plot of tests score and the student-teacher ratio

Estimating the coefficients of the SLR model

- How shall we fit a straight line through these observations?
- Let b_0 and b_1 be some chosen values for β_0 and β_1.
- The predicted value of the outcome variable for the ith observation from the regression line is $\widehat{Y}_i = b_0 + b_1 X_i$.
- The prediction error is $Y_i - \widehat{Y}_i = Y_i - b_0 - b_1 X_i$.
- Squaring these errors and summing over the observations we have

$$\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2 .$$

- The ordinary least squares (OLS) estimator is simply the values of b_0 and b_1 which minimize this quantity:

$$\min_{b_0, b_1} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2 . \quad (5)$$

Estimating the coefficients of the SLR model

- Closed form expressions for the solution for this minimization problem are available (see the hand-out for details):

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{s_{XY}}{s_X^2} \tag{6}$$

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{X} \tag{7}$$

where $\bar{Y}$ and $\bar{X}$ are the sample average for Y and X, respectively.

- The OLS predicted values $\widehat{Y}_i$ and residuals $\widehat{u}_i$ are

$$\text{Predicted values}: \widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i, \quad \text{for} \quad i = 1, 2, \ldots, n \tag{8}$$

$$\text{Residuals}: \widehat{u}_i = Y_i - \widehat{Y}_i, \quad \text{for} \quad i = 1, 2, \ldots, n. \tag{9}$$

Estimating the coefficients of the SLR model

- In R, we can use the `lm` function to estimate regression models.
- The basic syntax is `lm(y ~ x, data)`, where
  1. `y` is the dependent variable,
  2. `x` the explanatory variable, and
  3. `data` is the data frame that contains `y` and `x`.
- The `names` function can be used to see which quantities are produced by the `lm` function. We can use the `$` sign to access these quantities.

```
results = lm(testscr ~ str, data = CAschool)
names(results)
[1] "coefficients" "residuals"  "effects"     "rank"       "fitted.values" "assign"
[7] "qr"   "df.residual"    "xlevels"     "call"       "terms"          "model"

results$coefficients
(Intercept)         str
 698.932952   -2.279808
```

Estimating the coefficients of the SLR model

- In Python, we use statsmodels module to estimate the linear regression models.

```
import pandas as pd
import statsmodels.api as sm
import statsmodels.formula.api as smf

# Import data
CAschool=pd.read_csv("caschool.csv")

# Specify the model
model1=smf.ols(formula="testscr~str",data=CAschool)
# We need to use .fit() to obtain parameter estimates
result1=model1.fit()
# To view the OLS regression results, we can call the .summary() method
result1.summary()
```

- Alternatively, we can use the sm.OLS function:

```
# Arrange the data
CAschool["const"]=1 # add a constant column to wage_data
y= CAschool.testscr
X=CAschool[["const","str"]]
# Specify the model
model2=sm.OLS(endog=y,exog=X)
# Fit the model and print the results
result2=model2.fit()
result2.summary()
```

Estimating the coefficients of the SLR model

- After the estimation, we can write the estimated regression model as

$$\widehat{TestScore} = 698.93 - 2.28 \times STR \tag{10}$$

- How should we interpret these estimates?

- Districts with one more student per teacher on average (are predicted to) have test scores that are 2.28 points lower, that is,

$$\frac{\Delta TestScore}{\Delta STR} = -2.28.$$

- The intercept (taken literally) means that, according to this estimated line, districts with zero students per teacher would have a (predicted) test score of 698.9.

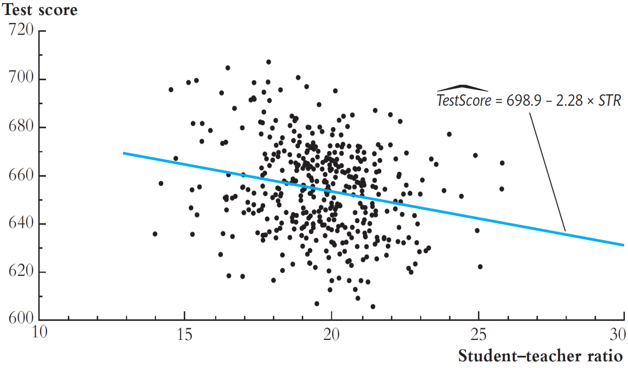
- But this interpretation of the intercept makes no sense, it extrapolates the line outside the range of the data (minimum STR was 14).

The SLR model

- Figure 4.3 shows the scatter plot of Y and X, and the line plot of $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$.

FIGURE 4.3 The Estimated Regression Line for the California Data

The estimated regression line shows a negative relationship between test scores and the student–teacher ratio. For two districts with class sizes that differ by one student per class, the district with the larger class has, on average, test scores that are lower by 2.28 points.



Algebraic properties of the OLS estimator

- There are three results that directly follow from the definition of the OLS estimator (see the hand-out).

    1. The OLS residuals sum to zero: $\sum_{i=1}^{n} \widehat{u}_i = 0$.
    2. The OLS residuals and the explanatory variable are uncorrelated: $\sum_{i=1}^{n} X_i \widehat{u}_i = 0$.
    3. The regression line passes through the sample mean: $\bar{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 \bar{X}$.

- These properties are proved in the handout.

Matrix form

- The model $Y_i = \beta_0 + \beta_1 X_i + u_i$ for $i = 1, \ldots, n$ can be written in matrix form:

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}}_{\boldsymbol{Y}} = \underbrace{\begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}}_{\boldsymbol{X}} \times \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}}_{\boldsymbol{U}} \tag{11}$$

- Thus, the model can be written in the following compact form:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}. \tag{12}$$

- Then, the OLS estimator is defined as

$$\hat{\boldsymbol{\beta}} = \text{argmin}_{\boldsymbol{b}} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})^{'} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}), \tag{13}$$

where $\boldsymbol{b} = (b_0, b_1)^{'}$.

Matrix form

- To find the OLS estimator, we need to set the first order derivative of $(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})^{'}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})$ with respect $\boldsymbol{b}$ to zero.
- Let $SSR = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})^{'}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})$. Note that we can express SER as

$$\begin{aligned} SSR &= \boldsymbol{Y}^{'}\boldsymbol{Y} - \boldsymbol{Y}^{'}\boldsymbol{X}\boldsymbol{b} - \boldsymbol{b}^{'}\boldsymbol{X}^{'}\boldsymbol{Y} + \boldsymbol{b}^{'}\boldsymbol{X}^{'}\boldsymbol{X}\boldsymbol{b} \\ &= \boldsymbol{Y}^{'}\boldsymbol{Y} - 2\boldsymbol{b}^{'}\boldsymbol{X}^{'}\boldsymbol{Y} + \boldsymbol{b}^{'}\boldsymbol{X}^{'}\boldsymbol{X}\boldsymbol{b} \end{aligned} \tag{14}$$

- Then, the OLS estimator is obtained from

$$\begin{aligned} &\frac{\partial SSR}{\partial \boldsymbol{b}} = -2\boldsymbol{X}^{'}\boldsymbol{Y} + 2\boldsymbol{X}^{'}\boldsymbol{X}\boldsymbol{b} = 0 \implies \boldsymbol{X}^{'}\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}^{'}\boldsymbol{Y} \\ &\implies \hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{'}\boldsymbol{X})^{-1}\boldsymbol{X}^{'}\boldsymbol{Y}. \end{aligned}$$

- This formula $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{'}\boldsymbol{X})^{-1}\boldsymbol{X}^{'}\boldsymbol{Y}$ is equivalent to those given in (6) and (7).

Measures of Fit

- After the estimation, you might wonder how well the fitted regression line describes the data.
- Does the regressor account for much or for little of the variation in the dependent variable?
- Are the observations tightly clustered around the regression line, or are they spread out?
- We can try to answer these questions using the R^2 and the standard error of the regression.
- The regression R^2 is the fraction of the sample variance of Y explained by (or predicted by) X.
- The standard error of the regression (SER) is an estimator of the standard deviation of the regression error terms.

Measures of Fit

- The regression R^2 is calculated as

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}\left(\widehat{Y}_i - \bar{\widehat{Y}}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2} = 1 - \frac{\sum_{i=1}^{n}\widehat{u}_i^2}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2} = 1 - \frac{SSR}{TSS},$$

  where

  - ESS stands for the explained sum of squares,
  - TSS stands for the total sum of squares,
  - SSR stands for the sum of squared residuals.

- The R^2 ranges between 0 and 1.

- The R^2 of the regression of Y on the single regressor X is the square of the correlation coefficient between Y and X.

Measures of Fit

- The standard error of the regression (SER) is an estimator of the standard deviation of the regression error terms, u_i's.
- The units of u_i and Y_i are the same, so the SER is a measure of the spread of the observations around the regression line, measured in the units of the outcome variable.
- For example, if the units of the dependent variable are dollars, then the SER measures the magnitude of a typical deviation from the regression line–that is, the magnitude of a typical regression error–in dollars.
- The SER is calculated as

$$SER = s_{\widehat{u}} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\left(\widehat{u}_i - \bar{\widehat{u}}\right)^2} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\widehat{u}_i^2} = \sqrt{\frac{SSR}{n-2}}$$

where the fourth equality is due to $\sum_{i=1}^{n}\widehat{u}_i = 0$ which means $\bar{\widehat{u}} = 0$.

Measures of Fit

- In our CA school example, recall that we saved the regression results in a list named `results`.
- The `summary` function can be used to summarize the OLS estimation results.

```
summary(results)

Call:
lm(formula = testscr ~ str, data = mydata)

Residuals:
    Min      1Q  Median      3Q     Max
-47.727 -14.251   0.483  12.822  48.540

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 698.9330     9.4675  73.825  < 2e-16 ***
str          -2.2798     0.4798  -4.751 2.78e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.58 on 418 degrees of freedom
Multiple R-squared:  0.05124,   Adjusted R-squared:  0.04897
F-statistic: 22.58 on 1 and 418 DF,  p-value: 2.783e-06
```

Measures of Fit

- The estimation results show that the R^2 is about 0.05 and the SER is 18.58.
- The R^2 of 0.05 means that the regressor `STR` explains 5.1% of the variance of the dependent variable `TestScore`.
- The scatter plot in Slide 15 shows that the `STR` explains some of the variation in test scores, but much variation remains unaccounted for.
- The SER of 18.6 means that the standard deviation of the regression residuals is 18.6, where the units are points on the standardized test.
- Because the standard deviation is a measure of spread, the SER of 18.6 means that there is a large spread of the scatterplot in Slide 15 around the regression line as measured in points on the test.
- This large spread means that predictions of test scores made using only the student- teacher ratio for that district will often be wrong by a large amount.

The Least Squares Assumptions for Causal Inference

- The following assumptions are necessary and sufficient to interpret the OLS estimates as causal.

Assumption 1 (Zero-conditional mean assumption)

$\mathrm{E}(u_i|X_i) = 0$, i.e., the conditional distribution of u_i given X_i has a mean of 0.

Assumption 2 (Random sampling assumption)

$(X_i, Y_i) : i = 1, 2, \ldots, n$ are independently and identically distributed (i.i.d.) across observations.

Assumption 3 (No large outliers assumption)

$\mathrm{E}(X_i^4) < \infty$, $\mathrm{E}(Y_i^4) < \infty$, i.e., observations with values of X_i, Y_i, or both, that are far outside the usual range of the data are unlikely.

- In ideal randomized controlled experiments, X is randomly assigned, thus Assumption 1 is plausible.
- With observational data, we will need to think hard about whether $\mathrm{E}(u|X = x) = 0$ holds.

The Least Squares Assumptions for Causal Inference

- How to think about $\mathrm{E}(u_i|X_i) = 0$?
- According to Example 8 in the lecture notes for the second week, the following result holds:

$$\mathrm{E}(u_i|X_i) = 0 \implies \mathrm{corr}(X, u_i) = 0 \tag{15}$$

- The contrapositive statement of (15) is

$$\mathrm{corr}(X, u_i) \neq 0 \implies \mathrm{E}(u_i|X_i) \neq 0 \tag{16}$$

- Thus, if we think that the error term includes a variable that is correlated with the regressor X, then we can claim that $\mathrm{E}(u_i|X_i) \neq 0$.
- For example, in our CA school example, one of the variable that can affect `TestScore` can be the average district income. Our model does not include this variable and thus its effect will show up in the error term.
- Since it is plausible to expect that the average district income will be correlated with `STR`, we can expect that the zero conditional mean assumption does not hold for this example.

The Least Squares Assumptions for Causal Inference

- The least squares assumptions play twin roles.
- Their first role is mathematical: If these assumptions hold, in large samples the OLS estimators are consistent and have sampling distributions that are normal (inference is feasible).
- Their second role is to organize the circumstances that pose difficulties for OLS estimation of the causal effect β_1.
- The first least squares assumption (the zero-conditional mean assumption) is the most important to consider in practice.
- The reasons why the first least squares assumption might not hold in practice are discussed in Chapters 6 and 9.

The Sampling Distribution of the OLS Estimators

- The OLS estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are random variables because they are formulas using the random variables Y_i's and X_i's.

- In other words, they will give different values from one sample to another drawn from the population.

- The sampling distribution of the OLS estimators describes the values they could take over different possible random samples.

- From $Y_i = \beta_0 + \beta_1 X_i + u_i$ and $\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{u}$, we obtain

$$Y_i - \bar{Y} = \beta_1 (X_i - \bar{X}) + (u_i - \bar{u}). \tag{17}$$

- Substituting (17) into (7) yields

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n (X_i - \bar{X}) \left(\beta_1 (X_i - \bar{X}) + (u_i - \bar{u})\right)}{\sum_{i=1}^n (X_i - \bar{X})^2} \\
&= \beta_1 + \frac{\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^n (X_i - \bar{X})^2}.
\end{aligned} \tag{18}$$

The Sampling Distribution of the OLS Estimators

- Note that $\sum_{i=1}^{n}(X_i-\bar{X})(u_i-\bar{u})=\sum_{i=1}^{n}(X_i-\bar{X})u_i$. Using this result in (18), we obtain

$$\widehat{\beta}_1=\beta_1+\frac{\sum_{i=1}^{n}(X_i-\bar{X})u_i}{\sum_{i=1}^{n}(X_i-\bar{X})^2} \tag{19}$$

- We can use (19) to compute $\mathrm{E}(\widehat{\beta}_1|X_1,\dots,X_n)$ as

$$\begin{aligned}\mathrm{E}(\widehat{\beta}_1|X_1,\dots,X_n)&=\beta_1+\mathrm{E}\left(\frac{\sum_{i=1}^{n}(X_i-\bar{X})u_i}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\middle|X_1,\dots,X_n\right)\\&=\beta_1+\frac{\sum_{i=1}^{n}(X_i-\bar{X})\,\mathrm{E}\left(u_i|X_1,\dots,X_n\right)}{\sum_{i=1}^{n}(X_i-\bar{X})^2}.\end{aligned} \tag{20}$$

- Under Assumptions 1 and 2, we have $\mathrm{E}\left(u_i|X_1,\dots,X_n\right)=\mathrm{E}(u_i|X_i)=0$, suggesting that $\mathrm{E}(\widehat{\beta}_1|X_1,\dots,X_n)=\beta_1$.
- Finally, the unbiasedness of $\hat{\beta}_1$ follows from the LIE:

$$\mathrm{E}(\widehat{\beta}_1)=\mathrm{E}\left(\mathrm{E}\left(\widehat{\beta}_1|X_1,\dots,X_n\right)\right)=E(\beta_1)=\beta_1. \tag{21}$$

- Similarly, it can be shown that $\mathrm{E}(\widehat{\beta}_0)=\beta_0$.

The Sampling Distribution of the OLS Estimators

- Let $v_i = (X_i - \bar{X})u_i$. Then, from (19), we have

$$\widehat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} v_i}{\left(\frac{n-1}{n}\right) s_X^2}, \tag{22}$$

where $s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is the sample variance of X.

- If n is large, we have $s_X^2 \approx \sigma_X^2$, $\bar{X} \approx \mu_X$, and $\frac{n-1}{n} \approx 1$. Thus,

$$\widehat{\beta}_1 - \beta_1 \approx \frac{\frac{1}{n}\sum_{i=1}^{n} v_i}{\sigma_X^2} \implies \operatorname{var}(\widehat{\beta}_1 - \beta_1) = \frac{\operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} v_i\right)}{(\sigma_X^2)^2}$$

$$\implies \operatorname{var}(\widehat{\beta}_1) = \frac{1}{n}\frac{\operatorname{var}(v_i)}{(\sigma_X^2)^2} = \frac{1}{n}\frac{\operatorname{var}\left((X_i - \mu_X)u_i\right)}{(\sigma_X^2)^2}. \tag{23}$$

- The result in (23) indicates that $\operatorname{var}(\widehat{\beta}_1)$ is inversely proportional to n.

The Sampling Distribution of the OLS Estimators

- Theorem 1 summarizes the large sample properties of $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

Theorem 1

Under the least squares assumptions, we have $\mathrm{E}(\widehat{\beta}_0) = \beta_0$ and $\mathrm{E}(\widehat{\beta}_1) = \beta_1$, i.e, unbiasedness. Furthermore, in large samples (as the sample size grows without a bound), $\widehat{\beta}_0$ and $\widehat{\beta}_1$ have a joint normal sampling distribution, i.e., asymptotic normality.

The Sampling Distribution of the OLS Estimators

- More specifically, under the least squares assumptions, in large samples, we have (see Appendix 4.3 in the textbook)

$$\widehat{\beta}_1 \overset{A}{\sim} N\left(\beta_1, \sigma^2_{\widehat{\beta}_1}\right), \text{ where } \quad \sigma^2_{\widehat{\beta}_1} = \frac{1}{n}\frac{\mathrm{var}\left((X_i - \mu_X)u_i\right)}{\left(\mathrm{var}(X_i)\right)^2}$$

$$\widehat{\beta}_0 \overset{A}{\sim} N\left(\beta_0, \sigma^2_{\widehat{\beta}_0}\right), \text{ where } \quad \sigma^2_{\widehat{\beta}_0} = \frac{1}{n}\frac{\mathrm{var}\left(H_i u_i\right)}{\left(\mathrm{E}(H_i^2)\right)^2} \text{ and } H_i = 1 - \frac{\mu_X}{\mathrm{E}(X_i^2)}X_i$$

- Note that since the variance of the OLS estimator is proportional to $1/n$, and as n grows without a bound, the variance shrinks to zero.

- Combining with unbiasedness, this implies that the OLS estimators are also consistent under the least squares assumptions.

- Note also that we cannot yet use the large sample distribution of the OLS estimators for inference, as they involve unknown terms.

Bibliography I

Hanck, Christoph et al. (2021). *Introduction to Econometrics with R*. URL: https://www.econometrics-with-r.org/index.html.

Stock, James H. and Mark W. Watson (2020). *Introduction to Econometrics*. Fourth. Pearson.