Introductory Econometrics
Ch2 The Simple Regression Model: Properties of Simple Regression Model

LIU Chenyuan

Spring 2024

Outline

Properties of OLS on Any Sample of Data

Goodness of Fit

Units of Measurement and Functional Form

Expected Values and Variances of the OLS Estimators

Outline

Properties of OLS on Any Sample of Data

Goodness of Fit

Units of Measurement and Functional Form

Expected Values and Variances of the OLS Estimators

Terminology

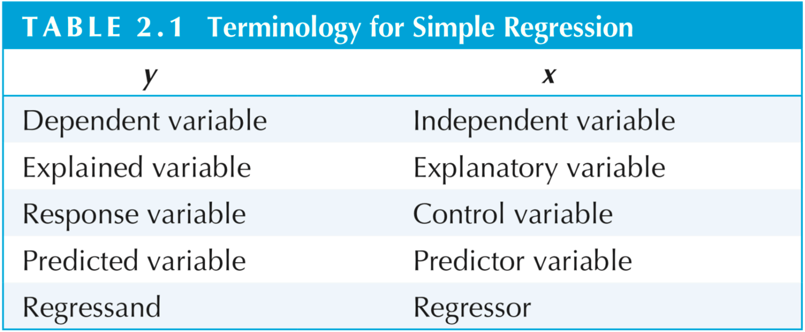
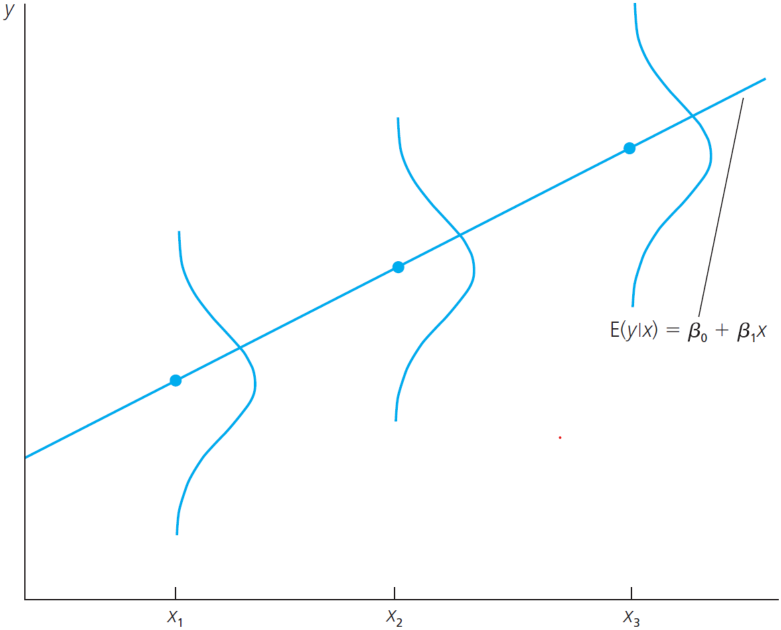


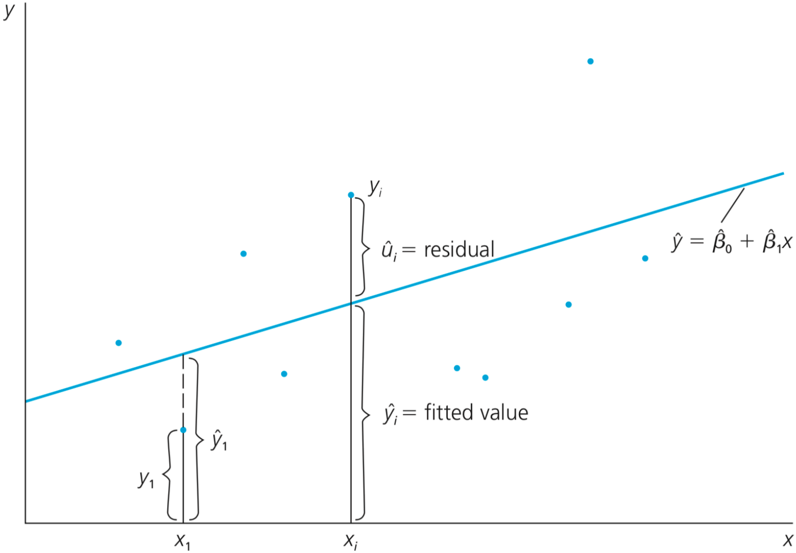
TABLE 2.1 Terminology for Simple Regression

y	*x*
Dependent variable	Independent variable
Explained variable	Explanatory variable
Response variable	Control variable
Predicted variable	Predictor variable
Regressand	Regressor

Population Regression Model



Sample Regression Model


y
y_i
$\hat{u}_i$ = residual
$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
$\hat{y}_i$ = fitted value
$\hat{y}_1$
y_1
x_1
x_i
x

- When we have a regression model

$$y = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{u},$$

  we call this “a regression of y on x”, or we “regress y on x”.
- We now derive the properties of the simple regression model. Note that these properties hold regardless of the interpretations.

Properties of OLS on Any Sample of Data

Prove the following properties

$$\sum_{i=1}^{N} \hat{u}_i = 0. \tag{1}$$

$$\sum_{i=1}^{N} x_i \hat{u}_i = 0. \tag{2}$$

$$\sum_{i=1}^{N} \hat{y}_i \hat{u}_i = 0. \tag{3}$$

$$\hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y}. \tag{4}$$

▸ Answer

Outline

Properties of OLS on Any Sample of Data

Goodness of Fit

Units of Measurement and Functional Form

Expected Values and Variances of the OLS Estimators

Goodness of Fit

- We want to measure how well our model fits the data.
- We can decompose y_i into two parts: the fitted value and the residual.

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

- The variation in y comes from two sources: the variation our model can explain ($\hat{y}_i$), and the variation the model does not capture ($\hat{u}_i$).

Goodness of Fit

Define the following terms:

- Total sum of squares (SST)

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

- Explained sum of squares(SSE)

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

Note that $\bar{\hat{y}}_i = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_i = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{u}_i) = \frac{1}{N}\sum_{i=1}^{N} y_i - \frac{1}{N}\sum_{i=1}^{N} \hat{u}_i = \bar{y}$.

- Residual sum of squares (SSR)

$$SSR = \sum_{i=1}^{N} \hat{u}_i^2.$$

- There is a nice relationship between the three terms:

$$\sum_{i=1}^{N}(y_i - \overline{y})^2 = \sum_{i=1}^{N}(\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{N} \hat{u}_i^2.$$

$$SST = SSE + SSR.$$

Proof:

$$
\begin{aligned}
\sum_{i=1}^{N}(y_i-\overline{y})^2 &= \sum_{i=1}^{N}[(y_i-\hat{y}_i)+(\hat{y}_i-\overline{y})]^2 \\
&= \sum_{i=1}^{N}[\hat{u}_i+(\hat{y}_i-\overline{y})]^2 \\
&= \sum_{i=1}^{N}\hat{u}_i^2+2\sum_{i=1}^{N}\hat{u}_i(\hat{y}_i-\overline{y})+\sum_{i=1}^{N}(\hat{y}_i-\overline{y})^2 \\
&= SSR+2\sum_{i=1}^{N}\hat{u}_i(\hat{y}_i-\overline{y})+SSE \\
&= SSR+SSE.
\end{aligned}
$$

where $\sum_{i=1}^{N}\hat{u}_i(\hat{y}_i-\overline{y})=\sum_{i=1}^{N}\hat{u}_i\hat{y}_i-\overline{y}\sum_{i=1}^{N}\hat{u}_i=0.$

Goodness of Fit: R^2

R^2 measures variation in y explained by the model

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

- R^2 measures how well the model fits the data
- R^2 is always between 0 and 1.
- R^2 measures the correlation between x and y. It does not say anything about the causal relationship between them.

Outline

Properties of OLS on Any Sample of Data

Goodness of Fit

Units of Measurement and Functional Form

Expected Values and Variances of the OLS Estimators

Units of Measurement

If we change the units of measurement of variables, what will happen to the OLS statistics?

Example: ROE and CEO salary.

$$\widehat{salary_k} = 963 + 18 \times ROE,$$

- $salary_k$: CEO salary in thousands of yuan
- ROE: defined in terms of net income as a percentage of common equity, e.g., $ROE = 10$ means the return on equity is 10%.

Question: if we measure *salary* as yuan, how will the coefficients change?

Changing Unit of Measurement

Consider the simple regression model:

$$y = \beta_0 + \beta_1 x + u.$$

Now suppose $y^* = w_1 y$ and $x^* = w_2 x$, Then for this model:

$$y^* = \beta_0^* + \beta_1^* x^* + u.$$

What's the relationship between the OLS estimators of these two models, $\hat{\beta}_0^*$ and $\hat{\beta}_0$, and $\hat{\beta}_1^*$ and $\hat{\beta}_1$?

Recall that

$$\begin{aligned}
\hat{\beta}_1^* &= \frac{\sum_{i=1}^N (x_i^* - \bar{x}^*)(y_i^* - \bar{y}^*)}{\sum_{i=1}^N (x_i^* - \bar{x}^*)^2} \\
&= \frac{\sum_{i=1}^N (w_2 x_i - w_2 \bar{x})(w_1 y_i - w_1 \bar{y})}{\sum_{i=1}^N (w_2 x_i - w_2 \bar{x})^2} \\
&= \frac{w_1 w_2 \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{w_2^2 \sum_{i=1}^N (x_i - \bar{x})^2} = \frac{w_1}{w_2} \hat{\beta}_1 . \\
\hat{\beta}_0^* &= \bar{y}^* - \hat{\beta}_1^* \bar{x}^* = w_1 \bar{y} - (\frac{w_1}{w_2} \hat{\beta}_1)(w_2 \bar{x}) \\
&= w_1 (\bar{y} - \hat{\beta}_1 \bar{x}) = w_1 \hat{\beta}_0 .
\end{aligned}$$

So we have

$$\hat{\beta}_0^* = w_1 \hat{\beta}_0,$$
$$\hat{\beta}_1^* = \frac{w_1}{w_2} \hat{\beta}_1.$$

Applying the formula:

$$\widehat{salary_y} = 963,000 + 18,000 \times ROE.$$

The formula aligns well with the interpretation. If ROE increases by one unit, then $salary_k$ increases by 18, which means salary increases by 18,000.

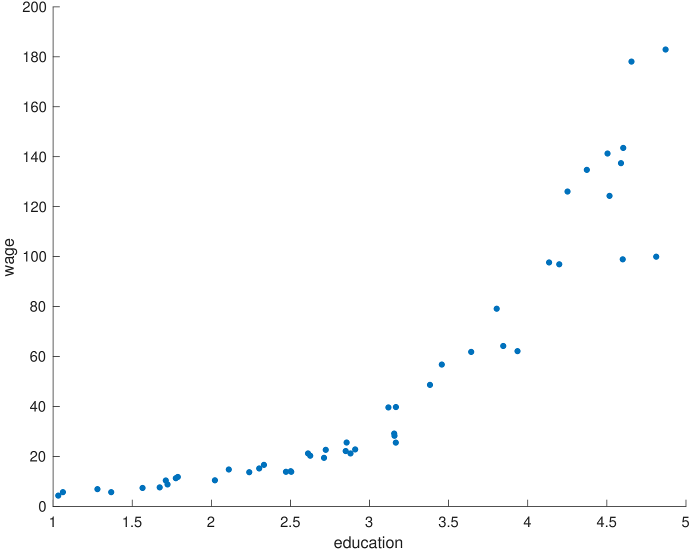
Functional Forms

- Recall that in the simple linear regression model,

  $$\beta_1 = \frac{\partial E(y|x)}{\partial x},$$

  β_1: when x increases by 1 unit, the change in y on average.
- What if x and y have a non-linear relationship?

Can we use what we have learned to model the relationship between *wage* and *edu*?



Adding Non-linear Factors into Regression Model

$$log(y) = \beta_0 + \beta_1 x + u.$$

Here, $log(\cdot)$ is the **natural logarithm** (to the base of e).

$$\beta_1 = \frac{dlog(y)}{dx} = \frac{dy}{y}\frac{1}{dx}$$

Let $\%\Delta y = 100\frac{\Delta y}{y_0}$. If $\Delta u = 0$, then

$$\%\Delta y \approx (100 \cdot \beta_1)\Delta x.$$

$100\beta_1$ represents when x increases by one unit, the percentage change in y.

Quiz: Education and Wage

$$\widehat{log(wage)} = 0.584 + 0.083edu.$$

What's the meaning of the slope coefficient?[1]

1. For every additional year of education, the wage increases by 0.083 yuan.
2. For every additional year of education, the wage increases by about 0.083%.
3. For every additional year of education, the wage increases by about 8.3%
4. If the year of education increases by 1%, the wage increases by about 0.083 yuan.

[1] Answers are on page 49.

Interpreting Regression Models

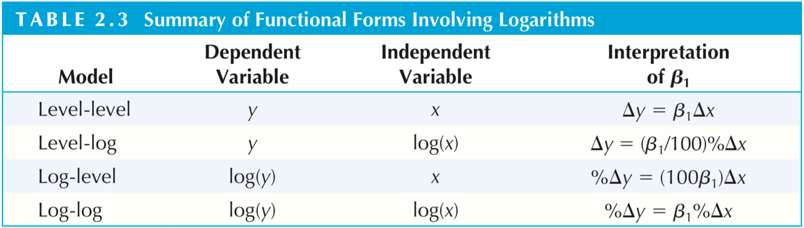


TABLE 2.3 Summary of Functional Forms Involving Logarithms

Model	Dependent Variable	Independent Variable	Interpretation of β_1
Level-level	y	x	$\Delta y = \beta_1 \Delta x$
Level-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$
Log-level	$\log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
Log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

Quiz: Constant Elasticity Model

$$\widehat{log(quantity)} = 4.822 + 0.257 log(price).$$

The elasticity of y with respect to x is $\frac{dy/y}{dx/x}$. What's the meaning of the slope coefficient?

1. When the price increases by 1 unit, the quantity increases by 0.257 units.
2. When the price increases by 1 unit, the quantity increases by 25.7%.
3. When the price increases by 1%, the quantity increases by 0.257 unit.
4. When price increases by 1%, the quantity increases by 0.257%.
5. When the price increases by 1%, the quantity increases by 25.7%.

Outline

Properties of OLS on Any Sample of Data

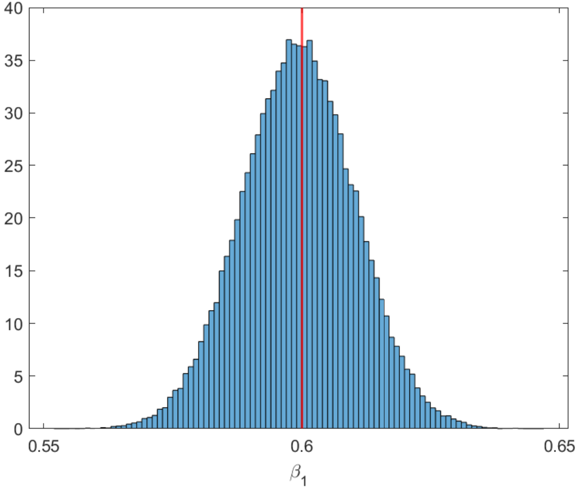
Goodness of Fit

Units of Measurement and Functional Form

Expected Values and Variances of the OLS Estimators

Distribution of the OLS Estimator

- $(\hat{\beta}_0, \hat{\beta}_1)$ are determined by the sample: different samples will produce different estimates.
- If we draw multiple samples, and for each sample, we calculate $(\hat{\beta}_0, \hat{\beta}_1)$, we can plot the values in a histogram.
- These estimates will have a probabilistic distribution, and we could calculate the mean and variance of them.
- Notice the difference between estimator, a random variable v.s estimates, the realization of the random variable.


40
35
30
25
20
15
10
5
0
0.55
0.6
0.65
β_1

Unbiasedness of OLS

- What is the property of a good estimator?
- **Unbiasedness** means the expectation of the estimator equals the true value.

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

To get this property, we need more assumptions.

Simple Linear Regression (SLR) Assumptions

1. SLR.1 Linear in Parameters
2. SLR.2 Random Sampling
3. SLR.3 Sample Variation in the Explanatory Variable
4. SLR.4 Zero Conditional Mean: $E(u|x) = 0$

We first define the basic model:

SLR.1-Linear in Parameters

In the population model, the dependent variable, y, is related to the independent variable, x, and the error (or disturbance), u, as

$$y = \beta_0 + \beta_1 x + u.$$

where β_0 and β_1 are the population intercept and slope parameters, respectively.

Notice that in making this assumption, we have really moved to the "structural world." That is we are really saying that this is the actual data generating process and our goal is to uncover the true parameters.

SLR.2 Random Sampling

We have a random sample of size N, $\{(x_i, y_i) : (i = 1, 2, ..., N)\}$, following the population model in SLR.1.

- We need to consider random sampling when choosing the sample.
- Time series data typically do not satisfy this assumption.
- Under this assumption $cov(u_i, u_j) = 0, \quad \forall i \neq j$.

SLR.3 Sample Variation in the Explanatory Variable

The sample outcomes on x, namely, $\{x_i : i = 1, 2, ..., N\}$, are not all the same value.

Recall the formula for $\hat{\beta}_1$: the denominator is $\sum_{i=1}^{N}(x_i - \bar{x})^2$, which is not zero only if x has some variation.

SLR.4 Zero Conditional Mean

The error u has an expected value of zero given any value of the explanatory variable. In other words,

$$E(u|x) = 0.$$

- This condition is critical to make sure of a causal interpretation.
- We should examine it case by case, sometimes using economic theory.

The Unbiasedness of the OLS Estimator

Unbiasedness of OLS

Using assumptions SLR.1 through SLR.4, for any value of β_0 and β_1, we have

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

In other words $\hat{\beta}_1$ is unbiased for β_1, $\hat{\beta}_0$ is unbiased for β_0.

Review: the Summation Operator

$$\sum_{i=1}^{N} c = Nc.$$

$$\sum_{i=1}^{N} cx_i = c \sum_{i=1}^{N} x_i.$$

$$\sum_{i=1}^{N} (ax_i + by_i) = a \sum_{i=1}^{N} x_i + b \sum_{i=1}^{N} y_i.$$

$$\sum_{i=1}^{N} (x_i - \overline{x}) = \sum_{i=1}^{N} x_i - N\overline{x} = 0.$$

$$\sum_{i=1}^{N} (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{N} (x_i - \overline{x}) y_i = \sum_{i=1}^{N} (x_i y_i - \overline{x}\overline{y}).$$

Proof of Unbiasedness: $\hat{\beta}_1$

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{N}(x_i - \bar{x})y_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^{N}(x_i - \bar{x})\beta_0 + \beta_1 \sum_{i=1}^{N}(x_i - \bar{x})x_i + \sum_{i=1}^{N}(x_i - \bar{x})u_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \\
&= \frac{\beta_1 \sum_{i=1}^{N}(x_i - \bar{x})^2 + \sum_{i=1}^{N}(x_i - \bar{x})u_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \beta_1 + \frac{\sum_{i=1}^{N}(x_i - \bar{x})u_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2}.
\end{aligned}$$

Consider a middle value $E(\hat{\beta}_1|x_1, x_2, ..., x_N)$:

$$\begin{aligned}
E(\hat{\beta}_1|x_1, x_2, ..., x_N) &= \beta_1 + \frac{\sum_{i=1}^{N}(x_i - \bar{x})E(u_i|x_1, x_2, ..., x_N)}{\sum(x_i - \bar{x})^2} \\
&= \beta_1 + \frac{\sum_{i=1}^{N}(x_i - \bar{x})E(u|x)}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \beta_1.
\end{aligned}$$

Then use the law of iterated expectations:

$$\begin{aligned} E(\hat{\beta}_1) &= E[E(\hat{\beta}_1 | x_1, x_2, ..., x_N)] \\ &= E(\beta_1) = \beta_1. \end{aligned}$$

- Note here we first fix $(x_1, x_2, ..., x_N)$, and view them as constants while taking the expectation, and then calculate the expectation for u.
- Random sampling assumption makes sure that for any u_i, its expectation is the same as u.
- Zero conditional mean assumption makes sure $E(u|x) = 0$.

Proof of Unbiasedness: $\hat{\beta}_0$

Calculate the average of $y_i = \beta_0 + \beta_1 x_i + u_i$:

$$\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{u}.$$

Define $\bar{u} = \frac{1}{N}\sum_{i=1}^{N} u_i$. Plug in the formula for $\hat{\beta}_0$:

$$\begin{aligned} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} &= \beta_0 + \beta_1 \bar{x} + \bar{u} - \hat{\beta}_1 \bar{x} \\ &= \beta_0 + (\beta_1 - \hat{\beta}_1)\bar{x} + \bar{u}. \end{aligned}$$

Take the expectation:

$$\begin{aligned} E(\hat{\beta}_0) &= \beta_0 + E[(\beta_1 - \hat{\beta}_1)\bar{x}] + E(\bar{u}) \\ &= \beta_0 + E[\bar{x}E(\beta_1 - \hat{\beta}_1 | x)] + \frac{1}{N}\sum_{i=1}^{N} E(u_i) \\ &= \beta_0. \end{aligned}$$

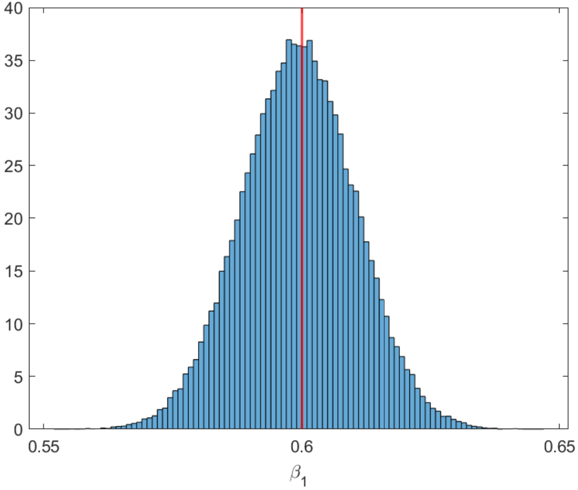
The random sampling assumption makes sure that $E(u_i) = E(u) = 0$.

The Interpretation of the Unbiasedness

Which of the following statements is correct?

1. The unbiasedness of the OLS estimator means the estimates calculated using the sample are the same as β in the population.
2. The unbiasedness of the OLS estimator guarantees that the estimates calculated using the sample are close to β in the population.
3. Though the OLS estimator is unbiased, it is still possible that the estimates calculated using the sample are very different from β in the population.

Variance of the OLS Estimator



Variance of the OLS Estimator

To derive the variance, we add another assumption.

Homoskedasticity

The error u has the same variance given any value of the explanatory variable. In other words,

$$Var(u|x) = \sigma^2.$$

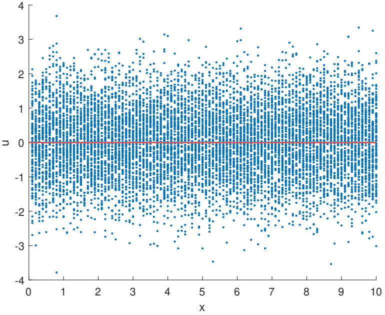


Figure: Homoskedasticity

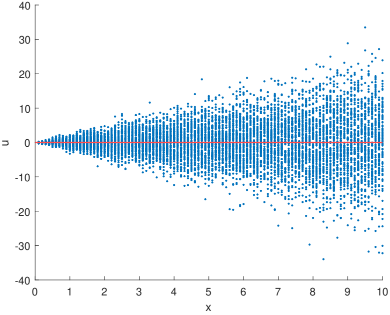


Figure: Heteroskedasticity

Sampling Variances of the OLS Estimators

Under assumptions SLR.1 through SLR.5,

$$Var(\hat{\beta}_1|x) = \frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \frac{\sigma^2}{SST_x},$$

$$Var(\hat{\beta}_0|x) = \frac{\sigma^2 \sum_{i=1}^{N} x_i^2}{N\sum_{i=1}^{N}(x_i - \bar{x})^2}.$$

- $Var(\hat{\beta}_1|x)$ is larger when σ^2 is large.
- $Var(\hat{\beta}_1|x)$ is smaller when x has more variation.
- We will prove the formula of $Var(\hat{\beta}_1|x)$ and leave $Var(\hat{\beta}_0|x)$ for after class exercise.

Review: Variance

Let X and X_i denote a random variable and a denote a constant.

$$Var(X + a) = Var(X).$$

$$Var(aX) = a^2 Var(X).$$

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y).$$

$$Var(\sum_{i=1}^{N} X_i) = \sum_{i,j=1}^{N} Cov(X_i, X_j) = \sum_{i=1}^{N} Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j).$$

Proof: $Var(\hat{\beta}_1|x)$

Using the formula when proving the unbiasedness of $\hat{\beta}_1$:

$$\begin{aligned}\hat{\beta}_1 &= \beta_1 + \frac{\sum_{i=1}^N (x_i - \bar{x}) u_i}{\sum_{i=1}^N (x_i - \bar{x})^2} \\ &= \beta_1 + \frac{\sum_{i=1}^N d_i u_i}{\sum_{i=1}^N d_i^2}.\end{aligned}$$

Note we define $d_i \equiv x_i - \bar{x}$. Then we have

$$\begin{aligned}Var(\hat{\beta}_1|x) &= Var(\beta_1 + \frac{\sum_{i=1}^N d_i u_i}{\sum_{i=1}^N d_i^2}|x) \\ &= \frac{\sum_{i=1}^N d_i^2 \, Var(u_i|x)}{(\sum_{i=1}^N d_i^2)^2} + \frac{\sum_{i\neq j} d_i d_j cov(u_i, u_j)}{(\sum_{i=1}^N d_i^2)^2} \\ &= \frac{\sum_{i=1}^N d_i^2 \sigma^2}{(\sum_{i=1}^N d_i^2)^2} = \frac{\sigma^2}{\sum_{i=1}^N (x_i - \bar{x})^2}.\end{aligned}$$

Estimating σ^2

σ^2 is unknown, so we need to estimate it using the sample:

$$s^2 \equiv \hat{\sigma}^2 = \frac{1}{N-2}\sum_{i=1}^{N}\hat{u}_i^2$$

- When calculating OLS, the residual $\hat{u}_i$ satisfy two conditions:$\sum_{i=1}^{N}\hat{u}_i = 0$ and $\sum_{i=1}^{N} x_i\hat{u}_i = 0$.
- So $\hat{u}_i$ has $N-2$ degrees of freedom while u_i has N degrees of freedom.
- By adjusting the degree of freedom, we can get an unbiased estimator of σ^2. (Proof not required.)

Summary

- How to interpret a simple regression model?
  - descriptive vs causal vs forecasting
  - change units of measurement
  - $log(y)$ or $log(x)$
- Estimating the model
  - method of moments
  - ordinary least square
- Properties of the OLS estimator
  - goodness of fit
  - five assumptions of the OLS estimator
  - expectation and variance

Answers to Quiz

- Property (1), (2), and (4) are conditions we used in estimation.
- Property (3): Plug in $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$:

$$\begin{aligned}\sum_{i=1}^{N} \hat{y}_i \hat{u}_i &= \sum_{i=1}^{N} (\hat{\beta}_0 + \hat{\beta}_1 x_i) \hat{u}_i \\ &= \hat{\beta}_0 \sum_{i=1}^{N} \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^{N} (x_i \hat{u}_i) \\ &= 0.\end{aligned}$$

- Answer to page 22: 3.
- Answer to page 25: 4.
- Answer to page 40: 3.

▸ Return

Proof: $Var(\hat{\beta}_0)$

Next, consider $Var(\hat{\beta}_0)$. First, note that:

$$\begin{aligned}
cov(\hat{\beta}_1, \bar{u}|x) &= cov(\beta_1 + \frac{\sum_{i=1}^N d_i u_i}{\sum_{i=1}^N d_i^2}, \frac{1}{N}\sum_{i=1}^N u_i|x) \\
&= \frac{1}{N\sum_{i=1}^N d_i^2}\sum_{i=1}^N d_i cov(u_i, u_i|x) \\
&= \frac{\sigma^2}{N\sum_{i=1}^N d_i^2}\sum_{i=1}^N d_i \\
&= 0.
\end{aligned}$$

Next, since $\hat{\beta}_0 = \bar{y} - \hat{\beta_1}\bar{x}$:

$$\hat{\beta}_0 = \beta_0 + \beta_1\bar{x} + \bar{u} - \hat{\beta}_1\bar{x}.$$

So

$$\begin{aligned} Var(\hat{\beta}_0|x) &= Var(\beta_0 + \beta_1\bar{x} + \bar{u} - \hat{\beta}_1\bar{x}|x) \\ &= Var(\bar{u}|x) + \bar{x}^2\,Var(\hat{\beta}_1|x) - 2\bar{x}cov(\hat{\beta}_1, \bar{u}|x) \\ &= \frac{1}{N}\sigma^2 + \frac{\bar{x}^2\sigma^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \\ &= \frac{\sigma^2 N^{-1}\sum_{i=1}^N x_i^2}{\sum_{i=1}^N (x_i - \bar{x})^2}. \end{aligned}$$