Lecture 7: Hypothesis Test and Confidence Intervals of Linear Regression with a Single Regressor

Zheng Tian

Contents

1 Introduction

This chapter consists of two parts. The first part concerns hypothesis testing for a single coefficient in a simple linear regression model. The basic concepts and ideas of hypothesis testing in this chapter can be naturally adopted in multiple regression models (Chapters 6 and 7). The second part goes back to some estimation issues, including a binary regressor, homoskedasticity versus heteroskedasticity, as well as the Gauss-Markov theorem, one of the most fundamental theories regarding the OLS estimation. Finally, this chapter ends up with the small sample properties of the t-statistics.

One of the features of this textbook is that it introduces the heteroskedasticity-robust standard error of the OLS estimators, which is considered as a general case and homoskedasticity as a special case. This is contrary to the common layouts of an Econometrics textbook that often first gives the assumption of homoskedasticity, which is a component of the classical OLS assumptions (equivalent to the three least squares assumptions plus the assumption of the homoskedastic and conditionally normally distributed errors). Then treat heteroskedasticity as a violation to these assumptions. Also, you should be aware that most discussions of the sample distributions in this textbook are in the context of a large sample, while the small sample statistical properties are not the focus.

2 Testing Hypotheses about One of the Regression Coefficients

2.1 A brief review of basic concepts in hypothesis tests

Let's quickly review what we have learned in Lecture 3 about hypothesis tests, taking an example of testing the true value of the population mean from a random sample.

The null versus alternative hypotheses

We want to test two contrasting hypotheses, the null hypothesis versus the alternative hypothesis.

- Two-sided tests:

$$H_0: \ E(Y) = \mu_{Y,0}$$

  v.s. $H_1: \ E(Y) \neq \mu_{Y,0}$

- One-sided test:

$$H_0: \ E(Y) = \mu_{Y,0}$$

v.s. $H_1: \ E(Y) > \mu_{Y,0}$

Test statistics

We need some tools for the testing, which is referred to as test statistics.

- When σ_Y is known, we use the z-statistics

$$z = \frac{\overline{Y} - \mu_{Y,0}}{\sigma_{\overline{Y}}} = \frac{\overline{Y} - \mu_{Y,0}}{\sigma_Y / \sqrt{n}} \xrightarrow{\text{d}} N(0,1)$$

- When σ_Y is unknown, we use the standard error of $\overline{Y}$ and compute the t-statistic[1]

$$t = \frac{\overline{Y} - \mu_{Y,0}}{SE(\overline{Y})} = \frac{\overline{Y} - \mu_{Y,0}}{s_Y / \sqrt{n}} \xrightarrow{\text{d}} N(0,1)$$

[1] In a small sample case, the exact distribution of the t-statistics is the Student-t distribution with $n-1$ degree of freedom.

The rules for hypothesis testing

We need to set up some rules for judging that under what circumstances, the null hypothesis is rejected or fails to be rejected.

- Type I and type II errors
  - **Type I error**. The null hypothesis is rejected when in fact it is true.
  - **Type II error**. The null hypothesis is not rejected when in fact it is false.
- The significance level, the critical value, and the p-value
  - **The significance level**. The pre-specified probability of type I error. $\alpha = 0.05, 0.10, \text{ or } 0.01$
  - **The critical value**. The value of the test statistic for which the test rejects the null hypothesis at the given significance level.

    For example. In a two-sided test, with the z statistic. The critical value at the 5% significance level is c_α such that $\Phi(c_\alpha) = 0.975$. Accordingly, we know $c_\alpha \approx 1.96$.
  - **The p-value**. The p-value is the probability of drawing a statistic at least as adverse to the null hypothesis as the one you actually computed in your sample, assuming the null hypothesis is correct.

    Equivalently, the p-value is the smallest significance level at which the null hypothesis could be rejected, based on the test statistic actually computed.

    Mathematically, the p-value is

$$\Pr_{H_0}\left(\left|\frac{\overline{Y} - \mu_{Y,0}}{SE(\overline{Y})}\right| > \left|\frac{\overline{Y}^{act} - \mu_{Y,0}}{SE(\overline{Y})}\right|\right) = 2\Phi(-|t^{act}|) \ .$$

- Rejection rules

  The following two statements are equivalent in terms of rejecting the null hypothesis at the 5% significance level.
  - We can reject the null if the test statistics falls into the rejection region delimited by the critical values at the 5% significance level, that is, when $|t^{act}| > c_\alpha = 1.96$,
  - We can reject the null if the p-value is less than the significance level that is 5% in this case.

  The rejection rule can be illustrated using Figure 1.

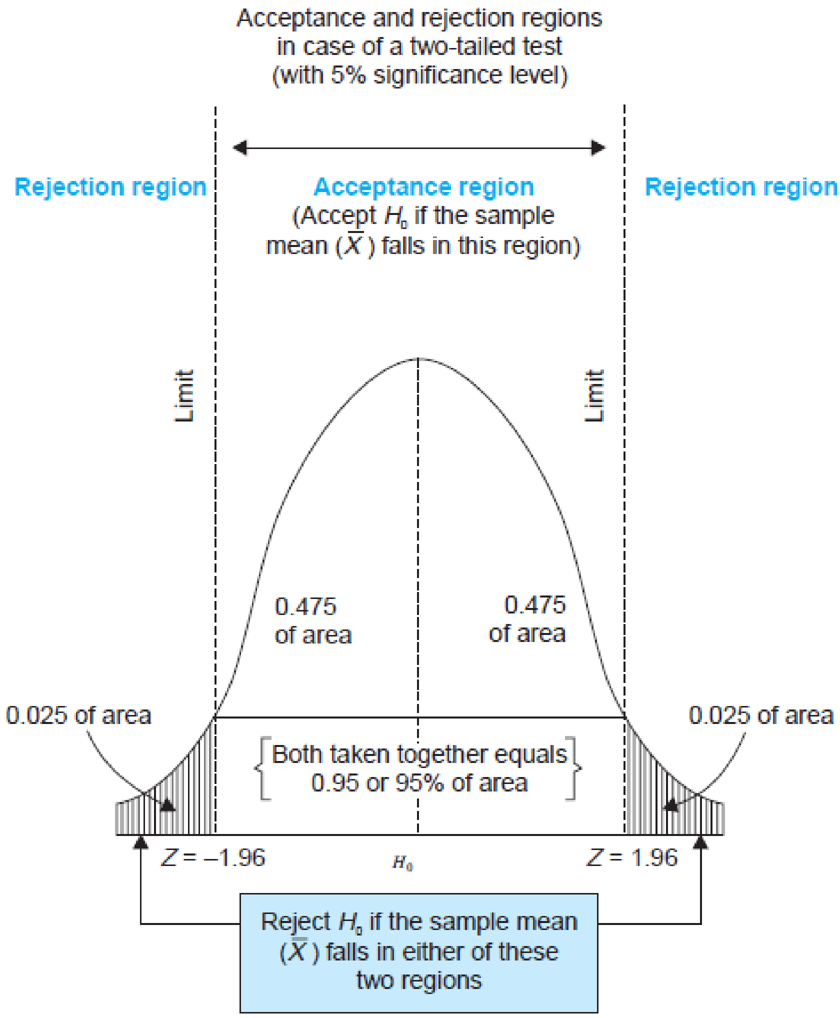


Figure 1: An illustration of a two-sided test

2.2 Two-sided hypotheses concerning β_1

Application to test scores

In the last lecture, we estimate a simple linear regression model for test scores and class sizes, which yields the following estimated sample regression function,

$$\widehat{TestScore} = 698.93 - 2.28 \times STR \tag{1}$$

Now the question faced by the superintendent of the California elementary school districts is whether the estimated coefficient on STR is valid. In the terminology of statistics, his question is whether β_1 is statistically significantly different from zero.

Testing hypotheses about the slope β_1

Note that all discussions about hypothesis testing that follows involve only the regression with a large sample size. The last section of this lecture touches upon the small sample properties of the test statistics.

- The two-sided hypothesis

  $$H_0 : \beta_1 = \beta_{1,0} \text{ vs. } H_1 : \beta_1 \neq \beta_{1,0}$$

  The null hypothesis is that β_1 is equal to a specific value $\beta_{1,0}$, and the alternative hypothesis is the opposite.

- The t-statistic

  The general form of the t-statistic is

  $$t = \frac{\text{estimator} - \text{hypothesized value}}{\text{standard error of the estimator}} \tag{2}$$

  The t-statistics for testing β_1 is then

  $$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)} \tag{3}$$

- The standard error of $\hat{\beta}_1$ is calculated as

  $$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}} \tag{4}$$

where

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \frac{\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \hat{u}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2} \tag{5}$$

- How to understand Equation 5
  - The population variance of β_1 is

    $$\sigma^2_{\hat{\beta}_1} = \frac{1}{n} \frac{\text{Var}\left((X_i - \mu_X)u_i\right)}{\left(\text{Var}(X_i)\right)^2}$$

  - The denominator in Equation (5) is a consistent estimator of $\text{Var}(X_i)^2$.
  - The numerator in Equation (5) is a consistent estimator of $\text{Var}((X_i - \mu_X)u_i)$, adjusted by $n-2$ degrees of freedom.
  - The standard error computed from Equation (5) is the **heteroskedasticity-robust standard error**, which will be explained in detail shortly in this lecture.
- Compute the p-value

  The p-value is the probability of observing a value of $\hat{\beta}_1$ at least as different from $\beta_{1,0}$ as the estimate actually computed ($\hat{\beta}_1^{act}$), assuming that the null hypothesis is correct. Accordingly, under the null hypothesis, the p-value for testing β_1 can be expressed with a probability function as

  $$\begin{aligned} p\text{-value} &= \Pr_{H_0}\left(|\hat{\beta}_1 - \beta_{1,0}| > |\hat{\beta}_1^{act} - \beta_{1,0}|\right) \\ &= \Pr_{H_0}\left(\left|\frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}\right| > \left|\frac{\hat{\beta}_1^{act} - \beta_{1,0}}{SE(\hat{\beta}_1)}\right|\right) \\ &= \Pr_{H_0}\left(|t| > |t^{act}|\right) \end{aligned}$$

  With a large sample, p-value $= \Pr\left(|t| > |t^{act}|\right) = 2\Phi(-|t^{act}|)$.

  The null hypothesis is rejected at the 5% significance level if the p-value < 0.05 or, equivalently, $|t^{act}| > 1.96$.
- Application to test scores

  The OLS estimation of the linear regression model of test scores against student-teacher ratios, together with the standard errors of all parameters in the model, can be represented using the following equation,

  $$\widehat{TestScore} = \underset{(10.4)}{698.9} - \underset{(0.52)}{2.28} \times STR,\ R^2 = 0.051,\ SER = 1.86$$

The **heteroskedasticity-robust** standard errors are reported in the parentheses, that is, $SE(\hat{\beta}_0) = 10.4$ and $SE(\hat{\beta}_1) = 0.52$.

The superintendent's question is whether β_1 is significant for which we can test the null hypothesis against the alternative one as

$$H_0 : \beta_1 = 0, H_1 : \beta_1 \neq 0$$

The t-statistics is

$$t = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{-2.28}{0.52} = -4.38 < -1.96$$

The p-value associated with $t^{act} = -4.38$ is approximately 0.00001, which is far less than 0.05.

Based on the t-statistics and the p-value, we can say the null hypothesis is rejected at the 5% significance level. In English, it means that the student-teacher ratios do have a significant effect on test scores.

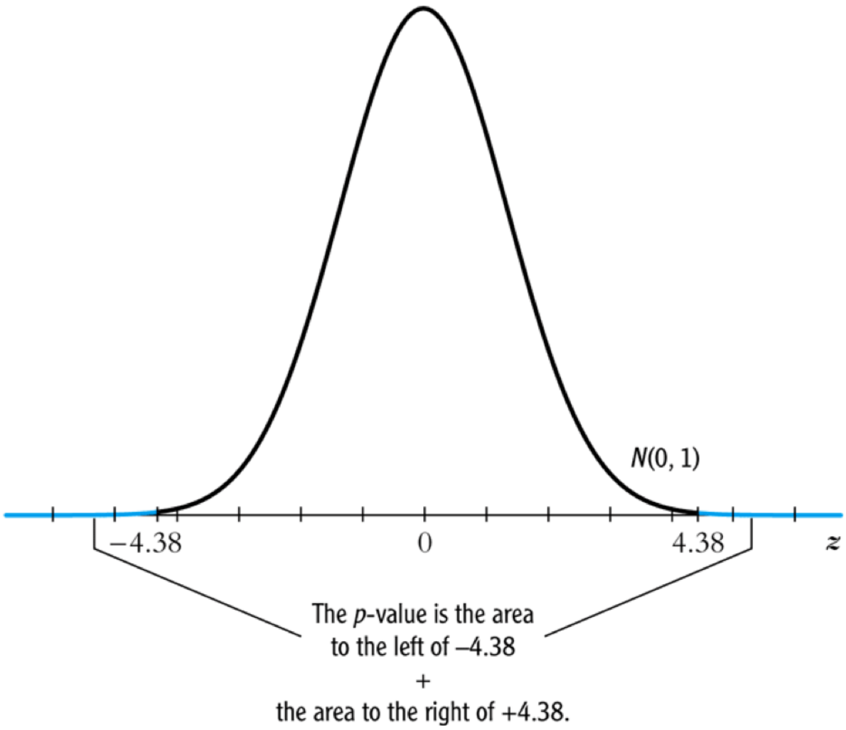


Figure 2: Calculating the p-value of a two-sided test when $t^{act} = -4.38$

2.3 The one-sided alternative hypothesis

The one-sided hypotheses

In some cases, it is appropriate to use a one-sided hypothesis test. For example, the superintendent of the California school districts want to know whether class sizes have a negative effect on test scores, that is, $\beta_1 < 0$.

For a one-sided test, the null hypothesis and the one-sided alternative hypothesis are [2]

$$H_0 : \beta_1 = \beta_{1,0} \text{ vs. } H_1 : \beta_1 < \beta_{1,0}$$

The one-sided left-tail test

- The t-statistic is the same as in a two-sided test

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

- Since we test $\beta_1 < \beta_{1,0}$, if this is true, the t-statistics should be statistically significantly less than zero.
- The p-value is computed as $\Pr(t < t^{act}) = \Phi(t^{act})$.
- The null hypothesis is rejected at the 5% significance level when p-value < 0.05 or $t^{act} < -1.645$.
- In the application of test scores, the t-statistics is -4.38, which is less than -1.645 and -2.33 (the critical value for a one-sided test with a 1% significance level). Thus, the null hypothesis is rejected at the 1% level.

3 Confidence Intervals for a Regression Coefficient

3.1 Two equivalent definitions of confidence intervals

Recall that a 95% **confidence interval** for β_1 has two equivalent definitions:

1. It is the set of values that cannot be rejected using a two-sided hypothesis test with a 5% significance level.
2. It is an interval that has a 95% probability of containing the true value of β_1.

Let's go back to Figure 1. According to the first definition, the acceptance region contains the values of the test statistics that fail to reject the null hypothesis, which corresponds to the values of β_1 that cannot be rejected.

[2]Note that the trick here is we put the desired hypothesis to the alternative place.

3.2 Construct the 95% confidence interval for β_1

The 95% confidence interval for β_1 can be constructed using the t-statistic, assuming that with large samples, the t-statistic is approximately normally distributed. The 95% critical value of a standard normal distribution is 1.96. Therefore, we can obtain the 95% confidence interval for β_1 by the following steps

$$-1.96 \leq \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \leq 1.96$$
$$\hat{\beta}_1 - 1.96SE(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + 1.96SE(\hat{\beta}_1)$$

The 95% confidence interval for β_1 is

$$\left[\hat{\beta}_1 - 1.96SE(\hat{\beta}_1),\ \hat{\beta}_1 + 1.96SE(\hat{\beta}_1)\right]$$

3.3 The application to test scores

In the application to test scores, given that $\hat{\beta}_1 = -2.28$ and $SE(\hat{\beta}_1) = 0.52$, the 95% confidence interval for β_1 is $-2.28 \pm 1.96 \times 0.52$, or $-3.30 \leq \beta_1 \leq -1.26$.

Note that the confidence interval only spans over the negative region with zero leaving outside the interval, which implies that the null hypothesis of $\beta_1 = 0$ can be rejected at the 5% significance level.

3.4 Confidence intervals for predicted effects of changing X

β_1 is the marginal effect of X on Y, that is,

$$\beta_1 = \frac{\mathrm{d}Y}{\mathrm{d}X} \Rightarrow \mathrm{d}Y = \beta_1 \mathrm{d}X$$

When X changes by ΔX, Y changes by $\beta_1 \Delta X$.

So the 95% confidence interval for $\beta_1 \Delta X$ is

$$\left[\hat{\beta}_1 \Delta X - 1.96SE(\hat{\beta}_1)\Delta X,\ \hat{\beta}_1 \Delta X + 1.96SE(\hat{\beta}_1)\Delta X\right]$$

4 Regression When X is a Binary Variable

4.1 A binary variable

A **binary variable** takes on values of one if some condition is true and zero otherwise, which is also called a **dummy variable**, a **categorical variable**, or an **indicator variable**.

For example,

$$D_i = \begin{cases} 1, & \text{if the } i^{th} \text{ subject is female} \\ 0, & \text{if the } i^{th} \text{ subject is male} \end{cases}$$

The linear regression model with a dummy variable as a regressor is

$$Y_i = \beta_0 + \beta_1 D_i + u_i,\ i = 1, \ldots, n \tag{6}$$

The coefficient on D_i is estimated by the OLS estimation method in the same way as a continuous regressor. The difference lies in how we interpret β_1.

4.2 Interpretation of the regression coefficients

Given that the assumption $E(u_i|D_i) = 0$ holds in Equation (6), we have two population regression functions for the two cases, that is,

- When $D_i = 1$, $E(Y_i|D_i = 1) = \beta_0 + \beta_1$
- When $D_i = 0$, $E(Y_i|D_i = 0) = \beta_0$

Therefore, $\beta_1 = E(Y_i|D_i = 1) - E(Y_i|D_i = 0)$, that is, **the difference in the population means** between two groups represented by $D_i = 1$ and $D_i = 0$, respectively.

4.3 Hypothesis tests and confidence intervals

The hypothesis tests and confidence intervals for the coefficient on a binary variable follows the same procedure of those for a continuous variable X.

Usually, the null and alternative hypotheses concerning a dummy variable are

$$H_0 : \beta_1 = 0 \text{ vs. } H_1 : \beta_1 \neq 0$$

Therefore, the t-statistic is

$$t = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)}$$

And the 95% confidence interval is

$$\hat{\beta}_1 \pm 1.96SE(\hat{\beta}_1)$$

5 Heteroskedasticity and Homoskedasticity

5.1 What are heteroskedasticity and homoskedasticity?

Homoskedasticity

The error term u_i is **homoskedastic** if the conditional variance of u_i given X_i is constant for $i = 1, \ldots, n$. Mathematically, it says $\text{Var}(u_i|X_i) = \sigma^2, \text{ for } i = 1, \ldots, n$, i.e., the variance of u_i for all i is a constant and does not depend on X_i.

Heteroskedasticity

In contrast, the error term u_i is **heteroskedastic** if the conditional variance of u_i given X_i changes on X_i for $i = 1, \ldots, n$. That is, $\text{Var}(u_i|X_i) = \sigma_i^2, \text{ for } i = 1, \ldots, n$.

e.g.. A multiplicative form of heteroskedasticity is $\text{Var}(u_i|X_i) = \sigma^2 f(X_i)$ where $f(X_i)$ is a function of X_i, for example, $f(X_i) = X_i$ as a simplest case.

Figure 3 for a visual comparison between homoskedasticity and heteroskedasticity.

5.2 Mathematical implications of homoskedasticity

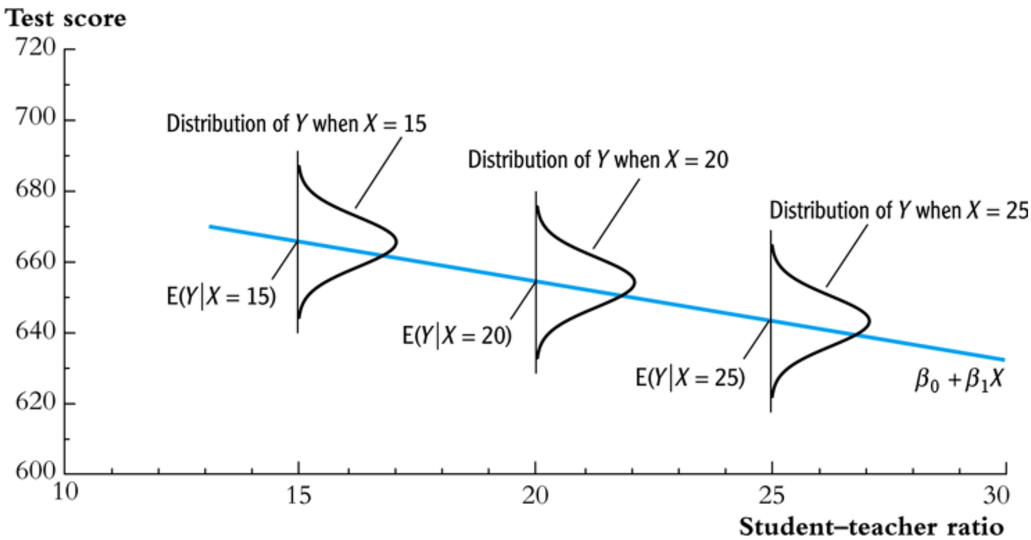
Unbiasedness, consistency, and the asymptotic distribution

As long as the least squares assumptions holds, whether the error term, u_i, is homoskedastic or heteroskedastic does not affect unbiasedness, consistency, and the asymptotic normal distribution of the OLS estimators.

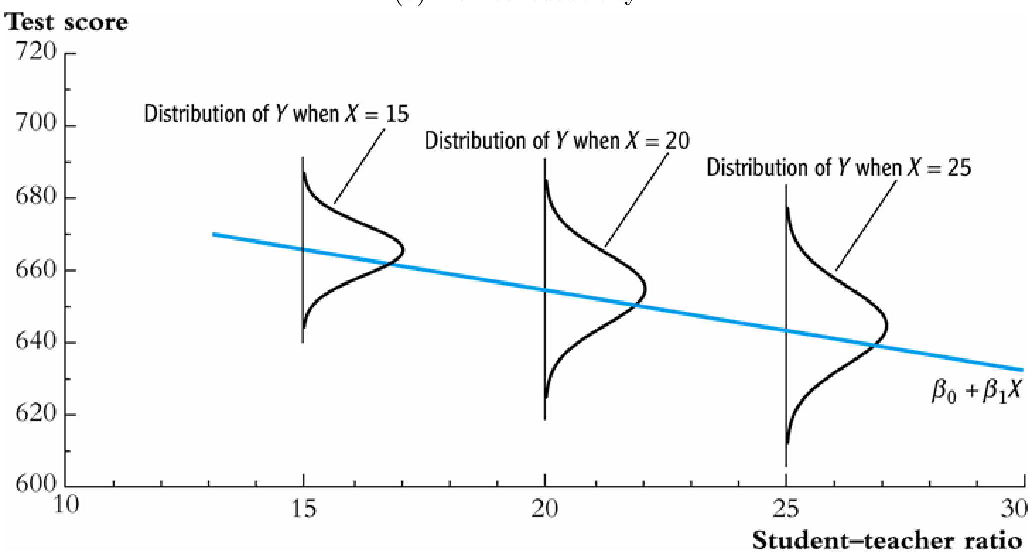
- The unbiasedness requires that $E(u_i|X_i) = 0$
- The consistency requires that $E(X_i u_i) = 0$, which is true if $E(u_i|X_i) = 0$.
- The asymptotic normal distribution requires additionally that $\text{Var}((X_i - \mu_X)u_i) < \infty$, which still holds as long as Assumption 3 holds, that is, no extreme outliers of X_i.

Efficiency

The existence of heteroskedasticity affects the efficiency of the OLS estimator



(a) Homoskedasticity



(b) Heteroskedasticity

Figure 3: Homoskedasticity Versus Heteroskedasticity

- Suppose $\hat{\beta}_1$ and $\tilde{\beta}_1$ are both unbiased estimators of β_1. Then, $\hat{\beta}_1$ is said to be more **efficient** than $\tilde{\beta}_1$ if $\text{Var}(\hat{\beta}_1) < \text{Var}(\tilde{\beta}_1)$.
- When the errors are homoskedastic, the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are efficient among all estimators that are linear in $Y_1, \ldots, Y_n$ and are unbiased, conditional on $X_1, \ldots, X_n$.
- See the Gauss-Markov Theorem below.

5.3 The homoskedasticity-only variance formula

Recall that we can write $\hat{\beta}_1$ as

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_i (X_i - \bar{X}) u_i}{\sum_i (X_i - \bar{X})^2}$$

Therefore, if u_i for $i = 1, \ldots, n$ is homoskedastic and σ^2 is known, then

$$\text{Var}(\hat{\beta}_1 | X_i) = \frac{\sum_i (X_i - \bar{X})^2 \text{Var}(u_i | X_i)}{\left[\sum_i (X_i - \bar{X})^2\right]^2} = \frac{\sigma^2}{\sum_i (X_i - \bar{X})^2} \tag{7}$$

When σ^2 is unknown, then we use $s_u^2 = 1/(n-2) \sum_i \hat{u}_i^2$ as an estimator of σ^2. Thus, the homoskedasticity-only estimator of the variance of $\hat{\beta}_1$ is

$$\tilde{\sigma}^2_{\hat{\beta}_1} = \frac{s_u^2}{\sum_i (X_i - \bar{X})^2} \tag{8}$$

And the homoskedasticity-only standard error is $SE(\hat{\beta}_1) = \sqrt{\tilde{\sigma}^2_{\hat{\beta}_1}}$.

Recall that the heteroskedasticity-robust standard error is

$$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}$$

where

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \frac{\frac{1}{n-2} \sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2}{\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right]^2}$$

which is also referred to as Eicker-Huber-White standard errors.

5.4 What does this mean in practice?

- Heteroskedasticity is common in cross-sectional data. If you do not have strong beliefs in homoskedasticity, then it is always safer to report the heteroskedasticity-robust standard errors and use these to compute the robust t-statistic.
- In most software, the default setting is to report the homoskedasticity-only standard errors. Therefore, you need to manually add the option for the robust estimation.

– In R, you can use the following codes

```
library(lmtest)
model1 <- lm(testscr ~ str, data = classdata)
coeftest(model1, vcov = vcovHC(model1, type="HC1"))
```

– In STATA, you can use

```
regress testscr str, robust
```

6 The Theoretical Foundations of Ordinary Least Squares

6.1 The Gauss-Markov conditions

We have already known the least squares assumptions: for $i = 1, \dots, n$, (1) $E(u_i|X_i) = 0$, (2) (X_i, Y_i) are i.i.d., and (3) large outliers are unlikely.

The Gauss-Markov conditions provide anther version of these assumptions plus the assumption of homoskedastic errors.

The Gauss-Markov conditions

For $\mathbf{X} = [X_1, \dots, X_n]$ [3]

1. $E(u_i|\mathbf{X}) = 0$
2. $\text{Var}(u_i|\mathbf{X}) = \sigma_u^2$, $0 < \sigma_u^2 < \infty$
3. $E(u_i u_j|\mathbf{X}) = 0$, $i \neq j$

From the three Least Squares Assumptions and the homoskedasticity assumption to the Gauss-Markov conditions

Note that the conditional expectations in the G-M conditions are in terms of all observations $\mathbf{X}$, not just one observation, X_i. However, all the G-M conditions can be derived from the least squares assumptions plus the homoskedasticity assumption. Specifically,

- Assumptions (1) and (2) imply $E(u_i|\mathbf{X}) = E(u_i|X_i) = 0$.
- Assumptions (1) and (2) imply $\text{Var}(u_i|\mathbf{X}) = \text{Var}(u_i|X_i)$. With the homoskedasticity assumption, $\text{Var}(u_i|X_i) = \sigma_u^2$, Assumption (3) then implies $0 < \sigma_u^2 < \infty$.
- Assumptions (1) and (2) imply that $E(u_i u_j|\mathbf{X}) = E(u_i u_j|X_i, X_j) = E(u_i|X_i)E(u_j|X_j) = 0$.

[3] Here I use the vector notation to represent all observations of X_i for $i = 1, \dots, n$. We will formally introduce the matrix notation for a linear regression model and the OLS estimation in the next lecture.

6.2 Linear conditionally unbiased estimator

The general form of a linear conditionally unbiased estimator of β_1

The class of linear conditionally unbiased estimators consists of all estimators of β_1 that are linear function of $Y_i, \ldots, Y_n$ and that are unbiased, conditioned on $X_1, \ldots, X_n$.

For any linear estimator $\tilde{\beta}_1$, it can be written as

$$\tilde{\beta}_1 = \sum_{i=1}^{n} a_i Y_i \tag{9}$$

where the weights a_i for $i = 1, \ldots, n$ depend on $X_1, \ldots, X_n$ but not on $Y_1, \ldots, Y_n$.

$\tilde{\beta}_1$ is conditionally unbiased means that

$$E(\tilde{\beta}_1|\mathbf{X}) = \beta_1 \tag{10}$$

By the Gauss-Markov conditions, from Equation (9), we can have

$$\begin{aligned} E(\tilde{\beta}_1|\mathbf{X}) &= \sum_i a_i E(\beta_0 + \beta_1 X_i + u_i|\mathbf{X}) \\ &= \beta_0 \sum_i a_i + \beta_1 \sum_i a_i X_i \end{aligned}$$

For Equation (10) being satisfied with any β_0 and β_1, we must have

$$\sum_i a_i = 0 \text{ and } \sum_i a_i X_i = 1$$

The OLS esimator $\hat{\beta}_1$ is a linear conditionally unbiased estimator

We have known that $\hat{\beta}_1$ is unbiased both conditionally and unconditionally. Next, we show that it is linear.

$$\hat{\beta}_1 = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2} = \frac{\sum_i (X_i - \bar{X}) Y_i}{\sum_i (X_i - \bar{X})^2} = \sum_i \hat{a}_i Y_i$$

where the weights are

$$\hat{a}_i = \frac{X_i - \bar{X}}{\sum_i (X_i - \bar{X})^2}, \text{ for } i = 1, \ldots, n$$

Since $\hat{\beta}_1$ is a linear conditionally unbiased estimator, we must have

$$\sum_i \hat{a}_i = 0 \text{ and } \sum_i \hat{a}_i X_i = 1$$

which can be simply verified.

6.3 The Gauss-Markov Theorem

The Gauss-Markov Theorem for $\hat{\beta}_1$ states

> If the Gauss-Markov conditions hold, then the OLS estimator $\hat{\beta}_1$ is the *B*est (most efficient) *L*inear conditionally *U*nbiased *E*stimator (BLUE).

The theorem can also be applied to $\hat{\beta}_0$.

The proof of the Gauss-Markov theorem is in Appendix 5.2. A key in this proof is that we can rewrite the expression of any linear conditionally unbiased estimator $\tilde{\beta}_1$ as

$$\tilde{\beta}_1 = \sum_i a_i Y_i = \sum_i (\hat{a}_i + d_i) Y_i = \hat{\beta}_1 + \sum_i d_i Y_i$$

And the goal of the proof is to show that

$$\text{Var}(\hat{\beta}_1|\mathbf{X}) \leq \text{Var}(\tilde{\beta}_1|\mathbf{X})$$

The equality holds only when $\tilde{\beta}_1 = \hat{\beta}_1$.

6.4 The limitations of the Gauss-Markov theorem

1. The Gauss-Markov conditions may hold in practice. Any violation of the Gauss-Markov conditions will result in the OLS estimator not being BLUE. The table below summarizes the cases in which a kind of violation occurs, the consequences of such violation to the OLS estimators, and possible remedies.

Table 1: Summary of Violations of the Gauss-Markov Theorem

Violation	Cases	Consequences	Remedies
$E(u \mid X) \neq 0$	omitted variables, endogeneity	biased	more \$X\$'s, IV method
$\text{Var}(u_i \mid X)$ not constant	heteroskedasticity	inefficient	WLS, GLS, HCCME
$E(u_i u_j \mid X) \neq 0$	autocorrelation	inefficient	GLS, HAC

2. There are other candidate estimators that are not linear and conditionally unbiased; under some conditions, these estimators are more efficient than the OLS estimators.

7 Using the t-Statistic in Regression When the Sample Size is Small

7.1 The classical assumptions of the least squares estimation

We first expand the OLS assumptions by two additional ones. One is the assumption of the homoskedastic errors, and another one is the assumption that the conditional distribution of u_i

given X_i is the normal distribution, i.e., $u_i \sim N(0, \sigma_u^2)$ for $i = 1, \ldots, n$.

All these assumptions together are often referred to as the classical assumptions of the least squares estimation. For $i = 1, 2, \ldots, n$

- Assumption 1: $E(u_i|X_i) = 0$ (exogeneity of X)
- Assumption 2: (X_i, Y_i) are i.i.d. (IID of $X, Y,$ and u)
- Assumption 3: $0 < E(X_i^4) < \infty$ and $0 < E(Y_i^4) < \infty$ (No large outliers)
- Extended Assumption 4: $\text{Var}(u_i|X_i) = \sigma_u^2,$ and $0 < \sigma_u^2 < \infty$ (homoskedasticity)
- Extended Assumption 5: $u_i|X_i \sim N(0, \sigma_u^2)$ (normality)

7.2 The t-Statistic and the Student-t Distribution

Under all the classical assumptions, we can construct the t-statistic for hypothesis testing of a single coefficient. Even with a small samples, the t-statistic has an exact Student-t distribution.

The t-statistic is for β_1

$$H_0 : \beta_1 = \beta_{1,0} \text{ vs } H_1 : \beta_1 \neq \beta_{1,0}$$

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{\hat{\sigma}_{\hat{\beta}_1}} \tag{11}$$

where

$$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{s_u^2}{\sum_i (X_i - \bar{X})^2} \text{ and } s_u^2 = \frac{1}{n-2} \sum_i \hat{u}_i^2$$

the former of which is the homoskedasticity-only standard error of $\hat{\beta}_1$ and the latter is the standard error of the regression.

The Student-t distribution of t

The t statistic can be rewritten as

$$t = \frac{(\hat{\beta}_1 - \beta_{1,0})/\sigma_{\hat{\beta}_1}}{\sqrt{\frac{\hat{\sigma}_{\hat{\beta}_1}^2}{\sigma_{\hat{\beta}_1}^2}}} = \frac{z_{\hat{\beta}_1}}{\sqrt{\frac{s_u^2}{\sigma_u^2}}} = \frac{z_{\hat{\beta}_1}}{\sqrt{\frac{W}{n-2}}} \tag{12}$$

where

$$\sigma_{\hat{\beta}_1}^2 = \frac{\sigma_u^2}{\sum_i (X_i - \bar{X})^2}$$

is the homoskedasticity-only variance of $\hat{\beta}_1$ when the variance of errors σ_u^2 is known.

$$z_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sigma_{\hat{\beta}_1}}$$

is the z-statistic which has a standard normal distribution, that is, $z_{\hat{\beta}_1} \sim N(0, 1)$

$$W = (n-2)\frac{s_u^2}{\sigma_u^2} = \frac{\sum_i \hat{u}_i^2}{\sigma_u^2} = \sum_i \left(\frac{\hat{u}_i}{\sigma_u}\right)^2$$

It can be shown that W is the sum of squares of $(n-2)$ independent standard normally distributed variables, which results in a chi-squared distribution with $(n-2)$ degrees of freedom. That is, $W \sim \chi^2(n-2)$, which is also independent of $z_{\hat{\beta}_1}$. Therefore, the t-statistic in Equation (12), as the ratio of $z_{\hat{\beta}_1}$ and $\sqrt{W/(n-2)}$, is distributed as $t(n-2)$.