Chapter 6: Ordinary Least Squares Estimation Procedure –
The Properties

Chapter 6 Outline

- Clint’s Assignment: Assess the Effect of Studying on Quiz Scores
- Review
  - Regression Model
  - Ordinary Least Squares (OLS) Estimation Procedure
  - The Estimates, $b_{Const}$ and $b_x$, Are Random Variables
- Strategy: General Properties and a Specific Application
  - Review: Assessing Clint’s Opinion Poll Results
  - Preview: Assessing Professor Lord’s Quiz Results
- Standard Ordinary Least Squares (OLS) Premises
- New Equation for the Ordinary Least Squares (OLS) Coefficient Estimate
- General Properties: Describing the Coefficient Estimate’s Probability Distribution
  - Mean (Center) of the Coefficient Estimate’s Probability Distribution
  - Variance (Spread) of the Coefficient Estimate’s Probability Distribution
- Estimation Procedures and the Estimate’s Probability Distribution
  - Importance of the Mean (Center)
  - Importance of the Variance (Spread)
- Reliability of the Coefficient Estimate
  - Variance (Spread) of the Error Term’s Probability Distribution
  - Sample Size: Number of Observations
  - Range of the Explanatory Variable
  - Reliability Summary
- Best Linear Unbiased Estimation Procedure (BLUE)

Chapter 6 Prep Questions

1. Run the Distribution of Coefficient Estimates simulation in the Econometrics Lab by clicking the following link:

   [Link to MIT-Lab 6P.1 goes here.]

After completing the lab, fill in the following blanks:

| Numerical Value Of Coefficient | Your Calculations | Simulation’s Calculations |
Repetition | Estimate | Mean | Variance | Mean | Variance
---|---|---|---|---|---
1 |   |   |   |   |   
2 |   |   |   |   |   
3 |   |   |   |   |   

NB: You must click the “Next Problem” button to get to the simulation’s problem 1.

2. Review the arithmetic of means:
   a. Mean of a constant times a variable: Mean[c\(x\)] =
   b. Mean of a constant plus a variable: Mean[c + \(x\)] =
   c. Mean of the sum of two variables: Mean[\(x + y\)] =

3. Review the arithmetic of variances:
   a. Variance of a constant times a variable: Var[c\(x\)] =
   b. Variance of the sum of a variable and a constant: Var[c + \(x\)] =
   c. Variance of the sum of two variables: Var[\(x + y\)] =
   d. Variance of the sum of two independent variables: Var[\(x + y\)] =

4. Consider an estimate’s probability distribution:
   a. Why is the mean (center) of the probability distribution important? Explain.
   b. Why is the variance (spread) of the probability distribution important? Explain.

Clint’s Assignment: Assess the Effect of Studying on Quiz Scores
Clint’s assignment is to assess the theory that additional studying increases quiz scores. To do so he must use data from Professor Lord’s first quiz, the number of minutes studied and the quiz score for each of the three students in the course:

<table>
<thead>
<tr>
<th>Student</th>
<th>Minutes Studied ((x))</th>
<th>Quiz Score ((y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>66</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>87</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>90</td>
</tr>
</tbody>
</table>

Table 6.1: First Quiz Results

**Project:** Use data from Professor Lord’s first quiz to assess the effect of studying on quiz scores.

**Review**

**Regression Model**

Clint uses the following regression model to complete his assignment:
\[ y_t = \beta_{\text{Const}} + \beta_x x_t + e_t \]

where

- \( y_t \) = quiz score of student \( t \)
- \( x_t \) = minutes studied by student \( t \)
- \( e_t \) = error term for student \( t \)

The Parameters: \( \beta_{\text{Const}} \) and \( \beta_x \)

\( \beta_{\text{Const}} \) and \( \beta_x \) are the parameters of the model. Let us review their interpretation:

- \( \beta_{\text{Const}} \) reflects the number of points Professor Lord gives students just for showing up.
- \( \beta_x \) reflects the number of additional points earned for each additional minute of studying.

The Error Term

The error term, \( e_t \), plays a crucial role in the model. The error term represents random influences. The mean of the error term’s probability distribution for each student equals 0:

\[ \text{Mean}[e_1] = 0 \quad \text{Mean}[e_2] = 0 \quad \text{Mean}[e_3] = 0 \]

Consequently, the error terms have no systematic effect on quiz scores. Sometimes the error term will be positive and sometimes it will be negative, but after many, many quizzes each student’s error terms will average out to 0. When the probability distribution of the error term is symmetric, the chances that a student will score better than “usual” on one quiz equal the chances that the student will do worse than “usual.”

**Ordinary Least Squares (OLS) Estimation Procedure**

As a consequence of the error terms (random influences), we can never determine the actual values of \( \beta_{\text{Const}} \) and \( \beta_x \); consequently, Clint has no choice but to estimate the values. The ordinary least squares (OLS) estimation procedure is the most commonly used procedure for doing this:

\[ b_x = \frac{\sum_{t=1}^{T} (y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \]

\[ b_{\text{Const}} = \bar{y} - b_x \bar{x} \]

Using the results of the first quiz, Clint estimates the values of the coefficient and constant:

<table>
<thead>
<tr>
<th>Student</th>
<th>( x )</th>
<th>( y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>87</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>90</td>
</tr>
</tbody>
</table>

**Ordinary Least Squares (OLS) Estimates:** \( E_{y_t} = 63 + 1.2x \)

\( b_{\text{Const}} \) = Estimated points for showing up = 63

\( b_x \) = Estimated points for each minute studied = 1.2
\[
\bar{x} = \frac{x_1 + x_2 + x_3}{3} = \frac{5 + 15 + 25}{3} = \frac{45}{3} = 15
\]
\[
\bar{y} = \frac{y_1 + y_2 + y_3}{3} = \frac{66 + 87 + 90}{3} = \frac{243}{3} = 81
\]

<table>
<thead>
<tr>
<th>Student</th>
<th>(y_i)</th>
<th>(\bar{y})</th>
<th>(y_i - \bar{y})</th>
<th>(x_i)</th>
<th>(\bar{x})</th>
<th>(x_i - \bar{x})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>66</td>
<td>81</td>
<td>-15</td>
<td>5</td>
<td>15</td>
<td>-10</td>
</tr>
<tr>
<td>2</td>
<td>87</td>
<td>81</td>
<td>6</td>
<td>15</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>90</td>
<td>81</td>
<td>9</td>
<td>25</td>
<td>15</td>
<td>10</td>
</tr>
</tbody>
</table>

\[
\sum_{i=1}^{T} (y_i - \bar{y})(x_i - \bar{x}) = 240
\]
\[
\sum_{i=1}^{T} (x_i - \bar{x})^2 = 200
\]
\[
b_x = \frac{\sum_{i=1}^{T} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{T} (x_i - \bar{x})^2} = \frac{240}{200} = \frac{6}{5} = 1.2
\]
\[
b_{Const} = \bar{y} - b_x \bar{x} = 81 - \frac{6}{5} \times 15 = 63
\]

**The Estimates, \(b_{Const}\) and \(b_x\), Are Random Variables**

In the previous chapter, we used the Econometrics Lab to show that the estimates for the constant and coefficient, \(b_{Const}\) and \(b_x\), are random variables. As a consequence of the error terms (random influences), we could not determine the numerical value of the estimates for the constant and coefficient, \(b_{Const}\) and \(b_x\), before we conduct the experiment, even if we knew the actual values of the constant and coefficient, \(\beta_{Const}\) and \(\beta_x\). Furthermore, we can never expect the estimates to equal the actual values. Consequently, we must assess the reliability of the estimates. We shall focus on the coefficient estimate:

**Estimate Reliability:** How reliable is the coefficient estimate calculated from the results of the first quiz? That is, how confident can Clint be that the coefficient estimate, 1.2, will be close to the actual value of the coefficient?

**Strategy: General Properties and a Specific Application**

**Review: Assessing Clint’s Opinion Poll Results**

Clint faced a similar problem when he polled a sample of the student population to estimate the fraction of students supporting him. 12 of the 16 randomly selected students polled, 75 percent, supported Clint thereby suggesting that he was leading. But we then observed that it was possible for this result to occur even if
the election was actually a tossup. In view of this, we asked how confident Clint should be in the results of his single poll. To address this issue, we turned to the general properties of polling procedures to assess the reliability of the estimate Clint obtained from his single poll:

<table>
<thead>
<tr>
<th>General Properties</th>
<th>versus</th>
<th>One Specific Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clint’s Estimation Procedure: Calculate the fraction of the 16 randomly selected students supporting Clint</td>
<td></td>
<td>Apply the polling procedure once to Clint’s sample of the 16 randomly selected students:</td>
</tr>
<tr>
<td>Mean[EstFrac] = p = ActFrac = Actual fraction of the population supporting Clint</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var[EstFrac] = ( \frac{p(1-p)}{T} = \frac{p(1-p)}{16} ) where T = Sample Size</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean and variance describe the center and spread of the estimate’s probability distribution. While we could not determine the numerical value of the estimated fraction, EstFrac, before the poll was conducted, we could describe its probability distribution. Using algebra, we derived the general equations for the mean and variance of the estimated fraction’s, EstFrac’s, probability distribution. Then, we checked our algebra with a simulation by exploiting the relative frequency interpretation of probability: after many, many repetitions, the distribution of the numerical values mirrors the probability distribution for one repetition.

What can we deduce before the poll is conducted?

General properties of the polling procedure described by EstFrac’s probability distribution.
Probability distribution is described by its mean (center) and variance (spread).

Use algebra to derive the equations for the probability distribution’s mean and variance.

\[
\text{Mean}[\text{EstFrac}] = p \\
\text{Var}[\text{EstFrac}] = \frac{p(1-p)}{T}
\]

Check the algebra with a simulation by exploiting the relative frequency interpretation of probability.

The estimated fraction’s probability distribution allowed us to assess the reliability of Clint’s poll.

**Preview: Assessing Professor Lord’s Quiz Results**

Using the ordinary least squares (OLS) estimation procedure we estimated the value of the coefficient to be 1.2. This estimate is based on a single quiz. The fact that the coefficient estimate is positive suggests that additional studying increases quiz scores. But how confident can we be that the coefficient estimate is close to the actual value? To address the reliability issue we shall focus on the general properties of the ordinary least squares (OLS) estimation procedure:

**General Properties**

- **OLS Estimation Procedure:**
  - Estimate \( \beta_{Const} \) and \( \beta_x \) by finding the \( b_{Const} \) and \( b_x \) that minimize the sum of squared residuals

**One Specific Application**

- Model:
  \[
y = \beta_{Const} + \beta_x x + e
\]

- OLS equations:
  \[
b_x = \frac{\sum_{i=1}^{T} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{T} (x_i - \bar{x})^2} \\
b_{Const} = \bar{y} - b_x \bar{x}
\]

**Before experiment**

Random Variable: Probability Distribution

**After experiment**

Estimate: Numerical Value

- \( b_x = \frac{240}{200} = \frac{6}{5} = 1.2 \)
- \( b_{Const} = 81 - \frac{6}{5} \times 15 = 63 \)
Mean and variance describe the center and spread of the estimate’s probability distribution. While we cannot determine the numerical value of the coefficient estimate before the quiz is given, we can describe its probability distribution. The probability distribution tells us how likely it is for the coefficient estimate based on a single quiz to equal each of the possible values. Using algebra, we shall derive the general equations for the mean and variance of the coefficient estimate’s probability distribution. Then, we will check our algebra with a simulation by exploiting the relative frequency interpretation of probability: after many, many repetitions, the distribution of the numerical values mirrors the probability distribution for one repetition.

What can we deduce before the poll is conducted?

General properties of the OLS estimation procedure described by the coefficient estimate’s probability distribution.

Probability distribution is described by its mean (center) and variance (spread).

Use algebra to derive the equations for the probability distribution’s mean and variance.

Check the algebra with a simulation by exploiting the relative frequency interpretation of probability.

The coefficient estimate’s probability distribution will allow us to assess the reliability of the coefficient estimate calculated from Professor Lord’s quiz.

**Standard Ordinary Least Squares (OLS) Regression Premises**

To derive the equations for the mean and variance of the coefficient estimate’s probability distribution, we shall apply the standard ordinary least squares (OLS) regression premises. As we mentioned Chapter 5, these premises make the analysis as straightforward as possible. In later chapters, we will relax these premises to study more general cases. In other words, we shall start with the most straightforward case and then move on to more complex ones later.

- **Error Term Equal Variance Premise:** The variance of the error term’s probability distribution for each observation is the same; all the variances equal \( \text{Var}[e] \):
  \[
  \text{Var}[e_1] = \text{Var}[e_2] = \ldots = \text{Var}[e_T] = \text{Var}[e]
  \]
• **Error Term/Error Term Independence Premise:** The error terms are independent: \( \text{Cov}[e_i, e_j] = 0 \).

  Knowing the value of the error term from one observation does not help us predict the value of the error term for any other observation.

• **Explanatory Variable/Error Term Independence Premise:** The explanatory variables, the \( x_t \)'s, and the error terms, the \( e_t \)'s, are not correlated.

  Knowing the value of an observation’s explanatory variable does not help us predict the value of that observation’s error term.

To keep the algebra manageable, we shall assume that the explanatory variables are constants in the derivations that follow. This assumption allows us to apply the arithmetic of means and variances easily. While this simplifies our algebraic manipulations, it does not affect the validity of our conclusions.

**New Equation for the Ordinary Least Squares (OLS) Coefficient Estimate**

In Chapter 5, we derived an equation that expressed the OLS coefficient estimate in terms of the \( x \)'s and \( y \)'s:

\[
b_x = \frac{\sum_{t=1}^{T} (y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^{T} (x_t - \bar{x})^2}
\]

It is advantageous to use a different equation to derive the equations for the mean and variance of the coefficient estimate’s probability distribution, however; we shall use an equivalent equation that expresses the coefficient estimate in terms of the \( x \)'s, \( e \)'s, and \( \beta \) rather than in terms of the \( x \)'s and \( y \)'s:

\[
b_x = \beta_x + \frac{\sum_{t=1}^{T} (x_t - \bar{x})e_t}{\sum_{t=1}^{T} (x_t - \bar{x})^2}
\]

To keep the notation as straightforward as possible, we shall focus on the 3 observation case. The logic for the general case is identical to the logic for the 3 observation case:

\[
b_x = \beta_x + \frac{(x_1 - \bar{x})e_1 + (x_2 - \bar{x})e_2 + (x_3 - \bar{x})e_3}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2}
\]

**General Properties: Describing the Coefficient Estimate’s Probability Distribution**

*Mean (Center) of the Coefficient Estimate’s Probability Distribution*
To calculate the mean of $b_x$’s probability distribution, review the arithmetic of means:

- Mean of a constant times a variable: $\text{Mean}[cx] = c \text{Mean}[x]$;
- Mean of a constant plus a variable: $\text{Mean}[c + x] = c + \text{Mean}[x]$;
- Mean of the sum of two variables: $\text{Mean}[x + y] = \text{Mean}[x] + \text{Mean}[y]$;

and recall that the error term represents random influences:

- The mean of each error term’s probability distribution is 0:
  $$\text{Mean}[e_1] = \text{Mean}[e_2] = \text{Mean}[e_3] = 0.$$ 

Now, we apply algebra to the new equation for the coefficient estimate, $b_x$:

$$\text{Mean}[b_x] = \text{Mean}\left[\beta_x + \frac{(x_1 - \bar{x})e_1 + (x_2 - \bar{x})e_2 + (x_3 - \bar{x})e_3}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2}\right]$$

Applying $\text{Mean}[c + x] = c + \text{Mean}[x]$:

$$= \beta_x + \text{Mean}\left[\frac{(x_1 - \bar{x})e_1 + (x_2 - \bar{x})e_2 + (x_3 - \bar{x})e_3}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2}\right]$$

Rewriting the fraction as a product:

$$= \beta_x + \text{Mean}\left[\frac{1}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2}\right]\left((x_1 - \bar{x})e_1 + (x_2 - \bar{x})e_2 + (x_3 - \bar{x})e_3\right)$$

Applying $\text{Mean}[cx] = c\text{Mean}[x]$:

$$= \beta_x + \frac{1}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2}\text{Mean}\left((x_1 - \bar{x})e_1 + (x_2 - \bar{x})e_2 + (x_3 - \bar{x})e_3\right)$$

Applying $\text{Mean}[x + y] = \text{Mean}[x] + \text{Mean}[y]$:

$$= \beta_x + \frac{1}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2}\text{Mean}\left((x_1 - \bar{x})e_1\right) + \text{Mean}\left((x_2 - \bar{x})e_2\right) + \text{Mean}\left((x_3 - \bar{x})e_3\right)$$

Applying $\text{Mean}[cx] = c\text{Mean}[x]$:

$$= \beta_x + \frac{1}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2}\left((x_1 - \bar{x})\text{Mean}[e_1] + (x_2 - \bar{x})\text{Mean}[e_2] + (x_3 - \bar{x})\text{Mean}[e_3]\right)$$

Since $\text{Mean}[e_1] = \text{Mean}[e_2] = \text{Mean}[e_3] = 0$:

$$= \beta_x$$

So, we have shown that

$$\text{Mean}[b_x] = \beta_x$$
Consequently, the ordinary least squares (OLS) estimation procedure for the value of the coefficient is unbiased. In any one repetition of the experiment, the mean (center) of the probability distribution equals the actual value of the coefficient. The estimation procedure does not systematically overestimate or underestimate the actual coefficient value, \( \beta_x \). If the probability distribution is symmetric, the chances that the estimate calculated from one quiz will be too high equal the chances that it will be too low.

*Econometrics Lab 6.1: Checking the Equation for the Mean*

We can use the Distribution of Coefficient Estimates simulation in our Econometrics Lab to replicate the quiz many, many times. But in reality, Clint only has information from one quiz, the first quiz. How then can a simulation be useful? The relative frequency interpretation of probability provides the answer. The relative frequency interpretation of probability tells us that the distribution of the numerical values after many, many repetitions of the experiments mirrors the probability distribution of one repetition. Consequently, repeating the experiment many, many times reveals the probability distribution for the one quiz:

\[
\text{Distribution of the Numerical Values} \downarrow \rightarrow \text{Probability Distribution}
\]

We can use the simulation to check the algebra we used to derive the equation for the mean of the coefficient estimate’s probability distribution:
Mean\(\mathbf{b}_x = \beta_x\)

If our algebra is correct, the mean (average) of the estimated coefficient values should equal the actual value of the coefficient, \(\beta_x\), after many, many repetitions.

Recall that a simulation allows us to do something that we cannot do in the real world. In the simulation, we can specify the actual values of the constant and coefficient, \(\beta_{\text{Const}}\) and \(\beta_x\). The default setting for the actual coefficient value is 2. Be certain that the Pause checkbox is checked. Click Start. Record the numerical value of the coefficient estimate for the first repetition. Click Continue to simulate the second quiz. Record the value of the coefficient estimate for the second repetition and calculate the mean and variance of the numerical estimates for the first two repetitions. Note that your calculations agree with those provided by the simulation. Click Continue again to simulate the third quiz. Calculate the mean and variance of the numerical estimates for the first three repetitions. Once again, note that your calculations and the simulation’s calculations agree. Continue to click Continue until you are convinced that the simulation is calculating the mean and variance of the numerical values for the coefficient estimates correctly.

Now, clear the Pause checkbox and click Continue. The simulation no longer pauses after each repetition. After many, many repetitions, click Stop.
**Question:** What does the mean (average) of the coefficient estimates equal?

**Answer:** It equals about 2.0.

This lends support to the equation for the mean of the coefficient estimate’s probability distribution that we just derived. Now, change the actual coefficient value from 2 to 4. Click Start and then after many, many repetitions, click Stop. What does the mean (average) of the estimates equal? Next, change the actual coefficient value to 6 and repeat the process.

![Probability Distribution](image)

**Figure 6.3: Histogram of Coefficient Value Estimates**

<table>
<thead>
<tr>
<th>Actual $\beta_x$</th>
<th>Prob Dist $\text{Mean}[b_x]$</th>
<th>Simulation Repetitions</th>
<th>Equation: Mean of Coef Estimate</th>
<th>Simulation: Mean (Average) of Estimated Coef Values, $b_x$, from the Experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>&gt;1,000,000</td>
<td>$\approx$2.0</td>
<td>$\approx$2.0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>&gt;1,000,000</td>
<td>$\approx$4.0</td>
<td>$\approx$4.0</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>&gt;1,000,000</td>
<td>$\approx$6.0</td>
<td>$\approx$6.0</td>
</tr>
</tbody>
</table>

Table 6.2: Distribution of Coefficient Estimate Simulation Results

**Conclusion:** In all cases, the mean (average) of the estimates for the coefficient value equals the actual value of the coefficient after many, many repetitions. The simulations confirm our algebra. The estimation procedure does not systematically underestimate or overestimate the actual value of the coefficient. The ordinary least squares (OLS) estimation procedure for the coefficient value is unbiased.
Variance (Spread) of the Coefficient Estimate’s Probability Distribution

Now, we turn our attention to the variance of the coefficient estimate’s probability distribution. To derive the equation for the variance, begin by reviewing the arithmetic of variances:

- Variance of a constant times a variable: \( \text{Var}(cx) = c^2 \text{Var}(x) \).
- Variance of the sum of a variable and a constant: \( \text{Var}(c + x) = \text{Var}(x) \).
- Variance of the sum of two independent variables: \( \text{Var}(x + y) = \text{Var}(x) + \text{Var}(y) \).

Focus on the first two standard ordinary least squares (OLS) premises:

- Error Term Equal Variance Premise: \( \text{Var}(e_1) = \text{Var}(e_2) = \text{Var}(e_3) = \text{Var}(e) \).
- Error Term/Error Term Independence Premise: The error terms are independent; that is, \( \text{Cov}(e_t, e_j) = 0 \).

Recall the new equation for \( b_x \):

\[
b_x = \beta_x + \frac{(x_1 - \bar{x})e_1 + (x_2 - \bar{x})e_2 + (x_3 - \bar{x})e_3}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2}
\]

Therefore,

\[
\text{Var}(b_x) = \text{Var}\left[ \beta_x + \frac{(x_1 - \bar{x})e_1 + (x_2 - \bar{x})e_2 + (x_3 - \bar{x})e_3}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2} \right]
\]

Applying \( \text{Var}(c + x) = \text{Var}(x) \)

\[
= \text{Var}\left[ \frac{(x_1 - \bar{x})e_1 + (x_2 - \bar{x})e_2 + (x_3 - \bar{x})e_3}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2} \right]
\]

Rewriting the fraction as a product

\[
= \text{Var}\left[ \frac{1}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2} \left( (x_1 - \bar{x})e_1 + (x_2 - \bar{x})e_2 + (x_3 - \bar{x})e_3 \right) \right]
\]

Applying \( \text{Var}(c x) = c^2 \text{Var}(x) \)

\[
= \frac{1}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2} \left( \text{Var}(e_1) + \text{Var}(e_2) + \text{Var}(e_3) \right)
\]

Error Term/Error Term Independence Premise:

\( \text{Var}(x + y) = \text{Var}(x) + \text{Var}(y) \)

\[
= \frac{1}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2} \left[ \text{Var}(e_1) + \text{Var}(e_2) + \text{Var}(e_3) \right]
\]

Applying \( \text{Var}(c x) = c^2 \text{Var}(x) \)

\[
= \frac{1}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2} \left[ (x_1 - \bar{x})^2 \text{Var}(e_1) + (x_2 - \bar{x})^2 \text{Var}(e_2) + (x_3 - \bar{x})^2 \text{Var}(e_3) \right]
\]

Error Term Equal Variance Premise:

\( \text{Var}(e_1) = \text{Var}(e_2) = \text{Var}(e_3) = \text{Var}(e) \)
We can generalize this: 

$$\text{Var}[b_x] = \frac{\text{Var}[e]}{\sum_{i=1}^{r} (x_i - \bar{x})^2}$$

The variance of the coefficient estimate’s probability distribution equals the variance of the error term’s probability distribution divided by the sum of squared $x$ deviations.

**Econometrics Lab 6.2: Checking the Equation for the Variance**

We shall now use the Distribution of Coefficient Estimates simulation to check the equation that we just derived for the variance of the coefficient estimate’s probability distribution.

The simulation automatically spreads the $x$ values uniformly between 0 and 30. We shall continue to consider three observations; accordingly, the $x$ values are 5,
15, and 25. To convince yourself of this, be certain that the Pause checkbox is checked. Click Start and then Continue a few times to observe that the values of \( x \) are always 5, 15, and 25.

Next, recall the equation we just derived for the variance of the coefficient estimate’s probability distribution:

\[
\text{Var}[\hat{\beta}_j] = \frac{\text{Var}[e]}{\sum_{i=1}^{T}(x_i - \bar{x})^2} = \frac{\text{Var}[e]}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2}
\]

By default, the variance of the error term probability distribution is 500; therefore, the numerator equals 500. Let us turn our attention to the denominator, the sum of squared \( x \) deviations. We have just observed that the \( x \) values are 5, 15, and 25. Their mean is 15 and their sum of squared deviations from the mean is 200:

\[
\bar{x} = \frac{x_1 + x_2 + x_3}{3} = \frac{5 + 15 + 25}{3} = \frac{45}{3} = 15
\]

\[
\begin{array}{c|ccc|c}
\text{Student} & x_i & \bar{x} & x_i - \bar{x} & (x_i - \bar{x})^2 \\
1 & 5 & 15 & -10 & (-10)^2 = 100 \\
2 & 15 & 15 & 0 & (0)^2 = 0 \\
3 & 25 & 15 & 10 & (10)^2 = 100 \\
\end{array}
\]

That is,

\[
(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 = 200
\]

When the variance of the error term’s probability distribution equals 500 and the sum of squared \( x \) deviations equals 200, the variance of the coefficient estimate’s probability distribution equals 2.50:

\[
\text{Var}[\hat{\beta}_j] = \frac{\text{Var}[e]}{\sum_{i=1}^{T}(x_i - \bar{x})^2} = \frac{500}{200} = 2.50
\]

To show that the simulation confirms this, be certain that the Pause checkbox is cleared and click Continue. After many, many repetitions click Stop. Indeed, after many, many repetitions of the experiment the variance of the numerical values is about 2.50. The simulation confirms the equation we derived for the variance of the coefficient estimate’s probability distribution.

**Estimation Procedures and the Estimate’s Probability Distribution: Importance of the Mean (Center) and Variance (Spread)**

Let us review what we learned about estimation procedures when we studied Clint’s opinion poll in Chapter 3:
• **Importance of the Probability Distribution’s Mean:** Formally, an estimation procedure is unbiased whenever the mean (center) of the estimate’s probability distribution equals the actual value. The relative frequency interpretation of probability provides intuition: If the experiment were repeated many, many times the average of the numerical values of the estimates will equal the actual value. An unbiased estimation procedure does not systematically underestimate or overestimate the actual value. If the probability distribution is symmetric, the chances that the estimate calculated from one repetition of the experiment will be too high equal the chances the estimate will be too low.

![Probability Distribution of Estimates – Importance of the Mean](image)

• **Importance of the Probability Distribution’s Variance:** When the estimation procedure is unbiased, the variance of the estimate’s probability distribution’s variance (spread) reveals the estimate’s reliability; the variance tells us how likely it is that the numerical value of the estimate calculated from one repetition of the experiment will be close to the actual value:
When the estimation procedure is unbiased, the variance of the estimate’s probability distribution determines reliability.

- As the variance decreases, the probability distribution becomes more tightly cropped around the actual value making it more likely for the estimate to be close to the actual value.
- On the other hand, as the variance increases, the probability distribution becomes less tightly cropped around the actual value making it less likely for the estimate to be close to the actual value.

Reliability of the Coefficient Estimate

We shall focus on the variance of the coefficient estimate’s probability distribution to explain what influences its reliability. We shall consider three factors:

- Variance of the error term’s probability distribution.
- Sample size.
- Range of the x’s.

Estimate Reliability and the Variance of the Error Term’s Probability Distribution

What is our intuition here? The error term represents the random influences. It is the error term that introduces uncertainty into the mix. As the variance of the error term’s probability distribution increases, uncertainty increases; consequently, the available information becomes less reliable. As the variance of the error term’s probability distribution increases, we would expect the coefficient estimate to become less reliable. On the other hand, as the variance of the error term’s
probability distribution decreases, the available information becomes more reliable, and we would expect the coefficient estimate to become more reliable.

To justify this intuition, recall the equation for the variance of the coefficient estimate’s probability distribution:

\[
\text{Var}[b_x] = \frac{\text{Var}[e]}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\text{Var}[e]}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2}
\]

The variance of the coefficient estimate’s probability distribution is directly proportional to the variance of the error term’s probability distribution. 

\textit{Econometrics Lab 6.3: Variance of the Error Term’s Probability Distribution}

We shall use the Distribution of Coefficient Estimates simulation to confirm the role played by the variance of the error term’s probability distribution. To do so, check the From-To checkbox. Two lists now appear: a From list and a To list. Initially, 1.0 is selected in the From list and 3.0 in the To list. Consequently, the simulation will report the percent of repetitions in which the coefficient estimate falls between 1.0 and 3.0. Since the default value for the actual coefficient, \(\beta_x\), equals 2.0, the simulation reports on the percent of repetitions in which the coefficient estimate falls within 1.0 of the actual value. The simulation reports the percent of repetitions in which the coefficient estimate is “close to” the actual value where “close to” is considered to be within 1.0.

By default, the variance of the error term’s probability distribution equals 500 and the sample size equals 3. Recall that the sum of the squared x deviations equals 200 and therefore the variance of the coefficient estimate’s probability distribution equals 2.50:

\[
\text{Var}[b_x] = \frac{\text{Var}[e]}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\text{Var}[e]}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2} = \frac{500}{200} = 2.50
\]

Be certain that the Pause checkbox is cleared. Click Start and then after many, many repetitions, click Stop. As Table 6.3 reports, the coefficient estimate lies within 1.0 of the actual coefficient value in 47.3 percent of the repetitions.

Now, reduce the variance of the error term’s probability distribution from 500 to 50. The variance of the coefficient estimate’s probability distribution now equals .25:

\[
\text{Var}[b_x] = \frac{\text{Var}[e]}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\text{Var}[e]}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2} = \frac{50}{200} = \frac{1}{4} = .25
\]
Click Start and then after many, many repetitions, click Stop. The histogram of the coefficient estimates is now more closely cropped around the actual value, 2.0. The percent of repetitions in which the coefficient estimate lies within 1.0 of the actual coefficient value rises from 47.3 percent to 95.5 percent.

<table>
<thead>
<tr>
<th>Actual Values</th>
<th>Sample Size</th>
<th>x</th>
<th>x</th>
<th>Probability Distribution Equations:</th>
<th>Estimated Coefficient Values, b_x</th>
<th>Percent Between 1.0 and 3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>β_x</td>
<td>Var[e]</td>
<td>Min</td>
<td>Max</td>
<td>Mean[b_x]</td>
<td>Var[b_x]</td>
<td>Mean (Average)</td>
</tr>
<tr>
<td>2</td>
<td>500</td>
<td>3</td>
<td>30</td>
<td>2.0</td>
<td>2.50</td>
<td>≈2.0</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>3</td>
<td>30</td>
<td>2.0</td>
<td>.25</td>
<td>≈2.0</td>
</tr>
</tbody>
</table>

Table 6.3: Distribution of Coefficient Estimate Simulation Reliability Results

Why is this important? The variance measures the spread of the probability distribution. This is important when the estimation procedure is unbiased. As the variance decreases, the probability distribution becomes more closely cropped around the actual coefficient value and the chances that the coefficient estimate obtained from one quiz will lie close to the actual value increases. The simulation confirms this; after many, many repetitions the percent of repetitions in which the coefficient estimate lies between 1.0 and 3.0 increases from 47.3 percent to 95.5 percent. Consequently, as the error term’s variance decreases, we can expect the estimate from one quiz to be more reliable. As the variance of the error term’s probability distribution decreases, the estimate is more likely to be “close to” the actual value. This is consistent with our intuition, is it not?

**Estimate Reliability and the Sample Size**

Next, we shall investigate the effect of the sample size, the number of observations, used to calculate the estimate. Increase the sample size from 3 to 5. What does our intuition suggest? As we increase the number of observations, we will have more information. With more information the estimate should become more reliable; that is, with more information the variance of the coefficient estimate’s probability distribution should decrease. Using the equation, let us now calculate the variance of the coefficient estimate’s probability distribution when there are 5 observations. With 5 observations the x values are spread uniformly at 3, 9, 15, 21, and 27; the mean (average) of the x’s, \( \bar{x} \), equals 15 and the sum of the squared x deviations equals 360:

\[
\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{3 + 9 + 15 + 21 + 27}{5} = \frac{75}{5} = 15
\]

Student

\[
\chi_i = \frac{x_i - \bar{x}}{} \quad (x_i - \bar{x})^2
\]
Applying the equation for the value of the coefficient estimate’s probability distribution:

\[
\text{Var}[b_x] = \frac{\text{Var}[e]}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\text{Var}[e]}{50}
\]

\[
= \frac{50}{(3-15)^2 + (9-15)^2 + (15-15)^2 + (21-15)^2 + (27-15)^2}
\]

\[
= \frac{50}{0 + 36 + 0 + 36 + 36} = \frac{50}{144} = \frac{50}{360} = .1388 \approx .14
\]

The variance of the coefficient estimate’s probability distribution falls from .25 to .14. The smaller variance suggests that the coefficient estimate will be more reliable.

*Econometrics Lab 6.4: Sample Size*

Are our intuition and calculations supported by the simulation? In fact, the answer is yes.

Note that the sample size has increased from 3 to 5. Click Start and then after many, many repetitions click Stop:

<table>
<thead>
<tr>
<th>Actual Values</th>
<th>Sample Size</th>
<th>Probability Distribution Equations: Estimated Coefficient Values, (b_x)</th>
<th>Simulations: Percent Between 1.0 and 3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_x)</td>
<td>(\text{Var}[e])</td>
<td>Sample Size</td>
<td>(x)</td>
</tr>
<tr>
<td>2</td>
<td>500</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 6.4: Distribution of Coefficient Estimate Simulation Reliability Results

After many, many repetitions the percent of repetitions in which the coefficient estimate lies between 1.0 and 3.0 increases from 95.5 percent to 99.3 percent. As the sample size increases, we can expect the estimate from one quiz to be more reliable. As the sample size increases, the estimate is more likely to be “close to” the actual value.

Estimate Reliability and the Range of x’s

Let us again begin by appealing to our intuition. As the range of x’s becomes smaller, we are basing our estimates on less variation in the x’s, less diversity; accordingly, we are basing our estimates on less information. As the range becomes smaller, the estimate should become less reliable and consequently, the variance of the coefficient estimate’s probability distribution should increase. To confirm this, increase the minimum value of x from 0 to 10 and decrease the maximum value from 30 to 20. The five x values are now spread uniformly between 10 and 20 at 11, 13, 15, 17, and 19; the mean (average) of the x’s, \( \bar{x} \), equals 15 and the sum of the squared x deviations equals 40:

\[
\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{11 + 13 + 15 + 17 + 19}{5} = 15
\]

<table>
<thead>
<tr>
<th>Student</th>
<th>( x_i )</th>
<th>( \bar{x} )</th>
<th>( x_i - \bar{x} )</th>
<th>( (x_i - \bar{x})^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>15</td>
<td>-4</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>15</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>15</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>15</td>
<td>4</td>
<td>16</td>
</tr>
</tbody>
</table>

Sum: \( \sum_{i=1}^{5} (x_i - \bar{x})^2 = 40 \)

Applying the equation for the value of the coefficient estimate’s probability distribution:

\[
\text{Var}[b] = \frac{\text{Var}[e]}{\sum_{i=1}^{5} (x_i - \bar{x})^2} = \frac{\text{Var}[e]}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2}
\]

\[
= \frac{50}{(11 - 15)^2 + (13 - 15)^2 + (15 - 15)^2 + (17 - 15)^2 + (19 - 15)^2}
\]

\[
= \frac{50}{(-4)^2 + (2)^2 + (0)^2 + (2)^2 + (4)^2} = \frac{50}{16 + 4 + 0 + 4 + 16} = \frac{50}{40} = \frac{5}{4} = 1.25
\]
The variance of the coefficient estimate’s probability distribution increases from about .14 to 1.25.

Econometrics Lab 6.5: Range of x’s

Our next lab confirms our intuition.

After changing the minimum value of x to 10 and the maximum value to 20, click the Start and then after many, many repetitions click Stop.

<table>
<thead>
<tr>
<th>Actual Values</th>
<th>Sample Size</th>
<th>x Min</th>
<th>x Max</th>
<th>Probability Distribution Equations:</th>
<th>Estimated Coefficient Values, $b_x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_x$</td>
<td>Var[e]</td>
<td>Mean $[b_x]$</td>
<td>Var $[b_x]$</td>
<td>Mean (Average)</td>
<td>Variance</td>
</tr>
<tr>
<td>2</td>
<td>500</td>
<td>2.0</td>
<td>2.50</td>
<td>≈2.0</td>
<td>≈2.50</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>2.0</td>
<td>.25</td>
<td>≈2.0</td>
<td>≈.25</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>2.0</td>
<td>.14</td>
<td>≈2.0</td>
<td>≈.14</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>2.0</td>
<td>1.25</td>
<td>≈2.0</td>
<td>≈1.25</td>
</tr>
</tbody>
</table>

Table 6.5: Distribution of Coefficient Estimate Simulation Reliability Results

After many, many repetitions the percent of repetitions in which the coefficient estimate lies between 1.0 and 3.0 decreases from 99.3 percent to 62.8 percent. An estimate from one repetition will be less reliable. As the range of the x’s decreases, the estimate is less likely to be “close to” the actual value.

Reliability Summary

Our simulation results illustrate relationships between information, the variance of the coefficient estimate’s probability distribution, and the reliability of an estimate:

More and/or more reliable information. \[\downarrow\]

Variance of coefficient estimate’s probability distribution smaller. \[\downarrow\]

Estimate more reliable; more likely the estimate is “close to” the actual value.

Less and/or less reliable information. \[\downarrow\]

Variance of coefficient estimate’s probability distribution larger. \[\downarrow\]

Estimate less reliable; less likely the estimate is “close to” the actual value.

Best Linear Unbiased Estimation Procedure (BLUE)
In Chapter 5 we introduced the mechanics of the ordinary least squares (OLS) estimation procedure and in this chapter we analyzed the procedure’s properties. Why have we devoted so much attention to this particular estimation procedure? The reason is straightforward. When the standard ordinary least squares (OLS) premises are satisfied, no other linear estimation procedure produces more reliable estimates. In other words, the ordinary least squares (OLS) estimation procedure is the best linear unbiased estimation procedure (BLUE). Let us now explain this more carefully.

If an estimation procedure is the best linear unbiased estimation procedure (BLUE), it must exhibit three properties:

1. The estimate must be a linear function of the dependent variable, the $y_i$’s.
2. The estimation procedure must be unbiased; that is, the mean of the estimate’s probability distribution must equal the actual value.
3. No other linear unbiased estimation procedure can be more reliable; that is, the variance of the estimate’s probability distribution when using any other linear unbiased estimation procedure cannot be less than the variance when the best linear unbiased estimation procedure is used.

The Gauss-Markov theorem proves that the ordinary least squares (OLS) estimation procedure is the best linear unbiased estimation procedure. We shall illustrate the theorem by describing two other linear unbiased estimation procedures that while unbiased, are not as reliable as the ordinary least squares (OLS) estimation procedure. Please note that while we would never use either of these estimation procedures to do serious analysis, they are useful pedagogical tools. They allow us to illustrate what we mean by the best linear unbiased estimation procedure.

Two New Estimation Procedures

We shall now consider the Any Two and the Min-Max estimation procedures:

- **Any Two Estimation Procedure**: Choose any two points on the scatter diagram; draw a straight line through the points. The coefficient estimate equals the slope of this line.
Econometrics Lab 6.6: Comparing the Ordinary Least Squares (OLS), Any Two, and Min-Max Estimation Procedures

We shall now use the BLUE simulation in our Econometrics Lab to justify our emphasis on the ordinary least squares (OLS) estimation procedure.

[Link to MIT-Lab 6.6 goes here.]
By default, the sample size equals 5 and the variance of the error term’s probability distribution equals 500. The From-To values are specified as 1.0 and 3.0:

<table>
<thead>
<tr>
<th>Estimation Procedure</th>
<th>Actual Values</th>
<th>Sample Size = 5</th>
<th>Simulations: Estimated Coefficient Values, $b_x$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_x$</td>
<td>Var[e]</td>
<td>Mean (Average)</td>
</tr>
<tr>
<td>OLS</td>
<td>2.0</td>
<td>500</td>
<td>$\approx 2.0$</td>
</tr>
<tr>
<td>Any Two</td>
<td>2.0</td>
<td>50</td>
<td>$\approx 2.0$</td>
</tr>
<tr>
<td>Min-Max</td>
<td>2.0</td>
<td>50</td>
<td>$\approx 2.0$</td>
</tr>
</tbody>
</table>

Table 6.6: BLUE Simulation Results

Initially, the ordinary least squares (OLS) estimation procedure is specified. Be certain that the Pause checkbox is cleared. Click Start and then after many, many repetitions click Stop. For the OLS estimation procedure, the average of the estimated coefficient values equals about 2.0 and the variance 1.4. 60.4 percent of the estimates lie with 1.0 of the actual value. Next, select the Any Two estimation procedure instead of OLS. Click Start and then after many, many repetitions click Stop. For the Any Two estimation procedure, the average of the estimated coefficient values equals about 2.0 and the variance 14.0; 29.0 percent of the estimates lie within 1.0 of the actual value. Repeat the process one last time after selecting the Min-Max estimation procedure; the average equals about 2.0 and the variance 1.7; 55.2 percent of the estimates lie with 1.0 of the actual value.

Let us summarize:

- In all three cases, the average of the coefficient estimates equal 2.0, the actual value; after many, many repetitions the mean (average) of the estimates equals the actual value. Consequently, all three estimation procedures for the coefficient value appear to be unbiased.
- The variance of the coefficient estimate’s probability distribution is smallest when the ordinary least squares (OLS) estimation procedure is used. Consequently, the ordinary least squares (OLS) estimation procedure produces the most reliable estimates.

What we have just observed can be generalized. When the standard ordinary least squares (OLS) regression premises are met, the ordinary least squares (OLS) estimation procedure is the best linear unbiased estimation procedure because no other linear unbiased estimation procedure produces estimates that are more reliable.
Appendix 6.1: New Equation for the OLS Coefficient Estimate

Begin by recalling the expression for $b_x$ that we derived previously in Chapter 5:

$$b_x = \frac{\sum_{t=1}^{T} (y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^{T} (x_t - \bar{x})^2}$$

$b_x$ is expressed in terms of the $x$’s and $y$’s. We wish to express $b_x$ in terms of the $x$’s, $e$’s, and $\beta_x$.

**Strategy:** Focus on the numerator of the expression for $b_x$ and substitute for the $y$’s to express the numerator in terms of the $x$’s, $e$’s, and $\beta_x$. As we shall shortly show, once we do this, our goal will be achieved.

We begin with the numerator, $\sum_{t=1}^{T} (y_t - \bar{y})(x_t - \bar{x})$, and substitute $\beta_{\text{Const}} + \beta_x x_t + e_t$ for $y_t$:

$$\sum_{t=1}^{T} (y_t - \bar{y})(x_t - \bar{x}) = \sum_{t=1}^{T} (\beta_{\text{Const}} + \beta_x x_t + e_t)(x_t - \bar{x})$$

Rearranging terms.

$$= \sum_{t=1}^{T} (\beta_{\text{Const}} - \bar{y} + \beta_x x_t + e_t)(x_t - \bar{x})$$

Adding and subtracting $\beta_x \bar{x}$.

$$= \sum_{t=1}^{T} (\beta_{\text{Const}} + \beta_x \bar{x} - \bar{y} + \beta_x x_t - \beta_x \bar{x} + e_t)(x_t - \bar{x})$$

Simplifying.

$$= \sum_{t=1}^{T} [(\beta_{\text{Const}} + \beta_x \bar{x} - \bar{y}) + \beta_x (x_t - \bar{x}) + e_t](x_t - \bar{x})$$

Splitting the summation into three parts.

$$= \sum_{t=1}^{T} (\beta_{\text{Const}} + \beta_x \bar{x} - \bar{y})(x_t - \bar{x}) + \sum_{t=1}^{T} \beta_x (x_t - \bar{x})^2 + \sum_{t=1}^{T} (x_t - \bar{x})e_t$$

Simplifying the first and second terms.

$$= (\beta_{\text{Const}} + \beta_x \bar{x} - \bar{y})\sum_{t=1}^{T} (x_t - \bar{x}) + \beta_x \sum_{t=1}^{T} (x_t - \bar{x})^2 + \sum_{t=1}^{T} (x_t - \bar{x})e_t$$
Now, focus on the first term, \((\beta_{\text{Const}} + \beta_x \bar{x} - \bar{y})\sum_{i=1}^{T} (x_i - \bar{x})\). What does \(\sum_{i=1}^{T} (x_i - \bar{x})\) equal?

\[
\sum_{i=1}^{T} (x_i - \bar{x}) = \sum_{i=1}^{T} x_i - \sum_{i=1}^{T} \bar{x} = \sum_{i=1}^{T} x_i - T\bar{x}
\]

Replacing \(\sum_{i=1}^{T} \bar{x}\) with \(T\bar{x}\).

\[
= \sum_{i=1}^{T} x_i - T\bar{x}
\]

Since \(\bar{x} = \frac{\sum_{i=1}^{T} x_i}{T}\).

\[
= \sum_{i=1}^{T} x_i - T \cdot \frac{\sum_{i=1}^{T} x_i}{T} = \sum_{i=1}^{T} x_i - \sum_{i=1}^{T} x_i = 0
\]

Next, return to the expression for the numerator, \(\sum_{i=1}^{T} (y_i - \bar{y})(x_i - \bar{x})\):

\[
\sum_{i=1}^{T} (y_i - \bar{y})(x_i - \bar{x}) = (\beta_{\text{Const}} + \beta_x \bar{x} - \bar{y})\sum_{i=1}^{T} (x_i - \bar{x}) + \beta_x \sum_{i=1}^{T} (x_i - \bar{x})^2 + \sum_{i=1}^{T} (x_i - \bar{x})e_i
\]

\[
\downarrow \sum_{i=1}^{T} (x_i - \bar{x}) = 0
\]

\[
= 0 + \beta_x \sum_{i=1}^{T} (x_i - \bar{x})^2 + \sum_{i=1}^{T} (x_i - \bar{x})e_i
\]

Therefore,

\[
\sum_{i=1}^{T} (y_i - \bar{y})(x_i - \bar{x}) = \beta_x \sum_{i=1}^{T} (x_i - \bar{x})^2 + \sum_{i=1}^{T} (x_i - \bar{x})e_i
\]

Last, apply this to the equation we derived for \(b_x\) in Chapter 5:
\[ b_x = \frac{\sum_{i=1}^{T} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{T} (x_i - \bar{x})^2} \]

Substituting for the numerator.

\[ \beta_x \frac{\sum_{i=1}^{T} (x_i - \bar{x})^2 + \sum_{i=1}^{T} (x_i - \bar{x})e_i}{\sum_{i=1}^{T} (x_i - \bar{x})^2} \]

Splitting the single fraction into two.

\[ = \frac{\beta_x \sum_{i=1}^{T} (x_i - \bar{x})^2}{\sum_{i=1}^{T} (x_i - \bar{x})^2} + \frac{\sum_{i=1}^{T} (x_i - \bar{x})e_i}{\sum_{i=1}^{T} (x_i - \bar{x})^2} \]

Simplifying the first term.

\[ = \beta_x + \frac{\sum_{i=1}^{T} (x_i - \bar{x})e_i}{\sum_{i=1}^{T} (x_i - \bar{x})^2} \]

We have now expressed \( b_x \) in terms of the \( x \)'s, \( e \)'s, and \( \beta_x \).
Appendix 6.2: Gauss-Markov Theorem

Gauss-Markov Theorem: When the standard ordinary least squares (OLS) premises are satisfied, the ordinary least squared (OLS) estimation procedure is the best linear unbiased estimation procedure.

Proof: Let

\[ b_{x}^{OLS} = \text{Ordinary least squares (OLS) estimate} \]

First, note that \( b_{x}^{OLS} \) is a linear function of the \( y \)'s:

\[
 b_{x}^{OLS} = \frac{\sum_{i=1}^{T} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{T} (x_i - \bar{x})^2}
\]

Let \( w_i^{OLS} \) equal the ordinary least squares (OLS) “linear weights”; more specifically,

\[
 b_{x}^{OLS} = \sum_{i=1}^{T} w_i^{OLS} (y_i - \bar{y}) \quad \text{where} \quad w_i^{OLS} = \frac{(x_i - \bar{x})}{\sum_{i=1}^{T} (x_i - \bar{x})^2}
\]

Now, let us derive two properties of \( w_i^{OLS} \):

- \( \sum_{i=1}^{T} w_i^{OLS} = 0 \)
- \( \sum_{i=1}^{T} w_i^{OLS} (x_i - \bar{x}) = 1 \)

First, \( \sum_{i=1}^{T} w_i^{OLS} = 0 \):

\[
 \sum_{i=1}^{T} w_i^{OLS} = \frac{\sum_{i=1}^{T} (x_i - \bar{x})}{\sum_{i=1}^{T} (x_i - \bar{x})^2}
\]

Placing the summation in the numerator.

\[
 = \frac{\sum_{i=1}^{T} (x_i - \bar{x})}{\sum_{i=1}^{T} (x_i - \bar{x})^2}
\]

Splitting the summations in the numerator.
\[
\sum_{i=1}^{T} x_i - T \bar{x} \\
= \frac{\sum_{i=1}^{T} x_i - T \bar{x}}{\sum_{i=1}^{T} (x_i - \bar{x})^2}
\]

Since there are \(T\) \(\bar{x}\) terms.

\[
\sum_{i=1}^{T} x_i - T \bar{x} \\
= \frac{\sum_{i=1}^{T} x_i}{\sum_{i=1}^{T} (x_i - \bar{x})^2}
\]

Since \(\bar{x} = \frac{\sum_{i=1}^{T} x_i}{T}\).

\[
\sum_{i=1}^{T} x_i - T \bar{x} \\
= \frac{\sum_{i=1}^{T} x_i - \sum_{i=1}^{T} x_i}{\sum_{i=1}^{T} (x_i - \bar{x})^2}
\]

Simplifying.

\[
\sum_{i=1}^{T} x_i - \sum_{i=1}^{T} x_i \\
= \frac{\sum_{i=1}^{T} (x_i - \bar{x})}{\sum_{i=1}^{T} (x_i - \bar{x})^2}
\]

Since the numerator equals 0.

Second, \(\sum_{i=1}^{T} w_i^{OLS} = 0\):

\[
\sum_{i=1}^{T} w_i^{OLS} (x_i - \bar{x}) = \sum_{i=1}^{T} \frac{(x_i - \bar{x})}{\sum_{i=1}^{T} (x_i - \bar{x})^2} (x_i - \bar{x})
\]

Simplifying.

\[
= \frac{\sum_{i=1}^{T} (x_i - \bar{x})^2}{\sum_{i=1}^{T} (x_i - \bar{x})^2}
\]

Placing the summation in the numerator.
\[
\frac{\sum_{i=1}^{T} (x_i - \bar{x})^2}{\sum_{i=1}^{T} (x_i - \bar{x})^2} = 1
\]

Since the numerator and denominator are equal.

Next, consider a new linear estimation procedure whose weights are \( w_i^{OLS} + w_i \).

Only when each \( w_i \) equals 0 will this procedure to identical to the ordinary least squares (OLS) estimation procedure. Let \( \hat{b}_x \) equal the coefficient estimate calculated using this new linear estimation procedure:

\[
\hat{b}_x = \sum_{i=1}^{T} (w_i^{OLS} + w_i) y_i
\]

Now, let us perform a little algebra:

\[
\hat{b}_x = \sum_{i=1}^{T} (w_i^{OLS} + w_i) y_i
\]

Since \( y_i = \beta_{\text{Const}} + \beta_x x_i + e_i \)

\[
= \sum_{i=1}^{T} (w_i^{OLS} + w_i)(\beta_{\text{Const}} + \beta_x x_i + e_i)
\]

Multiplying through.

\[
= \sum_{i=1}^{T} (w_i^{OLS} + w_i) \beta_{\text{Const}} + \sum_{i=1}^{T} (w_i^{OLS} + w_i) \beta_x x_i + \sum_{i=1}^{T} (w_i^{OLS} + w_i) e_i
\]

Factoring out \( \beta_{\text{Const}} \) from the first term and \( \beta_x \) from the second.

\[
= \beta_{\text{Const}} \sum_{i=1}^{T} (w_i^{OLS} + w_i) + \beta_x \sum_{i=1}^{T} (w_i^{OLS} + w_i) x_i + \sum_{i=1}^{T} (w_i^{OLS} + w_i) e_i
\]

Again, simplifying the first two terms.

\[
= \beta_{\text{Const}} \sum_{i=1}^{T} w_i^{OLS} + \beta_{\text{Const}} \sum_{i=1}^{T} w_i + \beta_x \sum_{i=1}^{T} w_i^{OLS} x_i + \beta_x \sum_{i=1}^{T} w_i x_i + \sum_{i=1}^{T} (w_i^{OLS} + w_i) e_i
\]

Since \( \sum_{i=1}^{T} w_i^{OLS} x_i = 0 \) and \( \sum_{i=1}^{T} w_i^{OLS} x_i = 1 \).
\[
0 + \beta_{\text{Const}} \sum_{t=1}^{T} w_i + \beta_x \sum_{t=1}^{T} w_i x_i + \sum_{t=1}^{T} (w_i^\text{OLS} + w_i) e_i
\]

Therefore,
\[
b_x' = \beta_{\text{Const}} \sum_{t=1}^{T} w_i + \beta_x \sum_{t=1}^{T} w_i x_i + \sum_{t=1}^{T} (w_i^\text{OLS} + w_i) e_i
\]

Now, calculate the mean of the new estimate’s probability distribution, Mean\[b_x']:
\[
\text{Mean}[b_x'] = \text{Mean}\left[\beta_{\text{Const}} \sum_{t=1}^{T} w_i + \beta_x \sum_{t=1}^{T} w_i x_i + \sum_{t=1}^{T} (w_i^\text{OLS} + w_i) e_i\right]
\]

Since Mean\[c + x] = c + \text{Mean}[x]
\[
= \beta_{\text{Const}} \sum_{t=1}^{T} w_i + \beta_x \sum_{t=1}^{T} w_i x_i + \text{Mean}\left[\sum_{t=1}^{T} (w_i^\text{OLS} + w_i) e_i\right]
\]

Focusing on the last term, Mean\[cx] = c\text{Mean}[x]
\[
= \beta_{\text{Const}} \sum_{t=1}^{T} w_i + \beta_x \sum_{t=1}^{T} w_i x_i + \sum_{t=1}^{T} (w_i^\text{OLS} + w_i) \text{Mean}[e_i]
\]

Focusing on the last term, since the error terms represents random influences, \text{Mean}[e_i] = 0
\[
= \beta_{\text{Const}} \sum_{t=1}^{T} w_i + \beta_x \sum_{t=1}^{T} w_i x_i
\]

The new linear estimation procedure must be unbiased:
\[
\text{Mean}[b_x'] = \beta_x
\]

Therefore,
\[
\sum_{t=1}^{T} w_i = 0 \quad \text{and} \quad \sum_{t=1}^{T} w_i x_i = 0
\]

Next, calculate the variance of \[b_x']:
\[
\text{Var}[b_x'] = \text{Var}\left[\beta_{\text{Const}} \sum_{t=1}^{T} w_i + \beta_x \sum_{t=1}^{T} w_i x_i + \sum_{t=1}^{T} (w_i^\text{OLS} + w_i) e_i\right]
\]

Since \text{Var}[c + x] = \text{Var}[x],
\[
= \text{Var}\left[\sum_{t=1}^{T} (w_i^\text{OLS} + w_i) e_i\right]
\]
Since the error terms are independent, covariances equal 0: \[ \text{Var}[x + y] = \text{Var}[x] + \text{Var}[y]. \]

\[
= \sum_{t=1}^{T} \text{Var}\left[ (w_t^{\text{OLS}} + w_t^i) e_t \right]
\]

Since \( \text{Var}[cx] = c^2 \text{Var}[x] \).

\[
= \sum_{t=1}^{T} (w_t^{\text{OLS}} + w_t^i)^2 \text{Var}[e_t]
\]

The variance of each error term’s probability distribution is identical, \( \text{Var}[e] \).

\[
= \sum_{t=1}^{T} (w_t^{\text{OLS}} + w_t^i)^2 \text{Var}[e]
\]

Factoring out \( \text{Var}[e] \).

\[
= \text{Var}[e] \sum_{t=1}^{T} (w_t^{\text{OLS}} + w_t^i)^2
\]

Expanding the squared terms.

\[
= \text{Var}[e] \left( \sum_{t=1}^{T} (w_t^{\text{OLS}})^2 + 2w_t^{\text{OLS}} w_t^i + (w_t^i)^2 \right)
\]

Splitting up the summation.

\[
= \text{Var}[e] \left( \sum_{t=1}^{T} (w_t^{\text{OLS}})^2 + 2 \sum_{t=1}^{T} w_t^{\text{OLS}} w_t^i + \sum_{t=1}^{T} (w_t^i)^2 \right)
\]

Now, focus on the cross product terms, \( \sum_{t=1}^{T} w_t^{\text{OLS}} w_t^i \):

\[
\sum_{t=1}^{T} w_t^{\text{OLS}} w_t^i = \sum_{t=1}^{T} \frac{(x_t - \bar{x})}{\sum_{i=1}^{T} (x_t - \bar{x})^2} w_t^i
\]

Placing the summation in the numerator.

\[
= \frac{\sum_{t=1}^{T} (x_t - \bar{x}) w_t^i}{\sum_{t=1}^{T} (x_t - \bar{x})^2}
\]

Splitting the summations in the numerator.

\[
= \frac{\sum_{t=1}^{T} (x_t w_t^i - \sum_{t=1}^{T} \bar{x} w_t^i)}{\sum_{t=1}^{T} (x_t - \bar{x})^2}
\]
Factor out $\bar{x}$ from the second term in the numerator.

$$\frac{\sum_{i=1}^{T} x_i w_i - \bar{x} \sum_{i=1}^{T} w'_i}{\sum_{i=1}^{T} (x_i - \bar{x})^2}$$

Since $\sum_{i=1}^{T} x_i w_i = 0$ and $\sum_{i=1}^{T} w'_i = 0$.

$$\frac{0 - 0}{\sum_{i=1}^{T} (x_i - \bar{x})^2}$$

Since the numerator equals 0.

$$= 0$$

Therefore,

$$\text{Var}[b'_x] = \text{Var}[e]\left(\sum_{i=1}^{T} (w'_i^{OLS})^2 + 2 \sum_{i=1}^{T} w'_i^{OLS} w'_i + \sum_{i=1}^{T} (w'_i)^2\right)$$

Since $\sum_{i=1}^{T} w'_i^{OLS} w'_i = 0$.

$$= \text{Var}[e]\left(\sum_{i=1}^{T} (w'_i^{OLS})^2 + \sum_{i=1}^{T} (w'_i)^2\right)$$

The variance of the estimate’s probability distribution is minimized whenever each $w'_i$ equals 0, whenever the estimation procedure is the ordinary least squares (OLS) estimation procedure.

---

1 Appendix 6.1 appearing at the end of this chapter shows how we can derive the second equation for the coefficient estimate, $b'_x$, from the first.
2 The proof appears at the end of this chapter in Appendix 6.2.
3 To reduce potential confusion, the summation index in the denominator has been changed from $t$ to $i$. 