Chapter 4

Properties of the Least Squares Estimators

Assumptions of the Simple Linear Regression Model

SR1. $y_t = \beta_1 + \beta_2 x_t + e_t$

SR2. $E(e_t) = 0 \Leftrightarrow E[y_t] = \beta_1 + \beta_2 x_t$

SR3. $\text{var}(e_t) = \sigma^2 = \text{var}(y_t)$

SR4. $\text{cov}(e_i, e_j) = \text{cov}(y_i, y_j) = 0$

SR5. x_t is not random and takes at least two values

SR6. $e_t \sim N(0, \sigma^2) \Leftrightarrow y_t \sim N[(\beta_1 + \beta_2 x_t), \sigma^2]$ (*optional*)

4.1 The Least Squares Estimators as Random Variables

To repeat an important passage from Chapter 3, when the formulas for b_1 and b_2, given in Equation (3.3.8), are taken to be rules that are used whatever the sample data turn out to be, then b_1 and b_2 are random variables since their values depend on the random variable y whose values are not known until the sample is collected. In this context we call b_1 and b_2 the ***least squares estimators***. When actual sample values, numbers, are substituted into the formulas, we obtain numbers that are *values of random variables*. In this context we call b_1 and b_2 the ***least squares estimates***.

4.2 The Sampling Properties of the Least Squares Estimators

The means (expected values) and variances of random variables provide information about the location and spread of their probability distributions (see Chapter 2.3). As such, the means and variances of b_1 and b_2 provide information about the range of values that b_1 and b_2 are likely to take. Knowing this range is important, because our objective is to obtain estimates that are *close* to the true parameter values. Since b_1 and b_2 are random variables, they may have covariance, and this we will determine as well. These "pre-data" characteristics of b_1 and b_2 are called **sampling properties**, because the randomness of the estimators is brought on by sampling from a population.

4.2.1 The Expected Values of b_1 and b_2

- The least squares *estimator* b_2 of the slope parameter β_2, based on a sample of T observations, is

$$b_2 = \frac{T\sum x_t y_t - \sum x_t \sum y_t}{T\sum x_t^2 - \left(\sum x_t\right)^2} \tag{3.3.8a}$$

- The least squares *estimator* b_1 of the intercept parameter β_1 is

$$b_1 = \bar{y} - b_2\bar{x} \tag{3.3.8b}$$

where $\bar{y} = \sum y_t / T$ and $\bar{x} = \sum x_t / T$ are the sample means of the observations on y and x, respectively.

- We begin by rewriting the formula in Equation (3.3.8a) into the following one that is more convenient for theoretical purposes:

$$b_2 = \beta_2 + \sum w_t e_t \tag{4.2.1}$$

where w_t is a constant (non-random) given by

$$w_t = \frac{x_t - \overline{x}}{\sum (x_t - \overline{x})^2} \tag{4.2.2}$$

Since w_t is a constant, depending only on the values of x_t, we can find the expected value of b_2 using the fact that the *expected value of a sum is the sum of the expected values* (see Chapter 2.5.1):

$$E(b_2) = E\left(\beta_2 + \sum w_t e_t\right) = E(\beta_2) + \sum E(w_t e_t)$$

$$= \beta_2 + \sum w_t E(e_t) = \beta_2 \quad [\text{since } E(e_t) = 0] \tag{4.2.3}$$

When the expected value of any estimator of a parameter equals the true parameter value, then that estimator is *unbiased*. Since $E(b_2) = \beta_2$, the least squares estimator b_2 is an unbiased estimator of β_2. If many samples of size T are collected, and the formula (3.3.8a) for b_2 is used to estimate β_2, then the average value of the estimates b_2 obtained from all those samples will be β_2, *if the statistical model assumptions are correct.*

- However, if the assumptions we have made are not correct, then the least squares estimator may not be unbiased. In Equation (4.2.3) note in particular the role of the assumptions SR1 and SR2. The assumption that $E(e_t) = 0$, for each and every t, makes

$\sum w_t E(e_t) = 0$ and $E(b_2) = \beta_2$. If $E(e_t) \neq 0$, then $E(b_2) \neq \beta_2$. Recall that e_t contains, among other things, factors affecting y_t that are *omitted* from the economic model. If we have omitted anything that is important, then we would expect that $E(e_t) \neq 0$ and $E(b_2) \neq \beta_2$. Thus, having an econometric model that is correctly specified, in the sense that it includes all relevant explanatory variables, is a must in order for the least squares estimators to be unbiased.

- The unbiasedness of the estimator b_2 is an important sampling property. When sampling repeatedly from a population, the least squares estimator is "correct," on average, and this is one desirable property of an estimator. This statistical property by itself does not mean that b_2 is a good estimator of β_2, but it is part of the story. The unbiasedness property depends on having *many* samples of data from the same population. The fact that b_2 is unbiased does not imply anything about what might happen in just one sample. An individual estimate (number) b_2 may be near to, or far from β_2. Since β_2 is *never* known, we will never know, given one sample, whether our estimate is "close" to β_2 or not. The least squares estimator b_1 of β_1 is also an *unbiased* estimator, and $E(b_1) = \beta_1$.

4.2.1a The Repeated Sampling Context

- To illustrate unbiased estimation in a slightly different way, we present in Table 4.1 least squares estimates of the food expenditure model from 10 random samples of size T = 40 from the same population. Note the variability of the least squares parameter estimates from sample to sample. This sampling variation is due to the simple fact that we obtained 40 different households in each sample, and their weekly food expenditure varies randomly.

Table 4.1 Least Squares Estimates from 10 Random Samples of size T=40

n	b_1	b_2
1	51.1314	0.1442
2	61.2045	0.1286
3	40.7882	0.1417
4	80.1396	0.0886
5	31.0110	0.1669
6	54.3099	0.1086
7	69.6749	0.1003
8	71.1541	0.1009
9	18.8290	0.1758
10	36.1433	0.1626

- The property of unbiasedness is about the *average* values of b_1 and b_2 if many samples of the same size are drawn from the same population. The average value of b_1 in these 10 samples is $\bar{b}_1 = 51.43859$. The average value of b_2 is $\bar{b}_2 = 0.13182$. If we took the averages of estimates from many samples, these averages would approach the true parameter values β_1 and β_2. Unbiasedness does not say that an estimate from any one sample is close to the true parameter value, and thus we can not say that an *estimate* is unbiased. We can say that the least squares estimation procedure (or the least squares estimator) is unbiased.

4.2.1b Derivation of Equation 4.2.1

- In this section we show that Equation (4.2.1) is correct. The first step in the conversion of the formula for b_2 into Equation (4.2.1) is to use some tricks involving summation signs. The first useful fact is that

$$\begin{aligned}\sum(x_t-\bar{x})^2 &= \sum x_t^2 - 2\bar{x}\sum x_t + T\bar{x}^2 = \sum x_t^2 - 2\bar{x}\left(T\frac{1}{T}\sum x_t\right) + T\bar{x}^2 \\ &= \sum x_t^2 - 2T\bar{x}^2 + T\bar{x}^2 = \sum x_t^2 - T\bar{x}^2\end{aligned} \tag{4.2.4a}$$

Then, starting from Equation(4.2.4a),

$$\sum(x_t - \overline{x})^2 = \sum x_t^2 - T\,\overline{x}^2 = \sum x_t^2 - \overline{x}\sum x_t = \sum x_t^2 - \frac{\left(\sum x_t\right)^2}{T} \qquad (4.2.4b)$$

To obtain this result we have used the fact that $\overline{x} = \sum x_t / T$, so $\sum x_t = T\,\overline{x}$.

- The second useful fact is

$$\sum(x_t - \overline{x})(y_t - \overline{y}) = \sum x_t y_t - T\overline{x}\,\overline{y} = \sum x_t y_t - \frac{\sum x_t \sum y_t}{T} \qquad (4.2.5)$$

This result is proven in a similar manner by using Equation (4.2.4b).

- If the numerator and denominator of b_2 in Equation (3.3.8a) are divided by T, then using Equations (4.2.4) and (4.2.5) we can rewrite b_2 in *deviation from the mean form* as

$$b_2 = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2} \tag{4.2.6}$$

This formula for b_2 is one that you should remember, as we will use it time and time again in the next few chapters. Its primary advantage is its theoretical usefulness.

- The sum of any variable about its average is zero, that is,

$$\sum (x_t - \bar{x}) = 0 \tag{4.2.7}$$

- Then, the formula for b_2 becomes

$$b_2 = \frac{\sum (x_t - \overline{x})(y_t - \overline{y})}{\sum (x_t - \overline{x})^2} = \frac{\sum (x_t - \overline{x}) y_t - \overline{y} \sum (x_t - \overline{x})}{\sum (x_t - \overline{x})^2}$$

(4.2.8)

$$= \frac{\sum (x_t - \overline{x}) y_t}{\sum (x_t - \overline{x})^2} = \sum \left[\frac{(x_t - \overline{x})}{\sum (x_t - \overline{x})^2} \right] y_t = \sum w_t y_t$$

where w_t is the constant given in Equation (4.2.2).

- To obtain Equation (4.2.1), replace y_t by $y_t = \beta_1 + \beta_2 x_t + e_t$ and simplify:

$$b_2 = \sum w_t y_t = \sum w_t (\beta_1 + \beta_2 x_t + e_t) = \beta_1 \sum w_t + \beta_2 \sum w_t x_t + \sum w_t e_t \qquad (4.2.9a)$$

First, $\sum w_t = 0$, this eliminates the term $\beta_1 \sum w_t$. Secondly, $\sum w_t x_t = 1$ (by using Equation (4.2.4b)), so $\beta_2 \sum w_t x_t = \beta_2$, and (4.2.9a) simplifies to Equation (4.2.1), which is what we wanted to show.

$$b_2 = \beta_2 + \sum w_t e_t \tag{4.2.9b}$$

The term $\sum w_t = 0$, because

$$\sum w_t = \sum \left[\frac{(x_t - \bar{x})}{\sum (x_t - \bar{x})^2} \right] = \frac{1}{\sum (x_t - \bar{x})^2} \sum (x_t - \bar{x}) = 0 \quad \left(\text{using } \sum (x_t - \bar{x}) = 0 \right)$$

To show that $\sum w_t x_t = 1$ we again use $\sum (x_t - \bar{x}) = 0$. Another expression for $\sum (x_t - \bar{x})^2$ is

$$\sum(x_t-\overline{x})^2=\sum(x_t-\overline{x})(x_t-\overline{x})=\sum(x_t-\overline{x})x_t-\overline{x}\sum(x_t-\overline{x})=\sum(x_t-\overline{x})x_t$$

Consequently,

$$\sum w_t x_t=\frac{\sum(x_t-\overline{x})x_t}{\sum(x_t-\overline{x})^2}=\frac{\sum(x_t-\overline{x})x_t}{\sum(x_t-\overline{x})x_t}=1$$

4.2.2 The Variances and Covariance of b_1 and b_2

- The variance of the random variable b_2 is the average of the squared distances between the values of the random variable and its mean, which we now know is $E(b_2) = \beta_2$. The variance (Chapter 2.3.4) of b_2 is defined as

$$\text{var}(b_2) = E[b_2 - E(b_2)]^2$$

It measures the spread of the probability distribution of b_2.

- In Figure 4.1 are graphs of two possible probability distribution of b_2, $f_1(b_2)$ and $f_2(b_2)$, that have the same mean value but different variances. The probability density function $f_2(b_2)$ has a smaller variance than the probability density function $f_1(b_2)$. Given a choice, we are interested in estimator precision and would *prefer* that b_2 have the probability distribution $f_2(b_2)$ rather than $f_1(b_2)$. With the distribution $f_2(b_2)$ the probability is more concentrated around the true parameter value β_2, giving, relative to $f_1(b_2)$, a *higher* probability of getting an estimate that is *close* to β_2. Remember, getting an estimate close to β_2 is our objective.
- The variance of an estimator measures the *precision* of the estimator in the sense that it tells us how much the estimates produced by that estimator can vary from sample to sample as illustrated in Table 4.1. Consequently, we often refer to the **sampling**

variance or **sampling precision** of an estimator. The lower the variance of an estimator, the greater the sampling precision of that estimator. One estimator is more precise than another estimator if its sampling variance is less than that of the other estimator.

- If the regression model assumptions SR1-SR5 are correct (SR6 is not required), then the variances and covariance of b_1 and b_2 are:

$$\operatorname{var}(b_1) = \sigma^2 \left[\frac{\sum x_t^2}{T \sum (x_t - \bar{x})^2} \right]$$

$$\operatorname{var}(b_2) = \frac{\sigma^2}{\sum (x_t - \bar{x})^2} \tag{4.2.10}$$

$$\operatorname{cov}(b_1,\ b_2) = \sigma^2 \left[\frac{-\bar{x}}{\sum (x_t - \bar{x})^2} \right]$$

- Let us consider the factors that affect the variances and covariance in Equation (4.2.10).
  1. The variance of the random error term, σ^2, appears in each of the expressions. It reflects the dispersion of the values *y* about their mean *E*(*y*). The greater the variance σ^2, the greater is the dispersion, and the greater the uncertainty about

where the values of y fall relative to their mean $E(y)$. The information we have about β_1 and β_2 is less precise the larger is σ^2. The larger the variance term σ^2, the greater the uncertainty there is in the statistical model, and the larger the variances and covariance of the least squares estimators.

2. The sum of squares of the values of x about their sample mean, $\sum(x_t - \overline{x})^2$, appears in each of the variances and in the covariance. This expression measures how spread out about their mean are the sample values of the independent or explanatory variable x. The more they are spread out, the larger the sum of squares. The less they are spread out the smaller the sum of squares. The larger the sum of squares, $\sum(x_t - \overline{x})^2$, the smaller the variance of least squares estimators and the more precisely we can estimate the unknown parameters. The intuition behind this is demonstrated in Figure 4.2. On the right, in panel (b), is a data scatter in which the values of x are widely spread out along the x-axis. In panel (a) the data are

"bunched." The data in panel (b) do a better job of determining where the least squares line must fall, because they are more spread out along the x-axis.

3. The larger the sample size T the smaller the variances and covariance of the least squares estimators; it is better to have more sample data than less. The sample size T appears in each of the variances and covariance because each of the sums consists of T terms. Also, T appears explicitly in var(b_1). The sum of squares term $\sum(x_t - \bar{x})^2$ gets larger and larger as T increases because each of the terms in the sum is positive or zero (being zero if x happens to equal its sample mean value for an observation). Consequently, as T gets larger, both var(b_2) and cov(b_1, b_2) get smaller, since the sum of squares appears in their denominator. The sums in the numerator and denominator of var(b_1) both get larger as T gets larger and offset one another, leaving the T in the denominator as the dominant term, ensuring that var(b_1) also gets smaller as T gets larger.

4. The term Σx^2 appears in var(b_1). The larger this term is, the larger the variance of the least squares estimator b_1. Why is this so? Recall that the intercept parameter β_1 is the expected value of y, given that $x = 0$. The farther our data from $x = 0$ the more difficult it is to interpret β_1, and the more difficult it is to accurately estimate β_1. The term Σx^2 measures the distance of the data from the origin, $x = 0$. If the values of x are near zero, then Σx^2 will be small and this will reduce var(b_1). But if the values of x are large in magnitude, either positive or negative, the term Σx^2 will be large and var(b_1) will be larger.
5. The sample mean of the x-values appears in cov(b_1, b_2). The covariance increases the larger in magnitude is the sample mean $\bar{x}$, and the covariance has the sign that is opposite that of $\bar{x}$. The reasoning here can be seen from Figure 4.2. In panel (b) the least squares fitted line must pass through the point of the means. Given a fitted line through the data, imagine the effect of increasing the estimated slope, b_2. Since the line must pass through the point of the means, the effect must be to lower the

point where the line hits the vertical axis, implying a reduced intercept estimate b_1. Thus, when the sample mean is positive, as shown in Figure 4.2, there is a negative covariance between the least squares estimators of the slope and intercept.

- **Deriving the variance of b_2:**

The starting point is Equation (4.2.1).

$$\begin{aligned}
\operatorname{var}(b_2) &= \operatorname{var}\left(\beta_2 + \sum w_t e_t\right) = \operatorname{var}\left(\sum w_t e_t\right) \qquad [\text{since } \beta_2 \text{ is a constant}] \\
&= \sum w_t^2 \operatorname{var}(e_t) \qquad [\text{using } \operatorname{cov}(e_i, e_j) = 0] \\
&= \sigma^2 \sum w_t^2 \qquad [\text{using } \operatorname{var}(e_t) = \sigma^2] \\
&= \frac{\sigma^2}{\sum (x_t - \overline{x})^2} \qquad (4.2.11)
\end{aligned}$$

The very last step uses the fact that

$$\sum w_t^2 = \sum \left[\frac{(x_t - \bar{x})^2}{\left\{ \sum (x_t - \bar{x})^2 \right\}^2} \right] = \frac{1}{\sum (x_t - \bar{x})^2} \qquad (4.2.12)$$

- **Deriving the variance of b_1**:

  From Equation (3.3.8b)

$$b_1 = \bar{y} - b_2 \bar{x} = \frac{1}{T} \sum (\beta_1 + \beta_2 x_t + e_t) - b_2 \bar{x}$$

$$= \beta_1 + \beta_2 \bar{x} + \bar{e} - b_2 \bar{x} = \beta_1 - (b_2 - \beta_2)\bar{x} + \bar{e}$$

$$\Rightarrow b_1 - \beta_1 = -(b_2 - \beta_2)\bar{x} + \bar{e}$$

Since $E(b_2) = \beta_2$ and $E(\bar{e}) = 0$, it follows that

$$E(b_1) = \beta_1 - (\beta_2 - \beta_2)\bar{x} + 0 = \beta_1$$

Then

$$\begin{aligned}
\operatorname{var}(b_1) &= E[(b_1 - \beta_1)^2] = E[(-(b_2 - \beta_2)\bar{x} + \bar{e})^2] \\
&= \bar{x}^2 E[(b_2 - \beta_2)^2] + E[\bar{e}^2] - 2\bar{x}E[(b_2 - \beta_2)\bar{e}] \\
&= \bar{x}^2 \operatorname{var}(b_2) + E[\bar{e}^2] - 2\bar{x}E[(b_2 - \beta_2)\bar{e}]
\end{aligned}$$

Now

$$\mathrm{var}(b_2) = \frac{\sigma^2}{\sum (x_t - \bar{x})^2}$$

and

$$E(\bar{e}^2) = E[(\frac{1}{T}\sum e_t)^2] = \frac{1}{T^2}E[(\sum e_t)^2]$$

$$= \frac{1}{T^2}E[\sum e_t^2 + \text{cross-product terms in } e_i e_j]$$

$$= \frac{1}{T^2}\{E[\sum e_t^2] + E[\text{cross-product terms in } e_i e_j]\}$$

$$= \frac{1}{T^2}(T\sigma^2) = \frac{1}{T}\sigma^2$$

(Note: $\mathrm{var}(e_t) = E[(e_t - E(e_t))^2] = E[e_t^2]$)

Finally,

$$E[(b_2 - \beta_2)\bar{e}] = E[(\sum w_t e_t)(\frac{1}{T}\sum e_t)] \quad \text{(from Equation (4.2.9b))}$$

$$= E[\frac{1}{T}(\sum w_t e_t^2 + \text{cross-product terms in } e_i e_j)]$$

$$= \frac{1}{T}\sum w_t E(e_t^2)$$

$$= \frac{1}{T}\sigma^2 \sum w_t$$

$$= 0 \quad (\sum w_t = 0)$$

Therefore,

$$\text{var}(b_1) = \bar{x}^2 E[(b_2 - \beta_2)^2] + E[\bar{e}^2]$$

$$= \frac{\sigma^2 \bar{x}^2}{\sum (x_t - \bar{x})^2} + \frac{\sigma^2}{T}$$

$$= \sigma^2 [\frac{T\bar{x}^2 + \sum (x_t - \bar{x})^2}{T \sum (x_t - \bar{x})^2}]$$

$$= \sigma^2 [\frac{T\bar{x}^2 + \sum x_t^2 - T\bar{x}^2}{T \sum (x_t - \bar{x})^2}]$$

$$= \sigma^2 [\frac{\sum x_t^2}{T \sum (x_t - \bar{x})^2}]$$

- **Deriving the covariance of b_1 and b_2**:

$$\begin{aligned}
\operatorname{cov}(b_1,\ b_2) &= E[(b_1-\beta_1)(b_2-\beta_2)] \\
&= E[(\bar{e}-(b_2-\beta_2)\bar{x})(b_2-\beta_2)] \\
&= E[\bar{e}(b_2-\beta_2)] - \bar{x}E[(b_2-\beta_2)^2] \\
&= -\frac{\sigma^2\bar{x}}{\sum(x-\bar{x})^2} \qquad \text{(Note: } E(b_2-\beta_2)=0\text{)}
\end{aligned}$$

4.2.3 Linear Estimators

- The least squares estimator b_2 is a weighted sum of the observations y_t, $b_2 = \sum w_t y_t$ [see Equation (4.2.8)]. In mathematics weighted sums like this are called linear combinations of the y_t; consequently, statisticians call estimators like b_2, that are linear combinations of an observable random variable, **linear estimators**.

- Putting together what we know so far, we can describe b_2 as a *linear*, *unbiased estimator* of β_2, with a variance given in Equation (4.2.10). Similarly, b_1 can be described as a linear, unbiased estimator of β_1, with a variance given in Equation (4.2.10).

4.3 The Gauss-Markov Theorem

Gauss-Markov Theorem: Under the assumptions SR1-SR5 of the linear regression model the estimators b_1 and b_2 have the *smallest variance of all linear and unbiased estimators* of β_1 and β_2. They are the **Best Linear Unbiased Estimators** (BLUE) of β_1 and β_2.

Let us clarify what the Gauss-Markov theorem does, and does not, say.

1. The estimators b_1 and b_2 are "best" when compared to *similar* estimators, those that are *linear* and *unbiased*. The Theorem does not say that b_1 and b_2 are the best of all *possible* estimators.
2. The estimators b_1 and b_2 are best within their class because they have the minimum variance.

3. In order for the Gauss-Markov Theorem to hold, the assumptions (SR1-SR5) must be true. If any of the assumptions 1-5 are *not* true, then b_1 and b_2 are *not* the best linear unbiased estimators of β_1 and β_2.
4. The Gauss-Markov Theorem does *not* depend on the assumption of normality (assumption SR6).
5. In the simple linear regression model, if we want to use a linear and unbiased estimator, then we have to do no more searching. The estimators b_1 and b_2 are the ones to use.
6. The Gauss-Markov theorem applies to the least squares estimators. It *does not* apply to the least squares *estimates* from a single sample.

Proof of the Gauss-Markov Theorem:

- Let $b_2^* = \sum k_t y_t$ (where the k_t are constants) be any other linear estimator of β_2. Suppose that $k_t = w_t + c_t$, where c_t is another constant and w_t is given in Equation

(4.2.2). While this is tricky, it is legal, since for any k_t that someone might choose we can find c_t.

- Into this new estimator substitute y_t and simplify, using the properties of w_t, $\sum w_t = 0$ and $\sum w_t x_t = 1$. Then it can be shown that

$$\begin{aligned} b_2^* &= \sum k_t y_t = \sum (w_t + c_t) y_t = \sum (w_t + c_t)(\beta_1 + \beta_2 x_t + e_t) \\ &= \sum (w_t + c_t)\beta_1 + \sum (w_t + c_t)\beta_2 x_t + \sum (w_t + c_t) e_t \\ &= \beta_1 \sum w_t + \beta_1 \sum c_t + \beta_2 \sum w_t x_t + \beta_2 \sum c_t x_t + \sum (w_t + c_t) e_t \\ &= \beta_1 \sum c_t + \beta_2 + \beta_2 \sum c_t x_t + \sum (w_t + c_t) e_t \end{aligned} \tag{4.3.1}$$

- Take the mathematical expectation of the last line in Equation (4.3.1), using the properties of expectation (see Chapter 2.5.1) and the assumption that $E(e_t) = 0$.

$$\begin{aligned} E(b_2^*) &= \beta_1 \sum c_t + \beta_2 + \beta_2 \sum c_t x_t + \sum (w_t + c_t) E(e_t) \\ &= \beta_1 \sum c_t + \beta_2 + \beta_2 \sum c_t x_t \end{aligned} \tag{4.3.2}$$

- In order for the linear estimator $b_2^* = \sum k_t y_t$ to be unbiased, it must be true that

$$\sum c_t = 0 \text{ and } \sum c_t x_t = 0 \tag{4.3.3}$$

- These conditions must hold in order for $b_2^* = \sum k_t y_t$ to be in the class of *linear* and *unbiased estimators.* So we will assume the conditions (4.3.3) hold and use them to simplify expression (4.3.1):

$$b_2^* = \sum k_t y_t = \beta_2 + \sum (w_t + c_t) e_t \tag{4.3.4}$$

We can now find the variance of the linear unbiased estimator b_2^* following the steps in Equation (4.2.11) and using the additional fact that

$$\sum c_t w_t = \sum \left[\frac{c_t (x_t - \bar{x})}{\sum (x_t - \bar{x})^2} \right] = \frac{1}{\sum (x_t - \bar{x})^2} \sum c_t x_t - \frac{\bar{x}}{\sum (x_t - \bar{x})^2} \sum c_t = 0$$

(The last step is from conditions (4.3.3))

Use the properties of variance to obtain:

$$\begin{aligned}
\operatorname{var}(b_2^*) &= \operatorname{var}\left(\beta_2 + \sum (w_t + c_t) e_t\right) = \sum (w_t + c_t)^2 \operatorname{var}(e_t) \\
&= \sigma^2 \sum (w_t + c_t)^2 = \sigma^2 \sum w_t^2 + \sigma^2 \sum c_t^2 \quad (\text{Note: } 2\sigma^2 \sum c_t w_t = 0) \\
&= \operatorname{var}(b_2) + \sigma^2 \sum c_t^2 \\
&\geq \operatorname{var}(b_2) \text{ since } \sum c_t^2 \geq 0 \text{ (and } \sigma^2 > 0)
\end{aligned} \tag{4.3.5}$$

The last line of Equation (4.3.5) establishes that, for the family of linear and unbiased estimators b_2^*, each of the alternative estimators has variance that is greater than or equal to that of the least squares estimator b_2. The only time that $\operatorname{var}(b_2^*) = \operatorname{var}(b_2)$ is when all the $c_t = 0$, in which case $b_2^* = b_2$. Thus, *there is no other linear and unbiased estimator* of β_2 that is better than b_2, which proves the Gauss-Markov theorem.

4.4 The Probability Distribution of the Least Squares Estimators

- If we make the normality assumption, that the random errors e_t are normally distributed with mean 0 and variance σ^2, then the probability distributions of the least squares estimators are also normal.

$$b_1 \sim N\left(\beta_1, \frac{\sigma^2 \sum x_t^2}{T\sum (x_t - \overline{x})^2}\right)$$

$$b_2 \sim N\left(\beta_2, \frac{\sigma^2}{\sum (x_t - \overline{x})^2}\right) \tag{4.4.1}$$

- This conclusion is obtained in two steps. First, based on assumption SR1, if e_t is normal, then so is y_t. Second, the least squares estimators are linear estimator, of the term $b_2 = \sum w_t y_t$, and weighted sums of normal random variables, using Equation (2.6.4), are normally distributed themselves. Consequently, if we make the normality assumption, assumption SR6 about the error term, then the least squares estimators are normally distributed.

- If assumptions SR1-SR5 hold, and if the sample size T is ***sufficiently large***, then the least squares estimators have a distribution that approximates the normal distributions shown in Equation (4.4.1).

4.5 Estimating the Variance of the Error Term

- The variance of the random variable e_t is the one unknown parameter of the simple linear regression model that remains to be estimated. The variance of the random variable e_t (see Chapter 2.3.4) is

$$\operatorname{var}(e_t) = \sigma^2 = E[(e_t - E(e_t))^2] = E(e_t^2) \tag{4.5.1}$$

  if the assumption $E(e_t) = 0$ is correct.

- Since the "expectation" is an average value we might consider estimating σ^2 as the average of the squared errors,

$$\hat{\sigma}^2 = \frac{\sum e_t^2}{T} \tag{4.5.2}$$

The formula in Equation (4.5.2) is unfortunately of no use, since the random error e_t are *unobservable*!

- While the random errors themselves are unknown, we do have an analogue to them, namely, the least squares residuals. Recall that the random errors are

$$e_t = y_t - \beta_1 - \beta_2 x_t$$

and the least squares residuals are obtained by replacing the unknown parameters by their least squares estimators,

$$\hat{e}_t = y_t - b_1 - b_2 x_t$$

- It seems reasonable to replace the random errors e_t in Equation (4.5.2) by their analogues, the least squares residuals, to obtain

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_t^2}{T} \tag{4.5.3}$$

- Unfortunately, the estimator in Equation (4.5.3) is a *biased* estimator of σ^2. Happily, there is a simple modification that produces an unbiased estimator, and that is

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_t^2}{T-2} \tag{4.5.4}$$

The "2" that is subtracted in the denominator is the number of regression parameters (β_1, β_2) in the model. The reason that the sum of the squared residuals is divided by $T-2$ is that while there are T data points or observations, the estimation of the intercept and slope puts two constraints on the data. This leaves $T-2$ unconstrained observations with which to estimate the residual variance. This subtraction makes the estimator $\hat{\sigma}^2$ unbiased, so that

$$E(\hat{\sigma}^2) = \sigma^2 \tag{4.5.5}$$

Consequently, before the data are obtained, we have an unbiased estimation procedure for the variance of the error term, σ^2, at our disposal.

4.5.1 Estimating the Variances and Covariances of the Least Squares Estimators

- Replace the unknown error variance σ^2 in Equation (4.2.10) by its estimator to obtain:

$$\widehat{\text{var}}(b_1) = \hat{\sigma}^2 \left[\frac{\sum x_t^2}{T \sum (x_t - \bar{x})^2} \right], \qquad \text{se}(b_1) = \sqrt{\widehat{\text{var}}(b_1)}$$

$$\widehat{\text{var}}(b_2) = \frac{\hat{\sigma}^2}{\sum (x_t - \bar{x})^2}, \qquad \text{se}(b_2) = \sqrt{\widehat{\text{var}}(b_2)} \tag{4.5.6}$$

$$\widehat{\text{cov}}(b_1, b_2) = \hat{\sigma}^2 \left[\frac{-\bar{x}}{\sum (x_t - \bar{x})^2} \right]$$

Therefore, having an unbiased estimator of the error variance, we can *estimate* the variances of the least squares estimators b_1 and b_2, and the covariance between them. The square roots of the estimated variances, se(b_1) and se(b_2), are the standard errors of b_1 and b_2.

4.5.2 The Estimated Variances and Covariances for the Food Expenditure Example

- The least squares estimates of the parameters in the food expenditure model are given in Chapter 3.3.2. In order to estimate the variance and covariance of the least squares estimators, we must compute the least squares residuals and calculate the estimate of the error variance in Equation (4.6.4). In Table 4.2 are the least squares residuals for the first five households in Table 3.1.

Table 4.2 Least Squares Residuals for Food Expenditure Data

y	$\hat{y} = b_1 + b_2 x$	$\hat{e} = y - y$
52.25	73.9045	–21.6545
58.32	84.7834	–26.4634
81.79	95.2902	–13.5002
119.90	100.7424	19.1576
125.80	102.7181	23.0819

- Using the residuals for all $T = 40$ observations, we estimate the error variance to be

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_t^2}{T-2} = \frac{54311.3315}{38} = 1429.2456$$

The estimated variances, covariances and corresponding standard errors are

$$\hat{\text{var}}(b_1) = \hat{\sigma}^2 \left[\frac{\sum x_t^2}{T \sum (x_t - \overline{x})^2} \right] = 1429.2456 \left[\frac{21020623}{40(1532463)} \right] = 490.1200$$

$$\text{se}(b_1) = \sqrt{\hat{\text{var}}(b_1)} = \sqrt{490.1200} = 22.1387$$

$$\hat{\text{var}}(b_2) = \frac{\hat{\sigma}^2}{\sum (x_t - \overline{x})^2} = \frac{1429.2456}{1532463} = 0.0009326$$

$$\mathrm{se}(b_2) = \sqrt{\hat{\mathrm{var}}(b_2)} = \sqrt{0.0009326} = 0.0305$$

$$\hat{\mathrm{cov}}(b_1, b_2) = \hat{\sigma}^2 \left[\frac{-\bar{x}}{\sum (x_t - \bar{x})^2} \right] = 1429.2456 \left[\frac{-698}{1532463} \right] = -0.6510$$

4.5.3 Sample Computer Output

Dependent Variable: FOODEXP
Method: Least Squares
Sample: 1 40
Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	40.76756	22.13865	1.841465	0.0734
INCOME	0.128289	0.030539	4.200777	0.0002

R-squared	0.317118	Mean dependent var	130.3130
Adjusted R-squared	0.299148	S.D. dependent var	45.15857
S.E. of regression	37.80536	Akaike info criterion	10.15149
Sum squared resid	54311.33	Schwarz criterion	10.23593
Log likelihood	-201.0297	F-statistic	17.64653
Durbin-Watson stat	2.370373	Prob(F-statistic)	0.000155

Table 4.3 EViews Regression Output

```
Dependent Variable: FOODEXP

                              Analysis of Variance

                                  Sum of          Mean
        Source          DF       Squares        Square       F Value       Prob>F

        Model            1   25221.22299   25221.22299        17.647       0.0002
        Error           38   54311.33145    1429.24556
        C Total         39   79532.55444

            Root MSE        37.80536     R-square       0.3171
            Dep Mean       130.31300     Adj R-sq       0.2991
            C.V.            29.01120

                              Parameter Estimates

                         Parameter         Standard    T for H0:
       Variable  DF       Estimate            Error   Parameter=0    Prob > |T|

       INTERCEP   1      40.767556      22.13865442         1.841        0.0734
       INCOME     1       0.128289       0.03053925         4.201        0.0002
```

Table 4.4 SAS Regression Output

```
VARIANCE OF THE ESTIMATE-SIGMA**2 = 1429.2

VARIABLE   ESTIMATED  STANDARD
  NAME    COEFFICIENT   ERROR
X          0.12829     0.3054E-01
CONSTANT    40.768       22.14
```

Table 4.5 SHAZAM Regression Output

```
                    Covariance of Estimates

          COVB                INTERCEP                 X

          INTERCEP       490.12001955       -0.650986935
          X              -0.650986935        0.000932646
```

Table 4.6 SAS Estimated Covariance Array

Appendix

Deriving Equation (4.5.4)

$$\hat{e}_t = y_t - b_1 - b_2 x_t = (\beta_1 + \beta_2 x_t + e_t) - b_1 - b_2 x_t$$

$$= -(b_1 - \beta_1) - (b_2 - \beta_2) x_t + e_t$$

$$= (b_2 - \beta_2)\bar{x} - \bar{e} - (b_2 - \beta_2) x_t + e_t \quad \text{(see derivation of var}(b_1))$$

$$= -(b_2 - \beta_2)(x_t - \bar{x}) + (e_t - \bar{e})$$

$$\hat{e}_t^2 = (b_2 - \beta_2)^2 (x_t - \bar{x})^2 + (e_t - \bar{e})^2 - 2(b_2 - \beta_2)(x_t - \bar{x})(e_t - \bar{e})$$

$$= (b_2 - \beta_2)^2 (x_t - \bar{x})^2 + (e_t - \bar{e})^2 - 2(b_2 - \beta_2)(x_t - \bar{x}) e_t + 2(b_2 - \beta_2)(x_t - \bar{x})\bar{e}$$

$$\sum \hat{e}_t^2 = (b_2 - \beta_2)^2 \sum (x_t - \bar{x})^2 + \sum (e_t - \bar{e})^2$$
$$- 2(b_2 - \beta_2) \sum (x_t - \bar{x}) e_t + 2(b_2 - \beta_2) \sum (x_t - \bar{x}) \bar{e}$$

$$= (b_2 - \beta_2)^2 \sum (x_t - \bar{x})^2 + \sum (e_t - \bar{e})^2 - 2(b_2 - \beta_2) \sum (x_t - \bar{x}) e_t \quad (\because \sum (x_t - \bar{x}) e_t = 0)$$

Note:

$$b_2 - \beta_2 = \sum w_t e_t = \frac{\sum (x_t - \bar{x}) e_t}{\sum (x_t - \bar{x})^2} \quad \text{(from Equations (4.2.1) and (4.2.2)}$$

$$\Rightarrow \sum (x_t - \bar{x}) e_t = (b_2 - \beta_2) \sum (x_t - \bar{x})^2$$

Note:

$$\sum (e_t - \bar{e})^2 = \sum (e_t^2 - 2 e_t \bar{e} + \bar{e}^2) = \sum e_t^2 - 2\bar{e} \sum e_t + T\bar{e}^2$$

$$= \sum e_t^2 - 2\left(\frac{1}{T} \sum e_t\right)\left(\sum e_t\right) + T\left(\frac{1}{T} \sum e_t\right)^2$$

$$= \sum e_t^2 - \frac{2}{T}\left(\sum e_t\right)^2 + \frac{1}{T}\left(\sum e_t\right)^2 = \sum e_t^2 - \frac{1}{T}\left(\sum e_t\right)^2$$

$$= \sum e_t^2 - \frac{1}{T}\left(\sum e_t^2 + 2 \sum_{i \neq j} e_i e_j\right) = \frac{T-1}{T} \sum e_t^2 - \frac{2}{T} \sum_{i \neq j} e_i e_j$$

Now,

$$\sum \hat{e}_t^2 = -(b_2 - \beta_2)^2 \sum (x_t - \bar{x})^2 + \sum (e_t - \bar{e})^2$$

$$= -(b_2 - \beta_2)^2 \sum (x_t - \bar{x})^2 + \frac{T-1}{T} \sum e_t^2 - \frac{2}{T} \sum_{i \neq i} e_i e_j$$

$$E[\sum \hat{e}_t^2] = -E[(b_2 - \beta_2)^2] \sum (x_t - \bar{x})^2 + \frac{T-1}{T} \sum E[e_t^2] - \frac{2}{T} \sum_{i \neq i} E(e_i e_j)$$

$$= -E[(b_2 - \beta_2)^2] \sum (x_t - \bar{x})^2 + \frac{T-1}{T} \sum E[e_t^2] \quad (\because E(e_i e_j) = 0)$$

$$= -\operatorname{var}(b_2) \sum (x_t - \bar{x})^2 + \frac{T-1}{T} (T\sigma^2)$$

$$= -\left(\frac{\sigma^2}{\sum (x_t - \bar{x})^2}\right) \sum (x_t - \bar{x})^2 + (T-1)\sigma^2 \quad \text{(from Equation (4.2.11))}$$

$$= -\sigma^2 + (T-1)\sigma^2 = (T-2)\sigma^2$$

Finally,

$$E[\hat{\sigma}^2] = \frac{1}{T-2} E[\sum e_t^2]$$

$$= \frac{1}{T-2}(T-2)\sigma^2 = \sigma^2$$

Exercise

4.1	4.2	4.3	4.4	4.5
4.6	4.13	4.14	4.15	4.16
4.17				