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. $E(e_t) = 0 \iff E(y_t) = \beta_1 + \beta_2 x_t$

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

4.1 The Least Squares Estimators as Random Variables

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

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

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

$$b_1 = \bar{y} - b_2\bar{x} \qquad (3.3.8b)$$

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.

- When the formulas for b_1 and b_2, 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***. 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

4.2.1 The Expected Values of b_1 and b_2

- 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

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

The expected value of a sum is the sum of the expected values (see Chapter 2.5.1):

$$\begin{aligned} E(b_2) &= E\left(\beta_2 + \sum w_t e_t\right) = E(\beta_2) + \sum E(w_t e_t) \\ &= \beta_2 + \sum w_t E(e_t) = \beta_2 \quad [\text{since } E(e_t) = 0] \end{aligned} \tag{4.2.3}$$

4.2.1a The Repeated Sampling Context

Table 4.1 contains least squares estimates of the food expenditure model from 10 random samples of size T=40 from the same population

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

4.2.1b Derivation of Equation 4.2.1

$$\sum (x_t - \overline{x})^2 = \sum x_t^2 - 2\overline{x}\sum x_t + T\,\overline{x}^2 = \sum x_t^2 - 2\overline{x}\left(T\frac{1}{T}\sum x_t\right) + T\,\overline{x}^2$$
$$= \sum x_t^2 - 2T\,\overline{x}^2 + T\,\overline{x}^2 = \sum x_t^2 - T\,\overline{x}^2 \qquad (4.2.4a)$$

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

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

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

b_2 in deviation from the mean form is:

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

- Recall that

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

- Then, the formula for b_2 becomes

$$\begin{aligned} b_2 &= \frac{\sum (x_t - \overline{x})(y_t - \overline{y})}{\sum (x_t - \overline{x})^2} = \frac{\sum (x_t - \overline{x})y_t - \overline{y}\sum (x_t - \overline{x})}{\sum (x_t - \overline{x})^2} \\ &= \frac{\sum (x_t - \overline{x})y_t}{\sum (x_t - \overline{x})^2} = \sum \left[\frac{(x_t - \overline{x})}{\sum (x_t - \overline{x})^2} \right] y_t = \sum w_t y_t \end{aligned} \tag{4.2.8}$$

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:

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

$\sum w_t = 0$, this eliminates the term $\beta_1 \sum w_t$.

$\sum w_t x_t = 1$, so $\beta_2 \sum w_t x_t = \beta_2$, and (4.2.9a) simplifies to equation 4.2.1

$$b_2 = \beta_2 + \sum w_t e_t \qquad (4.2.9b)$$

The term $\sum w_t = 0$, because

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

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

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

4.2.2 The Variances and Covariance of b_1 and b_2

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

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:

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

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

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

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.
2. The sum of squares of the values of x about their sample mean, $\sum(x_t - \bar{x})^2$, appears in each of the variances and in the covariance.
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*.
4. The term Σx^2 appears in $\mathrm{var}(b_1)$.
5. The sample mean of the x-values appears in $\mathrm{cov}(b_1,b_2)$.

Deriving the variance of *b*$_2$: The starting point is equation 4.2.1.

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

The very last step uses the fact that

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

4.2.3 Linear Estimators

- The least squares estimator b_2 is a weighted sum of the observations y_t, $b_2 = \sum w_t y_t$
- Estimators like b_2, that are linear combinations of an observable random variable, *linear estimators*

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

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
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
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.
- Into this new estimator substitute y_t and simplify, using the properties of w_t in equation 4.2.9.

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

since $\Sigma w_t = 0$ and $\Sigma w_t x_t = 1$.

$$E(b_2^*) = \beta_1 \sum c_t + \beta_2 + \beta_2 \sum c_t x_t + \sum (w_t + c_t) E(e_t)$$

$$= \beta_1 \sum c_t + \beta_2 + \beta_2 \sum c_t x_t \tag{4.3.2}$$

- 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

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

Use the properties of variance to obtain:

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

4.4 The Probability Distribution of the Least Squares Estimators

- *If* we make the normality assumption, assumption SR6 about the error term, then the least squares estimators are normally distributed.

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

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

- 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

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

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,

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

- Recall that the random errors are

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

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

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

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

- There is a simple modification that produces an unbiased estimator, and that is

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

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

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:

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

$$\widehat{\text{var}}(b_2) = \frac{\hat{\sigma}^2}{\sum (x_t - \overline{x})^2}, \qquad \text{se}(b_2) = \sqrt{\text{var}(b_2)} \qquad (4.6.6)$$

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

4.6.2 The Estimated Variances and Covariances for the Food Expenditure Example

Table 4.1 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

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

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

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

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

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

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

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

Source	DF	Sum of Squares	Mean 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

Variable	DF	Parameter Estimate	Standard Error	T for H0: 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 NAME	ESTIMATED COEFFICIENT	STANDARD 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