# Chapter 3

## The Simple Linear Regression Model:  Specification and Estimation

### 3.1   An Economic Model

- The simple regression function

$$E(y \mid x) = \mu_{y|x} = \beta_1 + \beta_2 x \qquad (3.1.1)$$

- Slope of regression line

$$\beta_2 = \frac{\Delta E(y \mid x)}{\Delta x} = \frac{dE(y \mid x)}{dx} \qquad (3.1.2)$$

"Δ" denotes "change in"

## 3.2 An Econometric Model

### Assumptions of the Simple Linear Regression Model-I

- The average value of *y*, for each value of *x*, is given by the *linear regression*

$$E(y) = \beta_1 + \beta_2 x$$

- For each value of *x*, the values of *y* are distributed about their mean value, following probability distributions that all have the same variance,

$$\text{var}(y) = \sigma^2$$

- The values of $y$ are all *uncorrelated*, and have zero *covariance*, implying that there is no linear association among them.

$$\text{cov}(y_i, y_j) = 0$$

This assumption can be made stronger by assuming that the values of $y$ are all *statistically independent.*

- The variable $x$ is not random and must take at least two different values

- (*optional*) The values of $y$ are *normally distributed* about their mean for each value of $x$,

$$y \sim N[(\beta_1 + \beta_2 x), \sigma^2]$$

## 3.2.1 Introducing the Error Term

The random error term is

$$e = y - E(y) = y - \beta_1 - \beta_2 x \qquad (3.2.1)$$

Rearranging gives

$$y = \beta_1 + \beta_2 x + e \qquad (3.2.2)$$

*y* is dependent variable; *x* is independent or explanatory variable

## Assumptions of the Simple Linear Regression Model-II

SR1 $y = \beta_1 + \beta_2 x + e$

SR2. $E(e) = 0 \Leftrightarrow E(y) = \beta_1 + \beta_2 x$

SR3. $\text{var}(e) = \sigma^2 = \text{var}(y)$

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

SR5. The variable *x* is not random and must take at least two different values.

SR6. (*optional*) The values of *e* are *normally distributed* about their mean

$$e \sim N(0, \sigma^2)$$

## 3.3 Estimating the Parameters for the Expenditure Relationship

### 3.3.1 The Least Squares Principle

- The fitted regression line is

$$\hat{y}_t = b_1 + b_2 x_t \tag{3.3.1}$$

- The least squares residual

$$\hat{e}_t = y_t - y_t = y_t - b_1 - b_2 x_t \tag{3.3.2}$$

- Any other fitted line

$$\hat{y}_t^* = b_1^* + b_2^* x_t \tag{3.3.3}$$

- Least squares line has smaller sum of squared residuals

$$\sum \hat{e}_t^2 = \sum (y_t - y_t)^2 \le \sum \hat{e}_t^{*2} = \sum (y_t - y_t^*)^2$$

- Least squares estimates are obtained by minimizing the sum of squares function

$$S(\beta_1, \beta_2) = \sum_{t=1}^{T} (y_t - \beta_1 - \beta_2 x_t)^2 \qquad (3.3.4)$$

- Math: Obtain partial derivatives

$$\frac{\partial S}{\partial \beta_1} = 2T\beta_1 - 2\sum y_t + 2\sum x_t \beta_2$$

$$(3.3.5)$$

$$\frac{\partial S}{\partial \beta_2} = 2\sum x_t^2 \beta_2 - 2\sum x_t y_t + 2\sum x_t \beta_1$$

- Set derivatives to zero

$$2(\sum y_t - Tb_1 - \sum x_t b_2) = 0$$

$$(3.3.6)$$

$$2(\sum x_t y_t - \sum x_t b_1 - \sum x_t^2 b_2) = 0$$

- Rearranging equation 3.3.6 leads to two equations usually known as the *normal equations*,

$$Tb_1 + \sum x_t b_2 = \sum y_t \qquad (3.3.7a)$$

$$\sum x_t b_1 + \sum x_t^2 b_2 = \sum x_t y_t \qquad (3.3.7b)$$

- Formulas for least squares estimates

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

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

- Since these formulas work for any values of the sample data, they are the least squares **estimators**.

## 3.3.2 Estimates for the Food Expenditure Function

$$b_2 = \frac{T\sum x_t y_t - \sum x_t \sum y_t}{T\sum x_t^2 - \left(\sum x_t\right)^2} = \frac{(40)(3834936.497) - (27920)(5212.520)}{(40)(21020623.02) - (27920)^2} \qquad (3.3.9a)$$

$$= 0.1283$$

$$b_1 = \bar{y} - b_2\bar{x} = 130.313 - (0.1282886)(698.0) = 40.7676 \qquad (3.3.9b)$$

A convenient way to report the values for $b_1$ and $b_2$ is to write out the *estimated* or *fitted* regression line:

$$\hat{y}_t = 40.7676 + 0.1283x_t \qquad (3.3.10)$$

### 3.3.3 Interpreting the Estimates

- The value $b_2 = 0.1283$ is an estimate of $\beta_2$, the amount by which weekly expenditure on food increases when weekly income increases by $1. Thus, we estimate that if income goes up by $100, weekly expenditure on food will increase by approximately $12.83.
- Strictly speaking, the intercept estimate $b_1 = 40.7676$ is an estimate of the weekly amount spent on food for a family with zero income

### 3.3.3a Elasticities

- The income elasticity of demand is a useful way to characterize the responsiveness of consumer expenditure to changes in income. From microeconomic principles the elasticity of any variable $y$ with respect to another variable $x$ is

$$\eta = \frac{\text{percentage change in } y}{\text{percentage change in } x} = \frac{\Delta y / y}{\Delta x / x} = \frac{\Delta y}{\Delta x} \cdot \frac{x}{y} \qquad (3.3.11)$$

- In the linear economic model given by equation 3.1.1 we have shown that

$$\beta_2 = \frac{\Delta E(y)}{\Delta x} \qquad (3.3.12)$$

- The elasticity of "average" expenditure with respect to income is

$$\eta = \frac{\Delta E(y)/E(y)}{\Delta x / x} = \frac{\Delta E(y)}{\Delta x} \cdot \frac{x}{E(y)} = \beta_2 \cdot \frac{x}{E(y)} \qquad (3.3.13)$$

- A frequently used alternative is to report the elasticity at the "point of the means" $(\overline{x}, \overline{y}) = (698.00, 130.31)$ since that is a representative point on the regression line.

$$\hat{\eta} = b_2 \cdot \frac{\overline{x}}{\overline{y}} = 0.1283 \times \frac{698.00}{130.31} = 0.687 \qquad (3.3.14)$$

3.3.3b Prediction

Suppose that we wanted to predict weekly food expenditure for a household with a weekly income of \$750. This prediction is carried out by substituting *x* = 750 into our estimated equation to obtain

$$\hat{y}_t = 40.7676 + 0.1283x_t = 40.7676 + 0.1283(750) = \$130.98 \qquad (3.3.15)$$

We *predict* that a household with a weekly income of \$750 will spend \$130.98 per week on food.

## 3.3.3c Examining 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 |

**Figure 3.10 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 |

**Figure 3.11 SAS Regression Output**

## 3.3.4 Other Economic Models

- The "log-log" model $\ln(y) = \beta_1 + \beta_2 \ln(x)$
- The derivative of ln(*y*) with respect to *x* is

$$\frac{d[\ln(y)]}{dx} = \frac{1}{y} \cdot \frac{dy}{dx}$$

- The derivative of $\beta_1 + \beta_2 \ln(x)$ with respect to *x* is

$$\frac{d[\beta_1 + \beta_2 \ln(x)]}{dx} = \frac{1}{x} \cdot \beta_2$$

- Setting these two pieces equal to one another, and solving for $\beta_2$ gives

$$\beta_2 = \frac{dy}{dx} \cdot \frac{x}{y} = \eta \qquad (3.3.16)$$