## Properties of the Least Squares Estimators

### Assumptions of the Simple Linear Regression Model

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR1.</td>
<td>$y_t = \beta_1 + \beta_2 x_t + e_t$</td>
</tr>
<tr>
<td>SR2.</td>
<td>$E(e_t) = 0 \iff E(y_t) = \beta_1 + \beta_2 x_t$</td>
</tr>
<tr>
<td>SR3.</td>
<td>$\text{var}(e_t) = \sigma^2 = \text{var}(y_t)$</td>
</tr>
<tr>
<td>SR4.</td>
<td>$\text{cov}(e_i, e_j) = \text{cov}(y_i, y_j) = 0$</td>
</tr>
<tr>
<td>SR5.</td>
<td>$x_t$ is not random and takes at least two values</td>
</tr>
<tr>
<td>SR6.</td>
<td>$e_t \sim N(0, \sigma^2) \iff y_t \sim N[(\beta_1 + \beta_2 x_t), \sigma^2]$ (optional)</td>
</tr>
</tbody>
</table>
4.1 The Least Squares Estimators as Random Variables

• The least squares estimator \( b_2 \) of the slope parameter \( \beta_2 \), based on a sample of \( T \) observations, is

\[
    b_2 = \frac{T \sum x_i y_i - \sum x_i \sum y_i}{T \sum x_i^2 - (\sum x_i)^2}
\]  \hspace{1cm} (3.3.8a)

• The least squares estimator \( b_1 \) of the intercept parameter \( \beta_1 \) is

\[
    b_1 = \bar{y} - b_2 \bar{x}
\]  \hspace{1cm} (3.3.8b)

where \( \bar{y} = \sum y_i / T \) and \( \bar{x} = \sum x_i / T \) 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,

$$b_2 = \beta_2 + \sum_{}^{} w_i e_i$$  \hspace{1cm} (4.2.1)

where $w_i$ is a constant (non-random) given by

$$w_i = \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}$$  \hspace{1cm} (4.2.2)

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

$$E(b_2) = E\left(\beta_2 + \sum w_i e_i\right) = E(\beta_2) + \sum E(w_i e_i)$$  \hspace{1cm} (4.2.3)

$$= \beta_2 + \sum w_i E(e_i) = \beta_2 \hspace{1cm} \text{[since } E(e_i) = 0\text{]}$$
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>
<thead>
<tr>
<th>$n$</th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51.1314</td>
<td>0.1442</td>
</tr>
<tr>
<td>2</td>
<td>61.2045</td>
<td>0.1286</td>
</tr>
<tr>
<td>3</td>
<td>40.7882</td>
<td>0.1417</td>
</tr>
<tr>
<td>4</td>
<td>80.1396</td>
<td>0.0886</td>
</tr>
<tr>
<td>5</td>
<td>31.0110</td>
<td>0.1669</td>
</tr>
<tr>
<td>6</td>
<td>54.3099</td>
<td>0.1086</td>
</tr>
<tr>
<td>7</td>
<td>69.6749</td>
<td>0.1003</td>
</tr>
<tr>
<td>8</td>
<td>71.1541</td>
<td>0.1009</td>
</tr>
<tr>
<td>9</td>
<td>18.8290</td>
<td>0.1758</td>
</tr>
<tr>
<td>10</td>
<td>36.1433</td>
<td>0.1626</td>
</tr>
</tbody>
</table>
4.2.1b Derivation of Equation 4.2.1

\[
\sum (x_t - \bar{x})^2 = \sum x_t^2 - 2\bar{x} \sum x_t + T\bar{x}^2 = \sum x_t^2 - 2\bar{x} \left( \frac{1}{T} \sum x_t \right) + T\bar{x}^2
\]  
\[= \sum x_t^2 - 2T\bar{x}^2 + T\bar{x}^2 = \sum x_t^2 - T\bar{x}^2 \tag{4.2.4a} \]

\[
\sum (x_t - \bar{x})^2 = \sum x_t^2 - T\bar{x}^2 = \sum x_t^2 - \bar{x} \sum x_t = \sum x_t^2 - \frac{\left( \sum x_t \right)^2}{T} \tag{4.2.4b} \]

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

\[
\sum (x_t - \bar{x})(y_t - \bar{y}) = \sum x_t y_t - T\bar{x}\bar{y} = \sum x_t y_t - \frac{\sum x_t \sum y_t}{T} \tag{4.2.5} \]
$b_2$ in deviation from the mean form is:

$$b_2 = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2}$$  \hspace{1cm} (4.2.6)

- Recall that

$$\sum (x_t - \bar{x}) = 0$$  \hspace{1cm} (4.2.7)

- Then, the formula for $b_2$ becomes

$$b_2 = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2} = \frac{\sum (x_t - \bar{x})y_t - \bar{y}\sum (x_t - \bar{x})}{\sum (x_t - \bar{x})^2}$$  \hspace{1cm} (4.2.8)

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

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:

\[
b_2 = \sum w_t y_t = \sum w_t (\beta_1 + \beta_2 x_t + e_t) = \beta_1 \sum w_t + \beta_2 \sum w_t x_t + \sum w_t e_t \tag{4.2.9a}
\]

\( \sum w_t = 0 \), 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

\[
b_2 = \beta_2 + \sum w_t e_t \tag{4.2.9b}
\]
The term $\sum w_t = 0$, because

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

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

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

Consequently
\[ \sum w_i x_i = \sum \frac{(x_i - \bar{x})x_i}{(x_i - \bar{x})^2} = \sum \frac{(x_i - \bar{x})x_i}{(x_i - \bar{x})} = 1 \]

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

\[
\text{var}(b_2) = \frac{\sigma^2}{\sum (x_t - \bar{x})^2}
\]

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

(4.2.10)
Let us consider the factors that affect the variances and covariance in equation 4.2.10.

1. The variance of the random error term, $\sigma^2$, appears in each of the expressions.
2. The sum of squares of the values of $x$ about their sample mean, $\sum (x_i - \bar{x})^2$, 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 $\Sigma x^2$ appears in $\text{var}(b_1)$.
5. The sample mean of the $x$-values appears in $\text{cov}(b_1, b_2)$.
Deriving the variance of \( b_2 \): The starting point is equation 4.2.1.

\[
\text{var}(b_2) = \text{var}(\beta_2 + \sum w_t e_t) = \text{var}(\sum w_t e_t) \quad \text{[since } \beta_2 \text{ is a constant]}
\]

\[
= \sum w_t^2 \text{var}(e_t) \quad \text{[using } \text{cov}(e_i, e_j) = 0]\]

\[
= \sigma^2 \sum w_t^2 \quad \text{[using } \text{var}(e_t) = \sigma^2]\]

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

The very last step uses the fact that

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

\[(4.2.12)\]
4.2.3 Linear Estimators

- The least squares estimator $b_2$ is a weighted sum of the observations $y_t$, $b_2 = \sum w_i y_t$
- Estimators like $b_2$, that are linear combinations of an observable random variable, are called 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 $\beta_1$ and $\beta_2$. They are the *Best Linear Unbiased Estimators* (BLUE) of $\beta_1$ and $\beta_2$
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 $\beta_1$ and $\beta_2$.

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 \( b_2^* = \sum k_i y_t \) (where the \( k_i \) are constants) be any other linear estimator of \( \beta_2 \).
- Suppose that \( k_i = w_i + c_i \), where \( c_i \) is another constant and \( w_i \) 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.

\[
b_2^* = \sum k_i y_t = \sum (w_i + c_i) y_t = \sum (w_i + c_i)(\beta_1 + \beta_2 x_t + e_t)
\]

\[
= \sum (w_i + c_i) \beta_1 + \sum (w_i + c_i) \beta_2 x_t + \sum (w_i + c_i) e_t
\]

\[
= \beta_1 \sum w_i + \beta_1 \sum c_i + \beta_2 \sum w_i x_t + \beta_2 \sum c_i x_t + \sum (w_i + c_i) e_t
\]

\[
= \beta_1 \sum c_i + \beta_2 + \beta_2 \sum c_i x_t + \sum (w_i + c_i) e_t
\]

since \( \Sigma w_t = 0 \) and \( \Sigma w_t x_t = 1 \).
\begin{align*}
E(b_2^*) &= \beta_1 \sum c_i + \beta_2 + \beta_2 \sum c_i x_i + \sum (w_t + c_i)E(e_t) \\
&= \beta_1 \sum c_i + \beta_2 + \beta_2 \sum c_i x_i
\end{align*}
(4.3.2)

• In order for the linear estimator \( b_2^* = \sum k_i y_t \) to be unbiased it must be true that

\[ \sum c_i = 0 \quad \text{and} \quad \sum c_i x_i = 0 \]
(4.3.3)

• These conditions must hold in order for \( b_2^* = \sum k_i y_t \) 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):

\[ b_2^* = \sum k_i y_i = \beta_2 + \sum (w_t + c_t)e_t \] (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

\[
\sum c_i w_t = \sum \left[ \frac{c_i (x_t - \bar{x})}{\sum (x_t - \bar{x})^2} \right] = \frac{1}{\sum (x_t - \bar{x})^2} \sum c_i x_t - \frac{\bar{x}}{\sum (x_t - \bar{x})^2} \sum c_t = 0
\]
Use the properties of variance to obtain:

\[ \text{var}(b^*_2) = \text{var}(\beta_2 + \sum (w_t + c_t) e_t) = \sum (w_t + c_t)^2 \text{var}(e_t) \]

\[ = \sigma^2 \sum (w_t + c_t)^2 = \sigma^2 \sum w_t^2 + \sigma^2 \sum c_t^2 \]

\[ = \text{var}(b_2) + \sigma^2 \sum c_t^2 \]

\[ \geq \text{var}(b_2) \text{ since } \sum c_t^2 \geq 0 \]
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.

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

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

- If assumptions SR1-SR5 hold, and if the sample size \( T \) is \textit{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

$$\text{var}(e_t) = \sigma^2 = E[e_t - E(e_t)]^2 = E(e_t^2)$$  \hspace{1cm} (4.5.1)

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

Since the “expectation” is an average value we might consider estimating $\sigma^2$ as the average of the squared errors,

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{T}$$  \hspace{1cm} (4.5.2)
Recall that the random errors are

\[ e_i = y_i - \beta_1 - \beta_2 x_i \]

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

\[ \hat{e}_i = y_i - \hat{b}_1 - \hat{b}_2 x_i \]

\[ \hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{T} \quad (4.5.3) \]

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

\[ \tilde{\sigma}^2 = \frac{\sum \hat{e}_i^2}{T - 2} \quad (4.5.4) \]

\[ E(\tilde{\sigma}^2) = \sigma^2 \quad (4.5.5) \]
4.5.1 Estimating the Variances and Covariances of the Least Squares Estimators

- Replace the unknown error variance $\sigma^2$ in equation 4.2.10 by its estimator to obtain:

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

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

$$
\hat{\text{cov}}(b_1, b_2) = \hat{\sigma}^2 \left[ \frac{-\bar{x}}{\sum (x_t - \bar{x})^2} \right]
$$
4.6.2 The Estimated Variances and Covariances for the Food Expenditure Example

**Table 4.1** Least Squares Residuals for Food Expenditure Data

<table>
<thead>
<tr>
<th>$y$</th>
<th>$\hat{y} = b_1 + b_2 x$</th>
<th>$\hat{\varepsilon} = y - \hat{y}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>52.25</td>
<td>73.9045</td>
<td>-21.6545</td>
</tr>
<tr>
<td>58.32</td>
<td>84.7834</td>
<td>-26.4634</td>
</tr>
<tr>
<td>81.79</td>
<td>95.2902</td>
<td>-13.5002</td>
</tr>
<tr>
<td>119.90</td>
<td>100.7424</td>
<td>19.1576</td>
</tr>
<tr>
<td>125.80</td>
<td>102.7181</td>
<td>23.0819</td>
</tr>
</tbody>
</table>

\[
\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_i^2}{T - 2} = \frac{54311.3315}{38} = 1429.2456
\]
\[
\text{vár}(b_1) = \hat{\sigma}^2 \left[ \frac{\sum x_t^2}{T \sum (x_t - \bar{x})^2} \right] = 1429.2456 \left[ \frac{21020623}{40(1532463)} \right] = 490.1200
\]

\[
\text{se}(b_1) = \sqrt{\text{vár}(b_1)} = \sqrt{490.1200} = 22.1387
\]

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

\[
\text{se}(b_2) = \sqrt{\text{vár}(b_2)} = \sqrt{0.0009326} = 0.0305
\]

\[
\text{côv}(b_1, b_2) = \hat{\sigma}^2 \left[ \frac{-\bar{x}}{\sum (x_t - \bar{x})^2} \right] = 1429.2456 \left[ \frac{-698}{1532463} \right] = -0.6510
\]
### 4.5.3 Sample Computer Output

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>40.76756</td>
<td>22.13865</td>
<td>1.841465</td>
<td>0.0734</td>
</tr>
<tr>
<td>INCOME</td>
<td>0.128289</td>
<td>0.030539</td>
<td>4.200777</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R-squared</td>
<td>0.317118</td>
<td>Mean dependent var</td>
<td>130.3130</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.299148</td>
<td>S.D. dependent var</td>
<td>45.15857</td>
<td></td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>37.80536</td>
<td>Akaike info criterion</td>
<td>10.15149</td>
<td></td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>54311.33</td>
<td>Schwarz criterion</td>
<td>10.23593</td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-201.0297</td>
<td>F-statistic</td>
<td>17.64653</td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson stat</td>
<td>2.370373</td>
<td>Prob(F-statistic)</td>
<td>0.000155</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.3 EViews Regression Output**
Dependent Variable: FOODEXP

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>25221.22299</td>
<td>25221.22299</td>
<td>17.647</td>
<td>0.0002</td>
</tr>
<tr>
<td>Error</td>
<td>38</td>
<td>54311.33145</td>
<td>1429.24556</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C Total</td>
<td>39</td>
<td>79532.55444</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Parameter Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>T for H0: Parameter=0</th>
<th>Prob &gt;</th>
<th>T</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEP</td>
<td>1</td>
<td>40.76756</td>
<td>22.13865442</td>
<td>1.841</td>
<td>0.0734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INCOME</td>
<td>1</td>
<td>0.128289</td>
<td>0.03053925</td>
<td>4.201</td>
<td>0.0002</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4 SAS Regression Output
VARIANCE OF THE ESTIMATE-SIGMA**2 = 1429.2

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>ESTIMATED</th>
<th>STANDARD</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAME</td>
<td>COEFFICIENT</td>
<td>ERROR</td>
</tr>
<tr>
<td>X</td>
<td>0.12829</td>
<td>0.3054E-01</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>40.768</td>
<td>22.14</td>
</tr>
</tbody>
</table>

Table 4.5 SHAZAM Regression Output

Covariance of Estimates

<table>
<thead>
<tr>
<th>COVB</th>
<th>INTERCEPT</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>490.12001955</td>
<td>-0.650986935</td>
</tr>
<tr>
<td>X</td>
<td>-0.650986935</td>
<td>0.000932646</td>
</tr>
</tbody>
</table>

Table 4.6 SAS Estimated Covariance Array