2 The simple regression model: estimation and properties

Ezequiel Uriel
University of Valencia
Version: 09-2013

2.1 Some definitions in the simple regression model

2.1.1 Population regression model and population regression function

In the simple regression model, the *population regression model* or, simply, the *population model* is the following:

$$y = \beta_1 + \beta_2 x + u \tag{2-1}$$

We shall look at the different elements of the model (2-1) and the terminology used to designate them. We are going to consider that there are three types of variables in the model: *y*, *x* and *u*. In this model there is only one factor *x* to explain *y*. All the other factors that affect *y* are jointly captured by *u*.

We typically refer to *y* as the endogenous (from the Greek: generated inside) variable or dependent variable. Other denominations are also used to designate *y*: left-hand side variable, explained variable, or regressand. In this model all these denominations are equivalent, but in other models, as we will see later on, there can be some differences.

In the simple linear regression of y on x, we typically refer to x as the exogenous (from the Greek: generated outside) variable or independent variable. Other denominations are also used to designate x: right-hand side variable, explanatory variable, regressor, covariate, or control variable. All these denominations are equivalent, but in other models, as we will see later, there can be some differences.

The variable u represents factors other than x that affect y. It is denominated error or random disturbance. The disturbance term can also capture measurement error in the dependent variable. The disturbance is an unobservable variable.

The parameters β_1 and β_2 are fixed and unknown.

On the right hand of (2-1) we can distinguish two parts: the systematic component $\beta_1 + \beta_2 x$ and the random disturbance u. Calling μ_y to the systematic component, we can write:

$$\mu_y = \beta_1 + \beta_2 x \tag{2-2}$$

This equation is known as the *population regression function* (*PRF*) or *population line*. Therefore, as can be seen in figure 2.1, μ_y is a linear function of x with intercept β_1 and slope β_2.

The linearity means that a one-unit increase in x changes the *expected value* of y - $\mu_y = E(y)$ - by β_1 units.

Now, let us suppose we have a random sample of size n $\{(y_i, x_i)$: $i = 1,\ldots,n\}$ from the studied population. In figure 2.2 the scatter diagram, corresponding to these data, have been displayed.

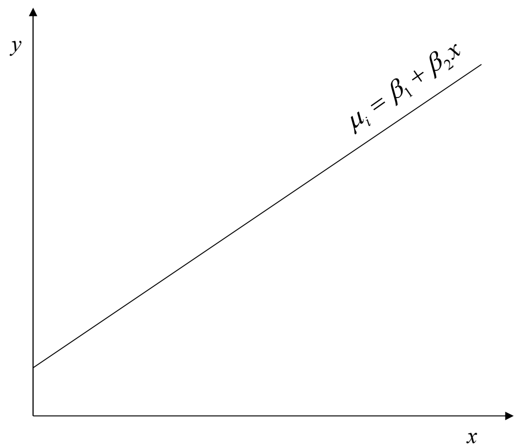


FIGURE 2.1. The population regression function. (PRF)

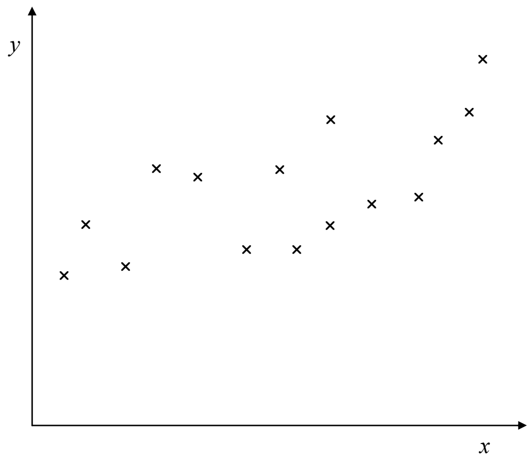


FIGURE 2.2. The scatter diagram.

We can express the population model for each observation of the sample:

$$y_i = \beta_1 + \beta_2 x_i + u_i \quad i = 1, 2, \cdots, n \tag{2-3}$$

In figure 2.3 the population regression function and the scatter diagram are put together, but it is important to keep in mind that although β_1 and β_2 are fixed, they are unknown. According to the model, it is possible to make the following decomposition from a theoretical point of view:

$$y_i = \mu_{yi} + u_i \quad i = 1, 2, \cdots, n \tag{2-4}$$

which is represented in figure 2.3 for the i^{th} observation. However, from an empirical point of view, it is not possible because β_1 and β_2 are unknown parameters and u_i is not observable.

2.1.2 Sample regression function

The basic idea of the regression model is to estimate the population parameters, β_2 and β_1, from a given sample.

The *sample regression function* (*SRF*) is the sample counterpart of the population regression function (*PRF*). Since the *SRF* is obtained for a given sample, a new sample will generate different estimates.

The *SRF*, which is an estimation of the PRF, given by

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i \tag{2-5}$$

allows us to calculate the *fitted value* ($\hat{y}_i$) for y when $x = x_i$. In the *SRF* $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimators of the parameters β_1 and β_2. For each x_i we have an observed value (y_i) and a fitted value ($\hat{y}_i$).

The difference between y_i and $\hat{y}_i$ is called the residual $\hat{u}_i$:

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i \tag{2-6}$$

In other words, the residual $\hat{u}_i$ is the difference between the sample value y_i and the fitted value of $\hat{y}_i$, as can be seen in figure 2.4. In this case, it is possible to calculate the decomposition:

$$y_i = \hat{y}_i + \hat{u}_i$$

for a given sample.

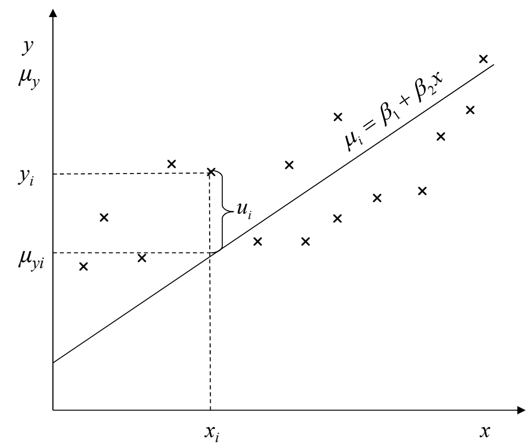


FIGURE 2.3. The population regression function and the scatter diagram.

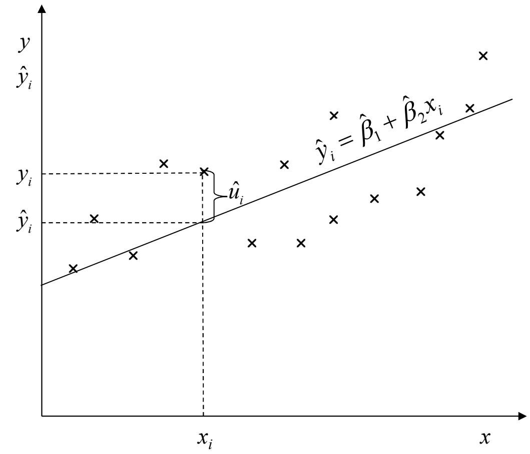


FIGURE 2.4. The sample regression function and the scatter diagram.

To sum up, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{y}_i$ and $\hat{u}_i$ are the sample counterpart of β_1, β_2, μ_{yi} and u_i respectively. It is possible to calculate $\hat{\beta}_1$ and $\hat{\beta}_2$ for a given sample, but the estimates will change for each sample. On the contrary, β_1 and β_2 are fixed, but unknown.

2.2 Obtaining the Ordinary Least Squares (OLS) Estimates

2.2.1 Different criteria of estimation

Before obtaining the least squares estimators, we are going to examine three alternative methods to illustrate the problem in hand. What these three methods have in common is that they try to minimize the residuals as a whole.

Criterion 1

The first criterion takes as estimators those values of $\hat{\beta}_1$ and $\hat{\beta}_2$ that make the sum of all the residuals as near to zero as possible. According to this criterion, the expression to minimize would be the following:

$$\text{Min}\left|\sum_{i=1}^{n}\hat{u}_i\right| \tag{2-7}$$

The main problem of this procedure is that the residuals of different signs can be compensated. Such a situation can be observed graphically in figure 2.5, in which three aligned observations are graphed, (x_1, y_1), (x_2, y_2) and (x_3, y_3). In this case the following happens:

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{y_3 - y_1}{x_3 - x_1}$$

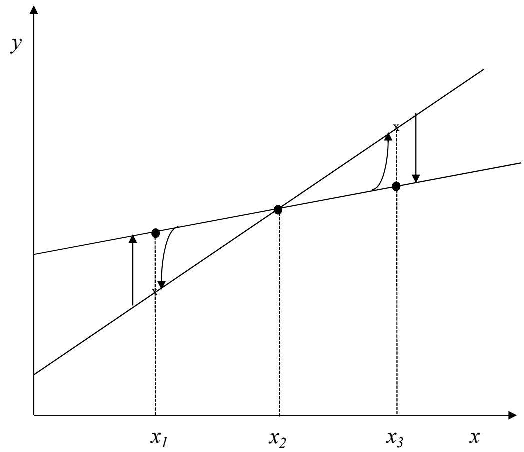


FIGURE 2.5. The problems of criterion 1.

If a straight line is fitted so that it passes through the three points, each one of the residuals will take value zero, and therefore

$$\left|\sum_{i=1}^{3}\hat{u}_i = 0\right|$$

This fit could be considered optimal. But it is also possible to obtain $\left|\sum_{i=1}^{3}\hat{u}_i = 0\right|$, by rotating the straight line - from the point x_2, y_2 - in any direction, as figure 2.5 shows, because $\hat{u}_3 = -\hat{u}_1$. In other words, by rotating this way the result $\left|\sum_{i=1}^{3} \hat{u}_i = 0\right|$ is always obtained. This simple example shows that this criterion is not appropriate for the estimation of the parameters given that, for any set of observations, an infinite number of straight lines exist, satisfying this criterion.

Criterion 2

In order to avoid the compensation of positive residuals with negative ones, the absolute values from the residuals are taken. In this case, the following expression would be minimized:

$$\text{Min} \sum_{i=1}^{n} |\hat{u}_i| \tag{2-8}$$

Unfortunately, although the estimators thus obtained have some interesting properties, their calculation is complicated and requires resolving the problem of linear programming or applying a procedure of iterative calculation.

Criterion 3

A third procedure is to minimize the sum of the square residuals, that is to say,

$$\text{Min } S = \text{Min} \sum_{i=1}^{n} \hat{u}_i^2 \tag{2-9}$$

The estimators obtained are denominated least square estimators (*LS*), and they enjoy certain desirable statistical properties, which will be studied later on. On the other hand, as opposed to the first of the examined criteria, when we square the residuals their compensation is avoided, and the least square estimators are simple to obtain, contrary to the second of the criteria. It is important to indicate that, from the moment we square the residuals, we proportionally penalize the bigger residuals more than the smaller ones (if a residual is double the size of another one, its square will be four times greater). This characterizes the least square estimation with respect to other possible procedures.

2.2.2 Application of least square criterion

Now, we are going to look at the process of obtaining the *LS* estimators. The objective is to minimize the residual sum of the squares (*S*). To do this, we are firstly going to express *S* as a function of the estimators, using (2-6):

Therefore, we must

$$\underset{\hat{\beta}_1,\hat{\beta}_2}{\text{Min}}\, S = \underset{\hat{\beta}_1,\hat{\beta}_2}{\text{Min}} \sum_{t=1}^{T} \hat{u}_t^2 = \underset{\hat{\beta}_1,\hat{\beta}_2}{\text{Min}} \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2 \tag{2-10}$$

To minimize *S*, we differentiate partially with respect to $\hat{\beta}_1$ and $\hat{\beta}_2$:

$$\frac{\partial S}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)$$

$$\frac{\partial S}{\partial \hat{\beta}_2} = -2\sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) x_i$$

The *LS* estimators are obtained by equaling the previous derivatives to 0:

$$\sum_{i=1}^{n}(y_i-\hat{\beta}_1-\hat{\beta}_2x_i)=0 \tag{2-11}$$

$$\sum_{i=1}^{n}(y_i-\hat{\beta}_1-\hat{\beta}_2x_i)x_i=0 \tag{2-12}$$

The equations (2-11) are denominated *normal equations* or *LS first order conditions*.

In operations with summations, the following rules must be taken into account:

$$\sum_{i=1}^{n}a=na$$

$$\sum_{i=1}^{n}ax_i=a\sum_{i=1}^{n}x_i$$

$$\sum_{i=1}^{n}(x_i+y_i)=\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}y_i$$

Operating with the normal equations, we have

$$\sum_{i=1}^{n}y_i=n\hat{\beta}_1+\hat{\beta}_2\sum_{i=1}^{n}x_i \tag{2-13}$$

$$\sum_{i=1}^{n}y_ix_i=\hat{\beta}_1\sum_{i=1}^{n}x_i+\hat{\beta}_2\sum_{i=1}^{n}x_i^2 \tag{2-14}$$

Dividing both sides of (2-13) by *n*, we have

$$\overline{y}=\hat{\beta}_1+\hat{\beta}_2\overline{x} \tag{2-15}$$

Therefore

$$\hat{\beta}_1=\overline{y}-\hat{\beta}_2\overline{x} \tag{2-16}$$

Substituting this value of $\hat{\beta}_1$ in the second normal equation (2-14), we have

$$\sum_{i=1}^{n}y_ix_i=(\overline{y}-\hat{\beta}_2\overline{x})\sum_{i=1}^{n}x_i+\hat{\beta}_2\sum_{i=1}^{n}x_i^2$$

$$\sum_{i=1}^{n}y_ix_i=\overline{y}\sum_{i=1}^{n}x_i-\hat{\beta}_2\overline{x}\sum_{i=1}^{n}x_i+\hat{\beta}_2\sum_{i=1}^{n}x_i^2$$

Solving for $\hat{\beta}_2$ we have:

$$\hat{\beta}_2=\frac{\sum_{i=1}^{n}y_ix_i-\overline{y}\sum_{i=1}^{n}x_i}{\sum_{i=1}^{n}x_i^2-\overline{x}\sum_{i=1}^{n}x_i} \tag{2-17}$$

Or, as can be seen in appendix 2.1,

$$\hat{\beta}_2=\frac{\sum_{i=1}^{n}(y_i-\overline{y})(x_i-\overline{x})}{\sum_{i=1}^{n}(x_i-\overline{x})^2} \tag{2-18}$$

Dividing the numerator and denominator of (2-18) by n, it can be seen that $\hat{\beta}_2$ is equal to the ratio between the two variables covariance and variance of x. Therefore, the sign of $\hat{\beta}_2$ is the same as the sign of the covariance.

Once $\hat{\beta}_2$ is calculated, then we can obtain $\hat{\beta}_1$ by using (2-16).

These are the *LS* estimators. Since other more complicated methods exist, also called least square methods, the method that we have applied is denominated ordinary least square (*OLS*), due to its simplicity.

In the precedent epigraphs $\hat{\beta}_1$ and $\hat{\beta}_2$ have been used to designate generic estimators. From now on, we will only designate *OLS* estimators with this notation.

EXAMPLE 2.1 Estimation of the consumption function

Given the Keynesian consumption function,

$$cons = \beta_1 + \beta_2 inc + u_i$$

we will estimate it using data from six households that appear in table 2.1.

TABLE 2.1. Data and calculations to estimate the consumption function.

Observ.	$cons_i$	inc_i	$cons_i \times inc_i$	inc_i^2	$cons_i - \overline{cons}$	$inc_i - \overline{inc}$	$(cons_i - \overline{cons}) \times (inc_i - \overline{inc})$	$(inc_i - \overline{inc})^2$
1	5	6	30	36	-4	-5	20	25
2	7	9	63	81	-2	-2	4	4
3	8	10	80	100	-1	-1	1	1
4	10	12	120	144	1	1	1	1
5	11	13	143	169	2	2	4	4
6	13	16	208	256	4	5	20	25
Sums	**54**	**66**	**644**	**786**	**0**	**0**	**50**	**60**

Calculating $\overline{cons}$ and $\overline{inc}$, and applying the formula (2-17), or alternatively (2-18), for the data table 2.1, we obtain

$$\overline{cons} = \frac{54}{6} = 9\,; \quad \overline{inc} = \frac{66}{6} = 11\,; \quad \text{(2-17): } \hat{\beta}_2 = \frac{644 - 9 \times 66}{786 - 11 \times 66} = 0.8\hat{3}\,; \quad \text{(2-18): } \hat{\beta}_2 = \frac{50}{60} = 0.8\hat{3}$$

Then by applying (2-16), we obtain $\hat{\beta}_1 = 9 - 0.8\hat{3} \times 11 = -0.1\hat{6}$

2.3 Some characteristics of *OLS* estimators

2.3.1 Algebraic implications of the estimation

The algebraic implications of the estimation are derived exclusively from the application of the *OLS* procedure to the simple linear regression model:

1. The sum of the OLS residuals is equal to 0:

$$\sum_{i=1}^{n} \hat{u}_i = 0 \tag{2-19}$$

From the definition of residual

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i \qquad i = 1, 2, \cdots, n \tag{2-20}$$

If we sum up the n observations, we get

$$\sum_{i=1}^{n}\hat{u}_i = \sum_{i=1}^{n}(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0 \qquad (2\text{-}21)$$

which is precisely the first equation (2-11) of the system of normal equations.

Note that, if (2-19) holds, it implies that

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n}\hat{y}_i \qquad (2\text{-}22)$$

and, dividing (2-19) and (2-22) by n, we obtain

$$\overline{\hat{u}} = 0 \qquad \overline{y} = \overline{\hat{y}} \qquad (2\text{-}23)$$

2. The OLS line always goes through the mean of the sample ($\overline{x},\overline{y}$ *).*

Effectively, dividing the equation (2-13) by n, we have:

$$\overline{y} = \hat{\beta}_1 + \hat{\beta}_2\overline{x} \qquad (2\text{-}24)$$

3. The sample cross product between each one of the regressors and the OLS residuals is zero.

That is to say,

$$\sum_{i=1}^{n} x_i\hat{u}_i = 0 \qquad (2\text{-}25)$$

We can see that (2-25) is equal to the second normal equation,

$$\sum_{i=1}^{n} x_i\hat{u}_i = \sum_{i=1}^{n} x_i(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0$$

given in (2-12).

4. The sample cross product between the fitted values ($\hat{y}$ *) and the OLS residuals is zero.*

That is to say,

$$\sum_{i=1}^{n}\hat{y}_i\hat{u}_i = 0 \qquad (2\text{-}26)$$

Proof

Taking into account the algebraic implications 1 -(2-19)- and 3 -(2-25)-, we have

$$\sum_{i=1}^{n}\hat{y}_i\hat{u}_i = \sum_{i=1}^{n}(\hat{\beta}_1 + \hat{\beta}_2 x_i)\hat{u}_i = \hat{\beta}_1\sum_{i=1}^{n}\hat{u}_i + \hat{\beta}_2\sum_{i=1}^{n}x_i\hat{u}_i = \hat{\beta}_1\times 0 + \hat{\beta}_2\times 0 = 0$$

2.3.2 Decomposition of the variance of *y*

By definition

$$y_i = \hat{y}_i + \hat{u}_i \qquad (2\text{-}27)$$

Subtracting $\bar{y}$ on both sides of the previous expression (remember that $\bar{\hat{y}}$ is equal to $\bar{y}$), we have

$$y_i - \bar{y} = \hat{y}_i - \bar{\hat{y}} + \hat{u}_i$$

Squaring both sides:

$$\left[y_i - \bar{y}\right]^2 = \left[(\hat{y}_i - \bar{\hat{y}}) + \hat{u}_i\right]^2 = (\hat{y}_i - \bar{\hat{y}})^2 + \hat{u}_i^2 + 2\hat{u}_i(\hat{y}_i - \bar{\hat{y}})$$

Summing for all i:

$$\sum\left[y_i - \bar{y}\right]^2 = \sum(\hat{y}_i - \bar{\hat{y}})^2 + \sum\hat{u}_i^2 + 2\sum\hat{u}_i(\hat{y}_i - \bar{\hat{y}})$$

Taking into account the algebraic properties 1 and 4, the third term of the right hand side is equal to 0. Analytically,

$$\sum\hat{u}_i(\hat{y}_i - \bar{\hat{y}}) = \sum\hat{u}_i\hat{y}_i - \bar{\hat{y}}\sum\hat{u}_i = 0 \qquad (2\text{-}28)$$

Therefore, we have

$$\sum\left[y_i - \bar{y}\right]^2 = \sum(\hat{y}_i - \bar{\hat{y}})^2 + \sum\hat{u}_i^2 \qquad (2\text{-}29)$$

In words,

Total sum of squares (*TSS*) =

Explained sum of squares (*ESS*)+Residual sum of squares (*RSS*)

It must be stressed that it is necessary to use the relation (2-19) to assure that (2-28) is equal to 0. We must remember that (2-19) is associated to the first normal equation: that is to say, to the equation corresponding to the intercept. If there is no intercept in the fitted model, then in general the decomposition obtained will not be fulfilled (2-29).

This decomposition can be made with variances, by dividing both sides of (2-29) by n:

$$\frac{\sum(y_i - \bar{y})^2}{n} = \frac{\sum(\hat{y}_i - \bar{\hat{y}})^2}{n} + \frac{\sum\hat{u}_i^2}{n} \qquad (2\text{-}30)$$

In words,

Total variance=explained variance+ residual variance

2.3.3 Goodness of fit: Coefficient of determination (R^2)

A priori we have obtained the estimators minimizing the sum of square residuals.

Once the estimation has been done, we can see how well our sample regression line fits our data.

The measures that indicate how well the sample regression line fits the data are denominated *goodness of fit* measures. We are going to look at the most well-known measure, which is called *coefficient of determination* or the *R*-square (R^2). This measure is defined in the following way:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (2\text{-}31)$$

Therefore, R^2 is the proportion of the total sum of squares (*TSS*) which is explained by the regression (*ESS*): that is to say, which is explained by the model. We can also say that $100\,R^2$ is the percentage of the sample variation in *y* explained by *x*.

Alternatively, taking into account (2-29), we have:

$$\sum(\hat{y}_i - \bar{\hat{y}})^2 = \sum(y_i - \bar{y})^2 - \sum\hat{u}_i^2$$

Substituting in (2-31), we have

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{\sum(y_i - \bar{y})^2 - \sum\hat{u}_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum\hat{u}_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{RSS}{TSS} \qquad (2\text{-}32)$$

Therefore, R^2 is equal to 1 minus the proportion of the total sum of squares (*TSS*) that is non-explained by the regression (*RSS*).

According to the definition of R^2, the following must be accomplished

$$0 \le R^2 \le 1$$

Extreme cases:

a) If we have a perfect fit, then $\hat{u}_i = 0 \quad \forall i$. This implies that

$$\hat{y}_i = y_i \quad \forall i \quad \Rightarrow \sum(\hat{y}_i - \bar{\hat{y}})^2 = \sum(y_i - \bar{y})^2 \quad \Rightarrow R^2 = 1$$

b) If $\hat{y}_i = c \quad \forall i$, it implies that

$$\bar{\hat{y}} = c \quad \Rightarrow \hat{y}_i - \bar{\hat{y}} = c - c = 0 \quad \forall i \quad \Rightarrow \sum(\hat{y}_i - \bar{\hat{y}})^2 = 0 \quad \Rightarrow R^2 = 0$$

If R^2 is close to 0, it implies that we have a poor fit. In other words, very little variation in *y* is explained by *x*.

In many cases, a high R^2 is obtained when the model is fitted using time series data, due to the effect of a common trend. On the contrary, when we use cross sectional data a low value is obtained in many cases, but it does not mean that the fitted model is bad.

What is the relationship between the coefficient of determination and the coefficient of correlation studied in descriptive statistics? The coefficient of determination is equal to the squared coefficient of correlation, as can be seen in appendix 2.2:

$$r_{xy}^2 = R^2 \qquad (2\text{-}33)$$

(This equality is only valid in the simple regression model, but not in multiple regression model).

Example 2.2 Fulfilling algebraic implications and calculating R^2 in the consumption function

In column 2 of table 2.2, $\widehat{cons}_i$ is calculated; in columns 3, 4 and 5, you can see the fulfillment of algebraic implications 1, 3 and 4 respectively. The remainder of the columns shows the calculations to obtain

$$TSS = 42 \qquad ESS = 41.67 \qquad RSS = 42 - 41.67 = 0.33 \qquad R^2 = \frac{41.67}{42} = 0.992$$

or, alternatively, $R^2 = 1 - \dfrac{0.33}{42} = 0.992$

TABLE 2.2. Data and calculations to estimate the consumption function.

Observ.	$\widehat{cons}_i$	$\hat{u}_i$	$\hat{u}_i \times inc_i$	$\widehat{cons}_i \times \hat{u}_i$	$cons_i^2$	$(cons_i - \overline{cons})^2$	$\widehat{cons}_i^2$	$(\widehat{cons}_i - \overline{\widehat{cons}})^2$
1	4.83	0.17	1.00	0.81	25	16	23.36	17.36
2	7.33	-0.33	-3.00	-2.44	49	4	53.78	2.78
3	8.17	-0.17	-1.67	-1.36	64	1	66.69	0.69
4	9.83	0.17	2.00	1.64	100	1	96.69	0.69
5	10.67	0.33	4.33	3.56	121	4	113.78	2.78
6	13.17	-0.17	-2.67	-2.19	169	16	173.36	17.36
	54.00	**0.00**	**0.00**	**0.00**	**528**	**42**	**527.67**	**41.67**

2.3.4 Regression through the origin

If we force the regression line to pass through the point (0,0), we are constraining the intercept to be zero, as can be seen in figure 2.6. This is called a regression through the origin.

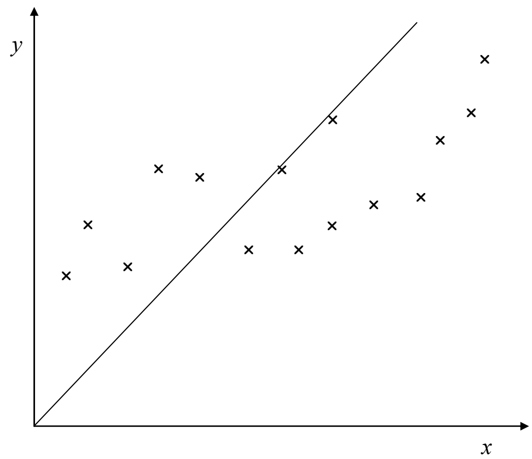


FIGURE 2.6. A regression through the origin.

Now, we are going to estimate a regression line through the origin. The fitted model is the following:

$$\tilde{y}_i = \tilde{\beta}_2 x_i \tag{2-34}$$

Therefore, we must minimize

$$\underset{\tilde{\beta}_2}{\text{Min}}\, S = \underset{\tilde{\beta}_2}{\text{Min}} \sum_{i=1}^{n} (y_i - \tilde{\beta}_2 x_i)^2 \tag{2-35}$$

To minimize S, we differentiate with respect to $\tilde{\beta}_2$ and equal to 0:

$$\frac{dS}{d\tilde{\beta}_2} = -2\sum_{i=1}^{n} (y_i - \tilde{\beta}_2 x_i) x_i = 0 \tag{2-36}$$

Solving for $\tilde{\beta}_2$

$$\tilde{\beta}_2 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} \tag{2-37}$$

Another problem with fitting a regression line through the origin is that the following generally happens:

$$\sum \left(y_i - \bar{y}\right)^2 \neq \sum (\hat{y}_i - \bar{\hat{y}})^2 + \sum \hat{u}_i^2$$

If the decomposition of the variance of y in two components (explained and residual) is not possible, then the R^2 is meaningless. This coefficient can take values that are negative or greater than 1 in the model without intercept.

To sum up, an intercept must always be included in the regressions, unless there are strong reasons against it supported by the economic theory.

2.4 Units of measurement and functional form

2.4.1 Units of Measurement

Changing the units of measurement (change of scale) in x

If x is multiplied/divided by a constant, $c\neq 0$, then the *OLS* slope is divided/multiplied by the same constant, c. Thus

$$\hat{y}_i = \hat{\beta}_1 + \left[\frac{\hat{\beta}_2}{c}\right](x_i \times c) \tag{2-38}$$

EXAMPLE 2.3

Let us suppose the following estimated consumption function, in which both variables are measured in thousands of euros:

$$\widehat{cons_i} = 0.2 + 0.85 \times inc_i \tag{2-39}$$

If we now express income in euros (multiplication by 1000) and call it *ince*, the fitted model with the new units of measurement of income would be the following:

$$\widehat{cons_i} = 0.2 + 0.00085 \times ince_i$$

As can be seen, changing the units of measurement of the explanatory variable does not affect the intercept.

Changing the units of measurement (change of scale) in y

If y is multiplied/divided by a constant, $c\neq 0$, then the *OLS* slope and intercept are both multiplied/divided by the same constant, c. Thus,

$$(\hat{y}_i \times c) = (\hat{\beta}_1 \times c) + (\hat{\beta}_2 \times c)x_i \tag{2-40}$$

EXAMPLE 2.4

If we express consumption in euros (multiplication by 1000) in model (2-39), and call it *conse*, the fitted model with the new units of measurement of consumption would be the following:

$$\widehat{conse_i} = 200 + 850 \times inc_i$$

Changing the origin

If one adds/subtracts a constant d to x and/or y, then the *OLS* slope is not affected. However, changing the origin of either x and/or y affects the intercept of the regression.

If one subtracts a constant d to x, the intercept will change in the following way:

$$\hat{y}_i = (\hat{\beta}_1 + \hat{\beta}_2 \times d) + \hat{\beta}_2 (x_i - d) \tag{2-41}$$

If one subtracts a constant d to y, the intercept will change in the following way:

$$\hat{y}_i - d = (\hat{\beta}_1 - d) + \hat{\beta}_2 x_i \tag{2-42}$$

EXAMPLE 2.5

Let us suppose that the average income is 20 thousand euros. If we define the variable $incd_i = inc_i - \overline{inc}$ and both variables are measured in thousands of euros, the fitted model with this change in the origin will be the following:

$$\widehat{cons_i} = (0.2 + 0.85 \times 20) + 0.85 \times (inc_i - 20) = 17.2 + 0.85 \times incd_i$$

EXAMPLE 2.6

Let us suppose that the average consumption is 15 thousands euros. If we define the variable $consd_i = cons_i - \overline{cons}$ and both variables are measured in euros, the fitted model with this change in the origin will be the following:

$$\widehat{cons_i} - 15 = 0.2 - 15 + 0.85 \times inc_i$$

that is to say,

$$\widehat{consd_i} = -14.8 + 0.85 \times inc_i$$

Note that R^2 is invariant to changes in the units of x and/or y, and also is invariant to the origin of the variables.

2.4.2 Functional Form

In many cases linear relationships are not adequate for economic applications. However, in the simple regression model we can incorporate nonlinearities (in variables) by appropriately redefining the dependent and independent variables.

Some definitions

Now we are going to look at some definitions of variation measures that will be useful in the interpretation of the coefficients corresponding to different functional forms. Specifically, we will look at the following: proportional change and change in logarithms.

The *proportional change* (or relative variation rate) between x_1 and x_0 is given by:

$$\frac{\Delta x_1}{x_0} = \frac{x_1 - x_0}{x_0} \tag{2-43}$$

Multiplying a proportional change by 100, we obtain a *proportional change in %*. That is to say:

$$100 \frac{\Delta x_1}{x_0} \% \tag{2-44}$$

The *change in logarithms* and *change in logarithms in %* between x_1 and x_0 are given by

$$\begin{gathered}\Delta\ln(x)=\ln(x_1)-\ln(x_0)\\ 100\Delta\ln(x)\%\end{gathered}\tag{2-45}$$

The *change in logarithms* is an approximation to the *proportional change*, as can be seen in appendix 2.3. This approximation is good when the proportional change is small, but the differences can be important when the proportional change is big, as can seen in table 2.3.

TABLE 2.3. Examples of proportional change and change in logarithms.

x_1	202	210	220	240	300
x_0	200	200	200	200	200
Proportional change in %	1%	5.0%	10.0%	20.0%	50.0%
Change in logarithms in %	1%	4.9%	9.5%	18.2%	40.5%

Elasticity is the ratio of the relative changes of two variables. If we use proportional changes, the elasticity of the variable *y* with respect to the variable *x* is given by

$$\varepsilon_{y/x}=\frac{\Delta y/y_0}{\Delta x/x_0}\tag{2-46}$$

If we use changes in logarithms and consider infinitesimal changes, then the elasticity of the variable *y* with respect to a variable *x* is given by

$$\varepsilon_{y/x}=\frac{dy/y}{dx/x}=\frac{d\ln(y)}{d\ln(x)}\tag{2-47}$$

In econometric models, elasticity is generally defined by using (2-47).

Alternative functional forms

The *OLS* method can also be applied to models in which the endogenous variable and/or the exogenous variable have been transformed. In the presentation of the model (2-1) we said that the exogenous variable and regressor were equivalent terms. But from now on, a regressor is the specific form in which an exogenous variable appears in the equation. For example, in the model

$$y=\beta_1+\beta_2\ln(x)+u$$

the exogenous variable is *x*, but the regressor is ln(*x*).

In the presentation of the model (2-1) we also said that the endogenous variable and the regressand were equivalent. But from now on, the regressand is the specific form in which an endogenous variable appears in the equation. For example, in the model

$$\ln(y)=\beta_1+\beta_2x+u$$

the endogenous variable is *y*, but the regressand is ln(*y*).

Both models are linear in the parameters, although they are not linear in the variable *x* (the first one) or in the variable *y* (the second one). In any case, if a model is linear in the parameters, it can be estimated by applying the *OLS* method. On the contrary, if a model is not linear in the parameters, iterative methods must be used in the estimation.

However, there are certain nonlinear models which, by means of suitable transformations, can become linear. These models are denominated linearizables.

Thus, on some occasions potential models are postulated in economic theory, such as the well-known Cobb-Douglas production function. A potential model with a unique explanatory variable is given by

$$y = e^{\beta_1} x^{\beta_2}$$

If we introduce the disturbance term in a multiplicative form, we obtain:

$$y = e^{\beta_1} x^{\beta_2} e^{u} \tag{2-48}$$

Taking natural logarithms on both sides of (2-48), we obtain a linear model in the parameters:

$$\ln(y) = \beta_1 + \beta_2 \ln(x) + u \tag{2-49}$$

On the contrary, if we introduce the disturbance term in an additive form, we obtain

$$y = e^{\beta_1} x^{\beta_2} + u$$

In this case, there is no transformation which allows this model to be turned into a linear model. This is a non-linearizable model.

Now we are going to consider some models with alternative functional forms, all of which are linear in the parameters. We will look at the interpretation of the coefficient $\hat{\beta}_2$ in each case.

a) Linear model

The $\hat{\beta}_2$ coefficient measures the effect of the regressor *x* on *y*. Let us look at this in detail. The observation *i* of the sample regression function is given according to (2-5) by

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i \tag{2-50}$$

Let us consider the observation *h* of the fitted model whereupon the value of the regressor and, consequently, of the regressand has changed with respect to (2-50):

$$\hat{y}_h = \hat{\beta}_1 + \hat{\beta}_2 x_h \tag{2-51}$$

Subtracting (2-51) from (2-50), we see that *x* has a linear effect on $\hat{y}$:

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x \tag{2-52}$$

where $\Delta\hat{y} = \hat{y}_i - \hat{y}_h$ and $\Delta x = x_i - x_h$

Therefore, $\hat{\beta}_2$ is the change in *y* (in the units in which *y* is measured) by a unit change of *x* (in the units in which *x* is measured).

For example, if income increases by 1 unit, consumption will increase by 0.85 units in the fitted function (2-39).

The linearity of this model implies that a one-unit change in x always has the same effect on y, regardless of the value of x considered.

***EXAMPLE 2.7* Quantity sold of coffee as a function of its price. Linear model**

In a marketing experiment[1] the following model has been formulated to explain the quantity sold of coffee per week (*coffqty*) as a function of the price of coffee (*coffpric*).

$$coffqty = \beta_1 + \beta_2 coffpric + u$$

The variable *coffpric* takes the value 1 for the usual price, and also 0.95 and 0.85 in two price actions whose effects are under investigation. This experiment lasted 12 weeks. *coffqty* is expressed in thousands of units and *coffpric* in French francs. Data appear in table 2.4 and in work file *coffee1*.

The fitted model is the following:

$$\widehat{coffqty} = 774.9 - 693.33 coffpric \quad R^2 = 0.95 \quad n = 12$$

TABLE 2.4. Data on quantities and prices of coffee.

week	*coffpric*	*coffqty*
1	1.00	89
2	1.00	86
3	1.00	74
4	1.00	79
5	1.00	68
6	1.00	84
7	0.95	139
8	0.95	122
9	0.95	102
10	0.85	186
11	0.85	179
12	0.85	187

Interpretation of the coefficient $\hat{\beta}_2$: if the price of coffee increases by 1 French franc, the quantity sold of coffee will decrease by 693.33 thousands of units. As the price of coffee is a small magnitude, the following interpretation is preferable: if the price of coffee increases by 1 cent of a French franc, the quantity sold will decrease by 6.93 thousands of units.

EXAMPLE 2.8 Explaining market capitalization of Spanish banks. Linear model

Using data from Bolsa de Madrid (*Madrid Stock Exchange*) on August 18, 1995 (file *bolmad95,* the first 20 observations), the following model has been estimated to explain the market capitalization of banks and financial institutions:

$$\widehat{marktval} = 29.42 + 1.219 bookval$$

$$R^2 = 0.836 \quad n = 20$$

where

- *marktval* is the capitalization the market value of a company. It is calculated by multiplying the price of the stock by the number of stocks issued.
- *bookval* is the book value or the net worth of the company. The book value is calculated as the difference between a company's assets and its liabilities.
- Data on *marktval* and *bookval* are expressed in millions of pesetas.

[1]The data of this exercise were obtained from a controlled marketing experiment in stores in Paris on coffee expenditure, as reported in A. C. Bemmaor and D. Mouchoux, "Measuring the Short-Term Effect of In-Store Promotion and Retail Advertising on Brand Sales: A Factorial Experiment", *Journal of Marketing Research*, 28 (1991), 202–14.

Interpretation of the coefficient β_2: if the book value of a bank increases by 1 million pesetas, the market capitalization of this bank will increase by 1.219 million of pesetas.

b) Linear-log model

A linear-log model is given by

$$y = \beta_1 + \beta_2 \ln(x) + u \tag{2-53}$$

The corresponding fitted function is the following:

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 \ln(x) \tag{2-54}$$

Taking first order differences in (2-54), and then multiplying and dividing the right hand side by 100, we have

$$\Delta\hat{y} = \frac{\hat{\beta}_2}{100} 100 \times \Delta\ln(x)\%$$

Therefore, if x increases by 1%, then $\hat{y}$ will increase by $(\hat{\beta}_2/100)$ units.

c) Log-linear model

A log-linear model is given by

$$\ln(y) = \beta_1 + \beta_2 x + u \tag{2-55}$$

The above model can be obtained by taking natural logs on both sides of the following model:

$$y = \exp(\beta_1 + \beta_2 x + u)$$

. For this reason, the model (2-55) is also called exponential.

The corresponding sample regression function to (2-55) is the following:

$$\widehat{\ln(y)} = \hat{\beta}_1 + \hat{\beta}_2 x \tag{2-56}$$

Taking first order differences in (2-56), and then multiplying both sides by 100, we have

$$100 \times \Delta\widehat{\ln(y)}\% \;=\; 100 \times \hat{\beta}_2 \Delta x$$

Therefore, if x increases by 1 unit, then $\hat{y}$ will increase by $100\,\hat{\beta}_2$ %.

d) Log-log model

The model given in (2-49) is a log-log model or, before the transformation, a potential model (2-48). This model is also called a constant elasticity model.

The corresponding fitted model to (2-49) is the following:

$$\widehat{\ln(y)} = \hat{\beta}_1 + \hat{\beta}_2 \ln(x) \tag{2-57}$$

Taking first order differences in (2-57), we have

$$\Delta\widehat{\ln(y)} = \hat{\beta}_2 \Delta\ln(x)$$

Therefore, if x increases by 1%, then $\hat{y}$ will increase by $\hat{\beta}_2$%. It is important to remark that, in this model, $\hat{\beta}_2$ is the estimated elasticity of y with respect to x, for any value of x and y. Consequently, in this model the elasticity is constant.

In annex 1 in a study case on the Engel curve for demand of dairy, six alternative functional forms are analyzed.

EXAMPLE 2.9 Quantity sold of coffee as a function of its price. Log- log model (Continuation example 2.7)

As an alternative to the linear model the following log-log model has been fitted:

$$\ln(\widehat{coffqty}) = 4.415 - 5.132\ln(coffpric) \quad R^2 = 0.90 \quad n = 12$$

Interpretation of the coefficient $\hat{\beta}_2$: if the price of coffee increases by 1%, the quantity sold of coffee will decrease by 5.13%. In this case $\hat{\beta}_2$ is the estimated demand/price elasticity.

EXAMPLE 2.10 Explaining market capitalization of Spanish banks. Log-log model (Continuation example 2.8)

Using data from example 2.8, the following log-log model has been estimated:

$$\widehat{\ln(marktval)} = 0.6756 + 0.938\ln(bookval)$$

$$R^2=0.928 \qquad n=20$$

Interpretation of the coefficient $\hat{\beta}_2$: if the book value of a bank increases by 1%, the market capitalization of this bank will increase by 0.938%. In this case $\hat{\beta}_2$ is the estimated market value/book value elasticity.

In table 2.5 and for the fitted model, the interpretation of $\hat{\beta}_2$ in these four models is shown. If we are considering the population model, the interpretation of β_2 is the same but taking into account that Δu must be equal to 0.

TABLE 2.5. Interpretation of $\hat{\beta}_2$ in different models.

Model	If x increases by	then y will increase by
linear	1 unit	$\hat{\beta}_2$ units
linear-log	1%	$(\hat{\beta}_2/100)$ units
log-linear	1 unit	$(100\hat{\beta}_2)$%
log-log	1%	$\hat{\beta}_2$%

2.5 Assumptions and statistical properties of *OLS*

We are now going to study the statistical properties of *OLS* estimators $\hat{\beta}_1$ and $\hat{\beta}_2$. But first we need to formulate a set of statistical assumptions. Specifically, the set of assumptions that we are going to formulate are called *classical linear model assumptions (CLM).* It is important to note that *CLM assumptions* are simple and that the *OLS* estimators have, under these assumptions, very good properties.

2.5.1 Statistical assumptions of the CLM in simple linear regression

a) Assumption on the functional form

1) The relationship between the regressand, the regressor and the random disturbance is linear in the parameters:

$$y = \beta_1 + \beta_2 x + u \tag{2-58}$$

The regressand and the regressors can be any function of the endogenous variable and the explanatory variables, respectively, provided that among regressors and regressand there is a linear relationship, i.e. the model is linear in the parameters. The additivity of the disturbance guarantees the linear relationship with the rest of the elements.

b) Assumptions on the regressor x

2) The values of x are fixed in repeated sampling:

According to this assumption, each observation of the regressor takes the same value for different samples of the regressand. This is a strong assumption in the case of the social sciences, where in general it is not possible to experiment. Data are obtained by observation, not by experimentation. It is important to remark that the results obtained using this assumption would remain virtually identical if we assume the repressors are stochastic, provided the additional assumption of independence between the regressors and the random disturbance is fulfilled. This alternative assumption can be formulated as:

2) The regressor x is distributed independently of the random disturbance.*

In any case, throughout this chapter and the following ones we will adopt assumption 2.

3) The regressor x does not contain measurement errors

This is an assumption that is not often fulfilled in practice, since the measurement instruments are unreliable in economy. Think, for example, of the multitude of errors that can be made in the collection of information, through surveys on families.

4) The sample variance of x is different from 0 and has a finite limit as n tends to infinity

Therefore, this assumption implies that

$$S_X^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n} \neq 0 \tag{2-59}$$

c) Assumptions on the parameters

5) The parameters β_1 and β_2 are fixed

If this assumption is not adopted, the regression model would be very difficult to handle. In any case, it may be acceptable to postulate that the model parameters are stable over time (if it is not a very long period) or space (if it is relatively limited).

d) Assumptions on the random disturbances

6) The disturbances have zero mean,

$$E(u_i) = 0, \quad i = 1,2,3,\ldots,n \tag{2-60}$$

This is not a restrictive assumption, since we can always use β_1 to normalize $E(u)$ to 0. Let us suppose, for example, that $E(u)=4$. We could then redefine the model in the following way:

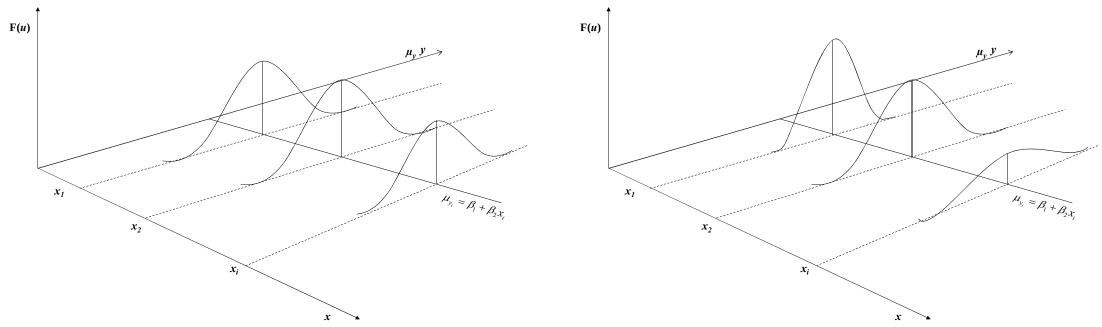
$$y=(\beta_1+4)+\beta_2 x+v$$

where $v=u-4$. Therefore, the expectation of the new disturbance, v, is 0 and the expectation of u has been absorbed by the intercept.

7) The disturbances have a constant variance

$$var(u_i)=\sigma^2 \qquad i=1,2,\cdots n \tag{2-61}$$

This assumption is called the *homoskedasticity* assumption**.** The word comes from the Greek: *homo* (equal) and *skedasticity* (spread). This means that the variation of y around the regression line is the same across the x values; that is to say, it neither increases or decreases as x varies. This can be seen in figure 2.7, part a), where disturbances are homoskedastic.



a) **b)**

FIGURE 2.7. Random disturbances: a) homoskedastic; b) heteroskedastic.

If this assumption is not satisfied, as happens in part b) of figure 2.7, the *OLS* regression coefficients are not efficient. Disturbances in this case are heteroskedastic (*hetero* means different).

8) The disturbances with different subscripts are not correlated with each other (no autocorrelation assumption):

$$E(u_i u_j)=0 \qquad i\neq j \tag{2-62}$$

That is, the disturbances corresponding to different individuals or different periods of time are not correlated with each other. This assumption of *no autocorrelation* or *no serial correlation*, like the previous one, is testable a posteriori. The transgression occurs quite frequently in models using time series data.

9) The disturbance u is normally distributed

Taking into account assumptions 6, 7 y 8, we have

$$u_i \sim NID(0,\sigma^2) \qquad i=1,2,\cdots,n \tag{2-63}$$

where *NID* states for *normally independently distributed*.

The reason for this assumption is that if u is normally distributed, so will y and the estimated regression coefficients, and this will be useful in performing tests of hypotheses and constructing confidence intervals for β_1 and β_2. The justification for the assumption depends on the Central Limit Theorem. In essence, this theorem states that, if a random variable is the composite result of the effects of an indefinite number of variables, it will have an approximately normal distribution even if its components do not, provided that none of them is dominant.

2.5.2 Desirable properties of the estimators

Before examining the properties of *OLS* estimators under the statistical assumptions of the *CLM*, we pose the following question: what are the desirable properties for an estimator?

Two desirable properties for an estimator are that it is unbiased and its variance is as small as possible. If this occurs, the inference process will be carried out in optimal conditions.

We will illustrate these properties graphically. Consider first the property of unbiasedness. In Figures 2.8 and 2.9 the density functions of two hypothetical estimators obtained by two different methods are shown.

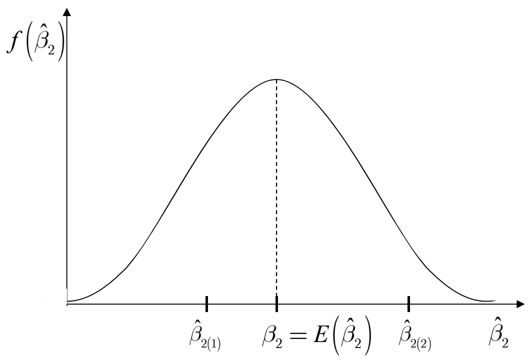


FIGURE 2.8. Unbiased estimator.

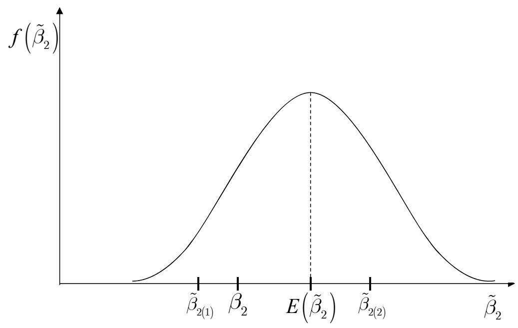


FIGURE 2.9. Biased estimator.

The estimator $\hat{\beta}_2$ is unbiased, i.e., its expected value is equal to the parameter that is estimated, β_2. The estimator $\hat{\beta}_2$ is a random variable. In each sample of *y*'s – the *x*'s are fixed in a repeated sample according to assumption 2- $\hat{\beta}_2$ taking a different value, but *on average* is equal to the parameter β_2, bearing in mind the infinite number of values $\hat{\beta}_2$ can take. In each sample of *y*'s a specific value of $\hat{\beta}_2$, that is to say, an *estimation* of $\hat{\beta}_2$ is obtained. In figure 2.8 two estimations of β_2 ($\hat{\beta}_{2(1)}$ and $\hat{\beta}_{2(2)}$) are obtained. The first estimate is relatively close to β_2, while the second one is much farther away. In any case, unbiasedness is a desirable property because it ensures that, on average, the estimator is centered on the parameter value.

The estimator $\tilde{\beta}_2$ is biased, since its expectation is not equal to β_2. The bias is precisely $E\left(\tilde{\beta}_2\right) - \beta_2$. In this case two hypothetical estimates, $\tilde{\beta}_{2(1)}$ and $\tilde{\beta}_{2(2)}$, are represented in figure 2.9. As can be seen $\tilde{\beta}_{2(1)}$ is closer to β_2 than the unbiased estimator $\hat{\beta}_{2(1)}$, but this is a matter of chance. In any case, when it is biased, it is not centered on the parameter value. An unbiased estimator will always be preferable, regardless of what happens in a specific sample, because it has no systematic deviation from the parameter value.

Another desirable property is efficiency. This property refers to the variance of the estimators. In figures 2.10 and 2.11 two hypothetical unbiased estimators, which are also called $\hat{\beta}_2$ and $\tilde{\beta}_2$, are represented. The first one has a smaller variance than the second one.

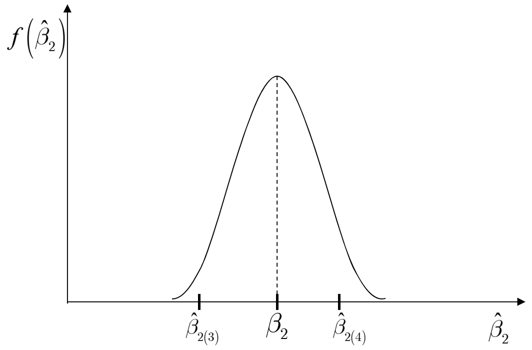


FIGURE 2.10. Estimator with small variance.

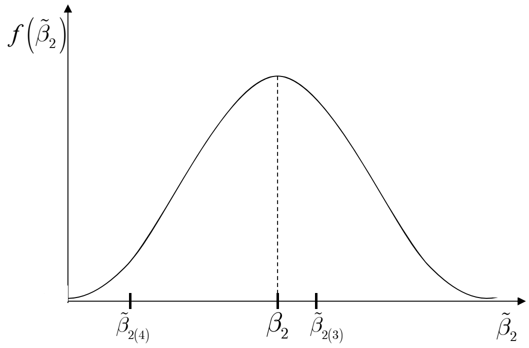


FIGURE 2.11. Estimator with big variance.

In both figures we have represented two estimates: $\hat{\beta}_{2(3)}$ and $\hat{\beta}_{2(4)}$ for the estimator with the smallest variance; and $\tilde{\beta}_{2(3)}$ and $\tilde{\beta}_{2(4)}$ for the estimator with the greatest variance. To highlight the role played by chance, the estimate that is closer to β_2 is precisely $\tilde{\beta}_{2(3)}$. In any case, it is preferable that the variance of the estimator is as small as possible. For example, when using the estimator $\hat{\beta}_2$ it is practically impossible that an estimate is so far from β_2 as it is in the case of $\hat{\beta}_2$, because the range of $\hat{\beta}_2$ is much smaller than the range of $\tilde{\beta}_2$

2.5.3 Statistical properties of *OLS* estimators

Under the above assumptions, the *OLS* estimators possess some ideal properties. Thus, we can say that the *OLS* are the best linear unbiased estimators.

Linearity and unbiasedness of the OLS

The *OLS* estimator $\hat{\beta}_2$ is unbiased. In appendix 2.4 we prove that $\hat{\beta}_2$ is an *unbiased* estimator using implicitly assumptions 3, 4 and 5, and explicitly assumptions 1, 2 and 6. In that appendix we can also see that $\hat{\beta}_2$ is a linear estimator using assumptions 1 and 2.

Similarly, one can show that the *OLS* estimator $\hat{\beta}_1$ is also unbiased. Remember that unbiasedness is a general property of the estimator, but in a given sample we may be “near” or “far” from the true parameter. In any case, its distribution will be centered at the population parameter.

Variances of the OLS estimators

Now we know that the sampling distribution of our estimator is centered around the true parameter. How spread out is this distribution? The variance (which is a measure of dispersion) of an estimator is an indicator of the accuracy of the estimator.

In order to obtain the variances of $\hat{\beta}_1$ and $\hat{\beta}_2$, assumptions 7 and 8 are needed, in addition to the first six assumptions. These variances are the following:

$$Var(\hat{\beta}_1) = \frac{\sigma^2 n^{-1} \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} \qquad Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} \tag{2-64}$$

Appendix 2.5 shows how the variance for $\hat{\beta}_2$ is obtained.

OLS estimators are BLUE

The *OLS* estimators have the least variance in the class of all linear and unbiased estimators. For this reason it is said that *OLS* estimators are the *best linear unbiased estimators (BLUE),* as illustrated in figure 2.12. This property is known as the Gauss–Markov theorem. For proof of this theorem assumptions 1-8 are used, as can be seen in appendix 2.6. This set of assumptions is known as the Gauss–Markov assumptions.

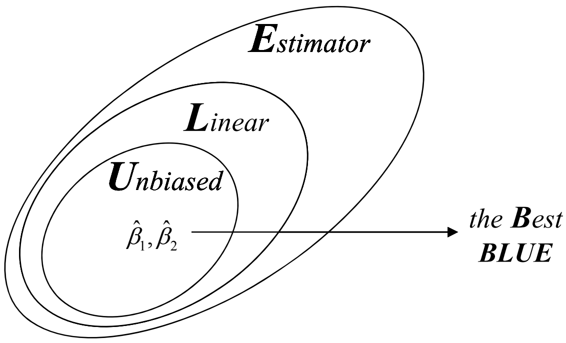


FIGURE 2.12. The *OLS* estimator is BLUE.

Estimating the disturbance variance and the variance of estimators

We do not know what the value of the disturbance variance, σ^2, is and thus we have to estimate it. But we cannot estimate it from the disturbances u_i, because they are not observable. Instead, we have to use the *OLS* residuals ($\hat{u}_i$).

The relation between disturbances and residuals is given by

$$\begin{aligned} \hat{u}_i &= y_i - \hat{y}_i = \beta_1 + \beta_2 x_i + u_i - \hat{\beta}_1 - \hat{\beta}_2 x_i \\ &= u_i - \left(\hat{\beta}_1 - \beta_1\right) - \left(\hat{\beta}_2 - \beta_2\right) x_i \end{aligned} \tag{2-65}$$

Hence $\hat{u}_i$ is not the same as u_i, although the difference between them- $\left(\hat{\beta}_1 - \beta_1\right) - \left(\hat{\beta}_2 - \beta_2\right) x_i$ - does have an expected value of zero. Therefore, a first estimator of σ^2 could be the residual variance:

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n} \tag{2-66}$$

However, this estimator is biased, essentially because it does not account for the two following restrictions that must be satisfied by the *OLS* residuals in the simple regression model:

$$\begin{cases} \sum_{i=1}^{n} \hat{u}_i = 0 \\ \sum_{i=1}^{n} x_i \hat{u}_i = 0 \end{cases} \tag{2-67}$$

One way to view these restrictions is the following: if we know n–2 of the residuals, we can get the other two residuals by using the restrictions implied by the normal equations.

Thus, there are only n–2 degrees of freedom in the *OLS* residuals, as opposed to n degrees of freedom in the disturbances. In the unbiased estimator of σ^2 shown below an adjustment is made taking into account the degrees of freedom:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2} \tag{2-68}$$

Under assumptions 1-8 (Gauss-Markov assumptions), and as can be seen in appendix 7, we obtain

$$E(\hat{\sigma}^2) = \sigma^2 \tag{2-69}$$

If $\hat{\sigma}^2$ is plugged into the variance formulas, we then have unbiased estimators of *var*($\hat{\beta}_1$) and *var*($\hat{\beta}_2$)

The natural estimator of σ is $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ and is called the *standard error of the regression*. The square root of the variance of $\hat{\beta}_2$ is called the *standard deviation* of $\hat{\beta}_2$, that is to say,

$$sd(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \tag{2-70}$$

Therefore, its natural estimator is called the *standard error* of $\hat{\beta}_2$:

$$se(\hat{\beta}_2) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \tag{2-71}$$

Note that $se(\hat{\beta}_2)$, due to the presence of the estimator $\hat{\sigma}$ in (2-71), is a random variable as is $\hat{\beta}_2$. The standard error of any estimate gives us an idea of how precise the estimator is.

Consistency of OLS and other asymptotic properties

Sometimes it is not possible to obtain an unbiased estimator. In any case *consistency* is a minimum requirement for an estimator. According to an intuitive approach, consistency means that as $n \to \infty$, the density function of the estimator collapses to the parameter value. This property can be expressed for the estimator $\hat{\beta}_2$ as:

$$\underset{n\to\infty}{\text{plim}}\, \hat{\beta}_2 = \beta_2 \tag{2-72}$$

where plim means probability limit. In other words, $\hat{\beta}_2$ converges in probability to β_2.

Note that the properties of unbiasedness and consistency are conceptually different. The property of unbiasedness can hold for any sample size, whereas consistency is strictly a large-sample property or an *asymptotic property*.

Under assumptions 1 through 6, the *OLS* estimators, $\hat{\beta}_1$ and $\hat{\beta}_2$, are consistent. The proof for $\hat{\beta}_2$ can be seen in appendix 2.8.

Other asymptotic properties of $\hat{\beta}_1$ and $\hat{\beta}_2$: Under the Gauss-Markov assumptions 1 through 8, $\hat{\beta}_1$ and $\hat{\beta}_2$ are *asymptotically normally distributed* and also *asymptotically efficient* within the class of consistent and asymptotically normal estimators.

OLS estimators are maximum likelihood estimators (ML) and minimum variance unbiased estimators (MVUE)

Now we are going to introduce the assumption 9 on normality of the disturbance *u*. The set of assumptions 1 through 9 is known as the *classical linear model* (*CLM*) assumptions.

Under the *CLM* assumptions, the *OLS* estimators are *also maximum likelihood estimators* (*ML*), as can be seen in appendix 2.8.

On the other hand, under *CLM* assumptions, *OLS* estimators are not only *BLUE*, but are the *minimum variance unbiased estimators (MVUE)*. This means that *OLS* estimators have the smallest variance among all unbiased, linear o nonlinear, estimators, as can be seen in figure 2.13. Therefore, we have no longer to restrict our comparison to estimators that are linear in the y_i's.

What also happens is that any linear combination of $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \cdots, \hat{\beta}_k$ is also normally distributed, and any subset of the $\hat{\beta}_j$'s has a joint normal distribution.

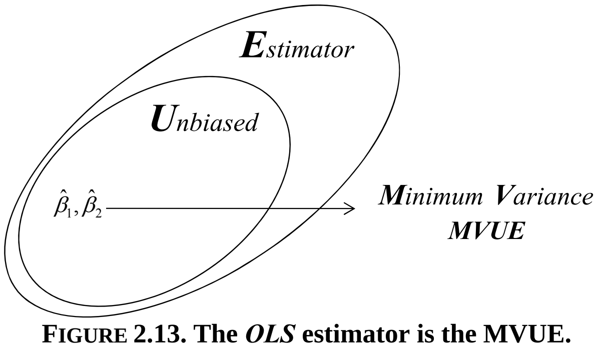


FIGURE 2.13. The *OLS* estimator is the MVUE.

In conclusion, we have seen that the *OLS* estimator has very desirable properties when the statistical basic assumptions are met.

Exercises

Exercise 2.1 The following model has been formulated to explain the annual sales (*sales*) of manufacturers of household cleaning products based as a function of a relative price index (*rpi*):

$$sales = \beta_1 + \beta_2 rpi + u$$

where the variable *sales* is expressed in a thousand million euros and *rpi* is an index obtained as the ratio between the prices of each firm and prices of the firm 1 of the sample). Thus, the value 110 in firm 2 indicates its price is 10% higher than in firm1.

Data on ten manufacturers of household cleaning products are the following:

firm	*sales*	*rpi*
1	10	100
2	8	110
3	7	130
4	6	100
5	13	80
6	6	80
7	12	90
8	7	120
9	9	120
10	15	90

a) Estimate β_1 and β_2 by *OLS*.
b) Calculate the *RSS*.
c) Calculate the coefficient of determination.
d) Check that the algebraic implications 1, 3 and 4 are fulfilled in the *OLS* estimation.

Exercise 2.2 To study the relationship between fuel consumption (y) and flight time (x) of an airline, the following model is formulated:

$$y = \beta_1 + \beta_2 x + u$$

where y is expressed in thousands of pounds and x in hours, using fractions of an hour as units of low-order decimal.

The statistics of "Flight times and fuel consumption" of an airline provides data on flight times and fuel consumption of 24 different trips made by an aircraft of the company. From these data the following statistics were drawn:

$$\sum y_i = 219.719; \ \sum x_i = 31.470; \ \sum x_i^2 = 51.075;$$

$$\sum x_i y_i = 349.486; \ \sum y_i^2 = 2396.504$$

a) Estimate β_1 and β_2 by *OLS*.
b) Decompose the variance of the variable y invariance explained by the regression and residual variance.
c) Calculate the coefficient of determination.
d) Estimate total consumption, in thousands of pounds, for a flight program consisting of 100 half-hour flights, 200 one hour flights and 100 two hours flights.

Exercise 2.3 An analyst formulates the following model:

$$y = \beta_1 + \beta_2 x + u$$

Using a given sample, the following results were obtained:

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n} = 20 \qquad \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} = 10 \qquad \bar{y}=8 \quad \bar{x}=4 \quad \hat{\beta}_2 = 3$$

Are these results consistent? Explain your answer.

Exercise 2.4 An econometrician has estimated the following model with a sample of five observations:

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

Once the estimation has been made, the econometrician loses all information except what appears in the following table:

Obs.	x_i	$\hat{u}_t$
1	1	2
2	3	-3
3	4	0
4	5	¿?
5	6	¿?

With the above information the econometrician must calculate the residual variance. Do it for them.

Exercise 2.5 Given the model

$$y_i = \beta_1 + \beta_2 x_i + u_i \qquad 1 = 1, 2, \ldots, n$$

the following results with a sample size of 11 were obtained:

$$\sum_{i=1}^{n} x_i = 0 \qquad \sum_{i=1}^{n} y_i = 0 \qquad \sum_{i=1}^{n} x_i^2 = B \qquad \sum_{i=1}^{n} y_i^2 = E \qquad \sum_{i=1}^{n} x_i y_i = F$$

a) Estimate β_2 and β_1

b) Calculate the sum of square residuals.

c) Calculate the coefficient of determination.

d) Calculate the coefficient of determination under the assumption that $2F^2 = BE$

Exercise 2.6 Company A is dedicated to mounting prefabricated panels for industrial buildings. So far, the company has completed eight orders, in which the number of square meters of panels and working hours employed in the assembly are as follows:

Number of square meters (thousands)	**Number of hours**
4	7400
6	9800
2	4600
8	12200
10	14000
5	8200
3	5800
12	17000

Company A wishes to participate in a tender to mount 14000m^2 of panels in a warehouse, for which a budget is required.

In order to prepare the budget, we know the following:

a) The budget must relate exclusively to the assembly costs, since the material is already provided.

b) The cost of the working hour for Company A is 30 euros.

c) To cover the remaining costs, Company A must charge 20% on the total cost of labor employed in the assembly.

Company A is interested in participating in the tender with a budget that only covers the costs. Under these conditions, and under the assumption that the number of hours worked is a linear function of the number of square meters of panels mounted, what would be the budget provided by company A?

Exercise 2.7 Consider the following equalities:

1. $E[u] = 0$.
2. $E[\hat{u}] = 0$.
3. $\bar{u} = 0$.
4. $\bar{\hat{u}} = 0$.

In the context of the basic linear model, indicate whether each of the above equalities are true or not. Justify your answer.

Exercise 2.8 The parameters β_1 and β_2 of the following model have been estimated by *OLS*:

$$y = \beta_1 + \beta_2 x + u$$

A sample of size 3 was used and the observations for x_i were $\{1,2,3\}$. It is also known that the residual for the first observation was 0.5.

From the above information, is it possible to calculate the sum of squared residuals and obtain an estimate of σ^2? If so, carry out the corresponding calculations.

Exercise 2.9 The following data are available to estimate a relationship between y and x:

y	x
-2	-2
-1	0
0	1
1	0
2	1

a) Estimate the parameters α and β of the following model by *OLS*:

$$y = \alpha + \beta x + \varepsilon$$

b) Estimate var(ε_i).

c) Estimate the parameters γ and δ of the following model by *OLS*:

$$x = \gamma + \delta y + \upsilon$$

d) Are the two fitted regression lines the same? Explain the result in terms of the least-square method.

Exercise 2.10 Answer the following questions:

a) One researcher, after performing the estimation of a model by *OLS*, calculates $\sum \hat{u}_i$ and verifies that it is not equal to 0. Is this possible? Are there any conditions in which this may occur?

b) Obtain an unbiased estimator of σ^2, indicating the assumption you have to use. Explain your answer.

Exercise 2.11 In the context of a linear regression model

$$y = \beta_1 + \beta_2 x + u$$

a) Indicate whether the following equalities are true. If so explain why

$$\bar{u} = \frac{\sum_{i=1}^{n} u_i}{n} = 0; \qquad \bar{\hat{u}} = \frac{\sum_{i=1}^{n} \hat{u}_i}{n} = 0; \qquad E\left[x_i u_i\right] = 0; \qquad E\left[u_i\right] = 0;$$

b) Establish the relationship between the following expressions:

$$E\left[u_i^2\right] = \sigma^2; \qquad \hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-k}$$

Exercise 2.12 Answer the following questions:

a) Define the probabilistic properties of *OLS* estimator under the basic assumptions of the linear regression model. Explain your answer.

b) What happens with the estimation of the linear regression model if the sample variance of the explanatory variable is null? Explain your answer.

Exercise 2.13 A researcher believes that the relationship between consumption (*cons*) and disposable income (*inc*) should be strictly proportional, and, therefore formulates the following model:

$$cons = \beta_2 inc + u$$

a) Derive the formula for estimating β_2.

b) Derive the formula for estimating σ^2.

c) In this model, is $\sum_{i=1}^{n} \hat{u}_i$ equal to 0?

Exercise 2.14 In the context of the simple linear regression model

$$y = \beta_1 + \beta_2 x + u$$

a) What assumptions must be met for the *OLS* estimators to be unbiased?

b) What assumptions are required for the estimators with variances which are the lowest within the set of linear unbiased estimators?

Exercise 2.15 In statistical terms it is often usual to make statements like the following:

"Let $x_2, \ldots x_n$, be a random sample of size *n* drawn from a population $N(\alpha,\sigma)$"

a) Express the previous statement with econometric language by introducing a disturbance term.

b) Derive the formula for estimating α.

c) Derive the formula for estimating σ^2.

d) In this model, is $\sum_{i=1}^{n} \hat{u}_i$ equal to 0?

Exercise 2.16 The following model relates expenditure on education (*exped*) and disposable income (*inc*):

$$exped=\beta_1+\beta_2 inc+u$$

Using the information obtained from a sample of 10 families, the following results have been obtained:

$$\overline{exped}=7 \quad \overline{inc}=50 \quad \sum_{i=1}^{10} inc_i^2=30.650 \quad \sum_{i=1}^{10} exped_i^2=622 \quad \sum_{i=1}^{10} inc_i \times exped_i=4.345$$

a) Estimate β_1 and β_2 by *OLS*.
b) Estimate the expenditure on education/ income elasticity for the sample average family.
c) Decompose the variance of the endogenous variable invariance explained by the regression and residual variance.
d) Calculate the coefficient of determination.
e) Estimate the variance of the disturbances.

Exercise 2.17 Given the population model

$$y_i=3+2x_i+u_i \qquad i=1, 2, 3$$

where $x_i=\{1,2,3\}$:

a) Using $N(0,1)$ random number, generate 15 samples of u_1, u_2 and u_3, and obtain the corresponding values of y
b) Carry out the corresponding estimates of β_1 and β_2 in the model:

$$y=\beta_1+\beta_2 x+u$$

c) Compare the sample means and variances of $\hat{\beta}_1$ y $\hat{\beta}_2$ with their population expectations and variances.

Exercise 2.18 Based on the information supplied in exercise 2.17, and the 15 pairs of estimates of β_1 and β_2 obtained:

a) Calculate the residuals corresponding to each of the estimates.
b) Show why the residuals always take the form

$$\hat{u}_1=-\hat{u}_2$$
$$\hat{u}_3=0$$

Exercise 2.19 The following model was formulated to explain sleeping time (*sleep*) as a function of time devoted to paid work (*paidwork*):

$$sleep=\beta_1+\beta_2 paidwork+u$$

where *sleep* and *paidwork* are measured in minutes per day.

Using a random subsample extracted from the file *timuse03*, the following results were obtained

$$\widehat{sleep_i}=550.17-0.1783 paidwork$$

$$R^2=0.2539 \qquad n=62$$

a) Interpret the coefficient on *paidwork*.
b) What is, on average, the predicted increment in sleep if time devoted to paid work decreases in an hour per day?
c) How much of the variation in *sleep* is explained by *paidwork*?

Exercise 2.20 Quantifying happiness is not an easy task. Researchers at the Gallup World Poll went about it by surveying thousands of respondents in 155 countries, between 2006 and 2009, in order to measure two types of well-being. They asked respondents to report on the overall satisfaction with their lives, and ranked their answers using a "life evaluation" score from 1 to 10. To explain the overall satisfaction (*stsfglo*) the following model has been formulated, where observations are averages of the variables in each country:

$$stsfglo = \beta_1 + \beta_2 lifexpec + u$$

where *lifexpec* is life expectancy at birth: that is to say, number of years a newborn infant is expected to live.

Using the work file *HDR2010*, the fitted model obtained is the following:

$$\widehat{stsfglo} = -1.499 + 0.1062 lifexpec$$

$$R^2 = 0.6135 \qquad n = 144$$

a) Interpret the coefficient on *lifexpec*.
b) What would be the average overall satisfaction for a country with 80 years of life expectancy at birth?
c) What should be the life expectancy at birth to obtain a global satisfaction equal to six?

Exercise 2.21 In economics, Research and Development intensity (or simply R&D intensity) is the ratio of a company's investment in Research and Development compared to its sales.

For the estimation of a model which explains R&D intensity, it is necessary to have an appropriate database. In Spain it is possible to use the Survey of Entrepreneurial Strategies (*Encuesta sobre Estrategias Empresariales*) produced by the Ministry of Industry. This survey, on an annual basis, provides in-depth knowledge of the industrial sector's evolution over time by means of multiple data concerning business development and company decisions. This survey is also designed to generate microeconomic information that enables econometric models to be specified and tested. As far as its coverage is concerned, the reference population of this survey is companies with 10 or more workers from the manufacturing industry. The geographical area of reference is Spain, and the variables have a timescale of one year. One of the most outstanding characteristics of this survey is its high degree of representativeness.

Using the work file *rdspain*, which is a dataset consisting of 1,983 Spanish firms for 2006, the following equation is estimated to explain expenditures on research and development (*rdintens*):

$$\widehat{rdintens} = -2.639 + 0.2123 \ln(sales)$$

$$R^2 = 0.0350 \qquad n = 1983$$

where *rdintens* is expressed as a percentage of sales, and *sales* are measured in millions of euros.

a) Interpret the coefficient on ln(*sales*).
b) If *sales* increase by 50%, what is the estimated percentage point change in *rdintens*?
c) What percentage of the variation of *rdintens* is explained by *sales*? Is it large? Justify your answer.

Exercise 2.22 The following model has been formulated to explain MBA graduated salary (*salMBAgr*) as a function of tuition fees (*tuition*)

$$salMBAgr = \beta_1 + \beta_2 tuition + u$$

where *salMBApr* is the median annual salary in dollars for students enrolled in 2,010 of the 50 best American business schools and *tuition* is tuition fees including all required fees for the entire program (but excluding living expenses).

Using the data in *MBAtui10*, this model is estimated:

$$\widehat{salMBAgr_i} = 54242 + 0.4313 tuition_i$$

$$R^2=0.4275 \qquad n=50$$

a) What is the interpretation of the intercept?

b) What is the interpretation of the slope coefficient?

c) What is the predicted value of *salMBAgr* for a graduate student who paid 110000$ *tuition* fees in a 2 years *MBA*?

Exercise 2.23 Using a subsample of the Structural Survey of Wages (*Encuesta de estructura salarial*) for Spain in 2006 (*wage06sp*), the following model is estimated to explain wages:

$$\widehat{\ln(wage)} = 1.918 + 0.0527 educ$$

$$R^2=0.2445 \qquad n=50$$

where *educ* (education) is measured in years and *wage* in euros per hour.

a) What is the interpretation of the coefficient on *educ*?

b) How many more years of education are required to have a 10% higher wage?

c) Knowing that $\overline{educ} = 10.2$, calculate the wage/education elasticity. Do you consider this elasticity to be high or low?

Exercise 2.24 Using data from the Spanish economy for the period 1954-2010 (work file *consump*), the Keynesian consumption function is estimated:

$$\widehat{conspc_t} = -288 + 0.9416 incpc_t$$

$$R^2=0.994 \qquad n=57$$

where consumption (*conspc*) and disposable income (*incpc*) are expressed in constant euros per capita, taking 2008 as reference year.

a) What is the interpretation of the intercept? Comment on the sign and magnitude of the intercept.

b) Interpret the coefficient on *incpc*. What is the economic meaning of this coefficient?

c) Compare the marginal propensity to consume with the average propensity to consume at the sample mean $(\overline{conspc} = 8084, \overline{incpc} = 8896)$. Comment on the result obtained.

d) Calculate the consumption/income elasticity for the sample mean.

Annex 2.1 Case study: Engel curve for demand of dairy products

The Engel curve shows the relationship between the various quantities of a good that a consumer is willing to purchase at varying income levels.

In a survey with 40 households, data were obtained on expenditure on dairy products and income. These data appear in table 2.6 and in work file *demand*. In order to avoid distortions due to the different size of households, both consumption and income have been expressed in terms of per capita. The data are expressed in thousands of euros per month.

There are several demand models. We will consider the following models: linear, inverse, semi-logarithmic, potential, exponential and inverse exponential. In the first three models, the regressand of the equation is the endogenous variable, whereas in the last three the regressand is the natural logarithm of the endogenous variable.

In all the models we will calculate the marginal propensity to expenditure, as well as the expenditure/income elasticity.

TABLE 2.6. Expenditure on dairy products (*dairy*), disposable income (*inc*) in terms of *per capita*. Unit: euros per month.

household	*dairy*	*inc*	*household*	*dairy*	*inc*
1	8.87	1.250	21	16.20	2.100
2	6.59	985	22	10.39	1.470
3	11.46	2.175	23	13.50	1.225
4	15.07	1.025	24	8.50	1.380
5	15.60	1.690	25	19.77	2.450
6	6.71	670	26	9.69	910
7	10.02	1.600	27	7.90	690
8	7.41	940	28	10.15	1.450
9	11.52	1.730	29	13.82	2.275
10	7.47	640	30	13.74	1.620
11	6.73	860	31	4.91	740
12	8.05	960	32	20.99	1.125
13	11.03	1.575	33	20.06	1.335
14	10.11	1.230	34	18.93	2.875
15	18.65	2.190	35	13.19	1.680
16	10.30	1.580	36	5.86	870
17	15.30	2.300	37	7.43	1.620
18	13.75	1.720	38	7.15	960
19	11.49	850	39	9.10	1.125
20	6.69	780	40	15.31	1.875

Linear model

The linear model for demand of dairy products will be the following:

$$dairy = \beta_1 + \beta_2 inc + u \tag{2-73}$$

The marginal propensity indicates the change in expenditure as income varies and it is obtained by differentiating the expenditure with respect to income in the demand equation. In the linear model the marginal propensity of the expenditure on dairy is given by

$$\frac{d\ dairy}{d\ inc} = \beta_2 \tag{2-74}$$

In other words, in the linear model the marginal propensity is constant and, therefore, it is independent of the value that takes the income. It has the disadvantage of not being adapted to describe the behavior of the consumers, especially when there are important differences in the household income. Thus, it is unreasonable that the marginal propensity of expenditure on dairy products is the same in a low-income family and a family with a high income. However, if the variation of the income is not very high in the sample, a linear model can be used to describe the demand of certain goods.

In this model the expenditure/income elasticity is the following:

$$\varepsilon_{dairy/inc}^{linear} = \frac{d\ dairy}{d\ inc}\frac{inc}{dairy} = \beta_2 \frac{inc}{dairy} \tag{2-75}$$

Estimating the model (2-73) with the data from table 2.6, we obtain

$$\widehat{dairy} = 4.012 + 0.005288 \times inc \qquad R^2 = 0.4584 \tag{2-76}$$

Inverse model

In an inverse model there is a linear relationship between the expenditure and the inverse of income. Therefore, this model is directly linear in the parameters and it is expressed in the following way:

$$dairy = \beta_1 + \beta_2 \frac{1}{inc} + u \tag{2-77}$$

The sign of the coefficient β_2 will be negative if the income is correlated positively with the expenditure. It is easy to see that, when the income tends towards infinite, the expenditure tends towards a limit which is equal to β_1. In other words, β_1 represents the maximum consumption of this good.

In figure 2.14, we can see a double representation of the population function corresponding to this model. In the first one, the relationship between the dependent variable and explanatory variable has been represented. In the second one, the relationship between the regressand and regressor has been represented. The second function is linear as can be seen in the figure.

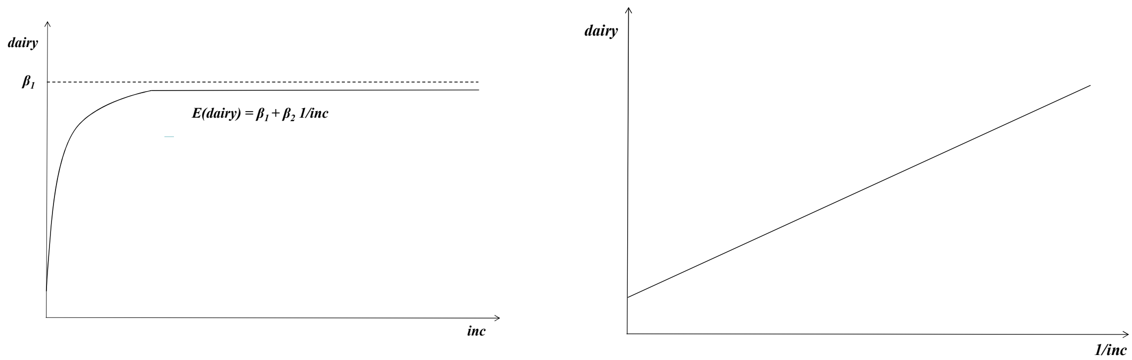


Figure 2.14. The inverse model.

In the inverse model, the marginal propensity to expenditure is given by

$$\frac{d\ dairy}{d\ inc} = -\beta_2 \frac{1}{(inc)^2} \tag{2-78}$$

According to (2-78), the marginal propensity is inversely proportional to the square of the income level.

On the other hand, the elasticity is inversely proportional to the product of expenditure and income, as can be seen in the following expression:

$$\varepsilon_{dairy/inc}^{inv} = \frac{d\ dairy}{d\ inc}\frac{inc}{dairy} = -\beta_2 \frac{1}{inc \times dairy} \tag{2-79}$$

Estimating the model (2-77) with the data of table 2.6, we obtain

$$\widehat{dairy} = 18.652 - 8702 \frac{1}{inc} \qquad R^2 = 0.4281 \tag{2-80}$$

In this case the coefficient $\hat{\beta}_2$ does not have an economic meaning.

Linear-log model

This model is denominated linear-log model, because the expenditure is a linear function of the logarithm of income, that is to say,

$$dairy = \beta_1 + \beta_2 \ln(inc) + u \tag{2-81}$$

In this model the marginal propensity to expenditure is given by

$$\frac{d\ dairy}{d\ inc} = \frac{d\ dairy}{d\ inc}\frac{inc}{inc} = \frac{d\ dairy}{d\ln(inc)}\frac{1}{inc} = \beta_2 \frac{1}{inc} \tag{2-82}$$

and the elasticity expenditure/income is given by

$$\varepsilon_{dairy/inc}^{lin\text{-}\log} = \frac{d\ dairy}{d\ inc}\frac{inc}{dairy} = \frac{d\ dairy}{d\ \ln(inc)}\frac{1}{dairy} = \beta_2 \frac{1}{dairy} \tag{2-83}$$

The marginal propensity is inversely proportional to the level of income in the linear-log model, while the elasticity is inversely proportional to the level of expenditure on dairy products.

In figure 2.15, we can see a double representation of the population function corresponding to this model.

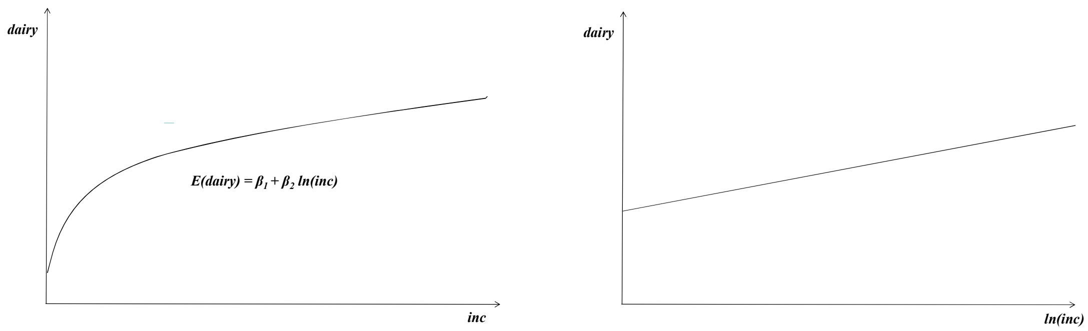


Figure 2.15. The linear-log model.

Estimating the model (2-81) with the data from table 2.6, we obtain

$$\widehat{dairy} = -41.623 + 7.399 \times \ln(inc) \quad R^2 = 0.4567 \tag{2-84}$$

The interpretation of $\hat{\beta}_2$ is the following: if the income increases by 1%, the demand of dairy products will increase by 0.07399 euros.

Log-log model or potential model

This exponential model is defined in the following way:

$$dairy = e^{\beta_1} inc^{\beta_2} e^{u} \tag{2-85}$$

This model is not linear in the parameters, but it is linearizable by taking natural logarithms, and the following is obtained:

$$\ln(dairy) = \beta_1 + \beta_2 \ln(inc) + u \tag{2-86}$$

This model is also called log-log model, because this is the structure of the corresponding linearized model.

In this model the marginal propensity to expenditure is given by

$$\frac{d\ dairy}{d\ inc} = \beta_2 \frac{dairy}{inc} \tag{2-87}$$

In the log-log model, the elasticity is constant. Therefore, if the income increases by 1%, the expenditure will increase by β_2%, since

$$\varepsilon_{dairy/inc}^{log\text{-}log} = \frac{d\ dairy}{d\ inc}\frac{inc}{dairy} = \frac{d\ \ln(dairy)}{d\ \ln(inc)} = \beta_2 \tag{2-88}$$

In figure 2.16, we can see a double representation of the population function corresponding to this model.

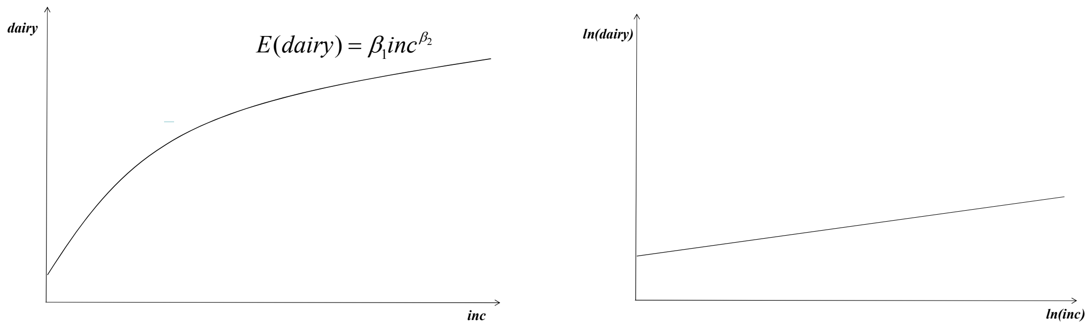


Figure 2.16. The log-log model.

Estimating the model (2-86) with the data from table 2.6, we obtain

$$\widehat{\ln(dairy)} = -2.556 + 0.6866 \times \ln(inc) \quad R^2 = 0.5190 \tag{2-89}$$

In this case $\hat{\beta}_2$ is the expenditure/income elasticity. Its interpretation is the following: if the income increases by 1%, the demand of dairy products will increase by 0.68%.

Log-linear or exponential model

This exponential model is defined in the following way:

$$dairy = \exp(\beta_1 + \beta_2 inc + u) \tag{2-90}$$

By taking natural logarithms on both sides of (2-90), we obtain the following model that is linear in the parameters:

$$\ln(dairy) = \beta_1 + \beta_2 inc + u \tag{2-91}$$

In this model the marginal propensity to expenditure is given by

$$\frac{d\ dairy}{d\ inc} = \beta_2 dairy \tag{2-92}$$

In the exponential model, unlike other models seen previously, the marginal propensity increases when the level of expenditure does. For this reason, this model is adequate to describe the demand of luxury products. On the other hand, the elasticity is proportional to the level of income:

$$\varepsilon_{dairy/inc}^{exp} = \frac{d\ dairy}{d\ inc}\frac{inc}{dairy} = \frac{d\ \ln(dairy)}{d\ inc} inc = \beta_2 inc \tag{2-93}$$

In figure 2.17, we can see a double representation of the population function corresponding to this model.

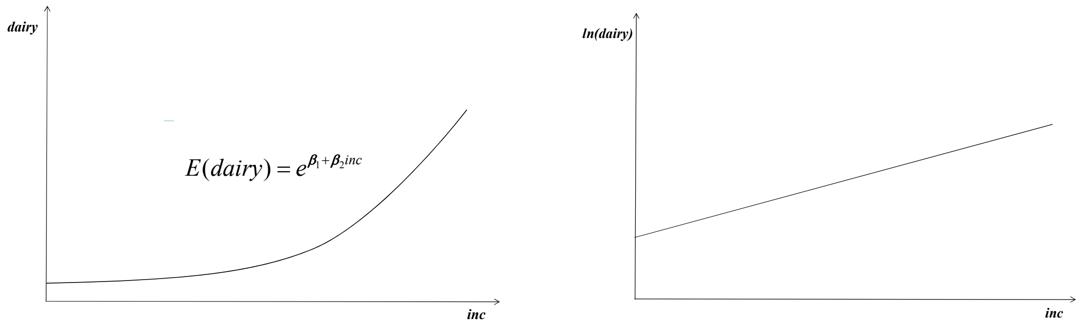


Figure 2.17. The log-linear model.

Estimating the model (2-91) with the data from table 2.6, we obtain

$$\widehat{\ln(dairy)} = 1.694 + 0.00048 \times inc \quad R^2 = 0.4978 \tag{2-94}$$

The interpretation of $\hat{\beta}_2$ is the following: if the income increases by a euro the demand of dairy products will increase by 0.048%.

Inverse exponential model

The inverse exponential model, which is a mixture of the exponential model and the inverse model, has properties that make it suitable for determining the demand for products in which there is a saturation point. This model is given by

$$dairy = \exp(\beta_1 + \beta_2 \frac{1}{inc} + u) \tag{2-95}$$

By taking natural logarithms on both sides of (2-95), we obtain the following model that is linear in the parameters:

$$\ln(dairy) = \beta_1 + \beta_2 \frac{1}{inc} + u \tag{2-96}$$

In this model the marginal propensity to expenditure is given by

$$\frac{d\ dairy}{d\ inc} = -\beta_2 \frac{dairy}{(inc)^2} \tag{2-97}$$

and the elasticity by

$$\varepsilon_{dairy/inc}^{invexp} = \frac{d\ dairy}{d\ inc}\frac{inc}{dairy} = \frac{d\ \ln(dairy)}{d\ inc} inc = -\beta_2 \frac{1}{inc} \tag{2-98}$$

Estimating the model (2-96) with the data from table 2.6, we obtain

$$\widehat{\ln(dairy)} = 3.049 - 822.02\frac{1}{inc} \qquad R^2 = 0.5040 \tag{2-99}$$

In this case, as in the inverse model, the coefficient $\hat{\beta}_2$ does not have an economic meaning.

In table 2.7, the results of the marginal propensity, the expenditure/income elasticity and R^2 in the six fitted models are shown

Table 2.7. Marginal propensity, expenditure/income elasticity and R^2 in the fitted models.

Model	*Marginal propensity*	*Elasticity*	R^2
Linear	$\hat{\beta}_2$=0.0053	$\hat{\beta}_2\frac{\overline{inc}}{\overline{dairy}}$=0.6505	0.4440
Inverse	$-\hat{\beta}_2\frac{1}{\left[\overline{inc}\right]^2}$=0.0044	$-\hat{\beta}_2\frac{1}{\overline{dairy\times inc}}$=0.5361	0.4279
Linear-log	$\hat{\beta}_2\frac{1}{\overline{inc}}$=0.0052	$\hat{\beta}_2\frac{1}{\overline{dairy}}$=0.6441	0.4566
Log-log	$\hat{\beta}_2\frac{\overline{dairy}}{\overline{inc}}$=0.0056	$\hat{\beta}_2$=0.6864	0.5188
Log-linear	$\hat{\beta}_2\times\overline{dairy}$=0.0055	$\hat{\beta}_2\times\overline{inc}$=0.6783	0.4976
Inverse-log	$-\hat{\beta}_2\frac{\overline{dairy}}{\left[\overline{inc}\right]^2}$=0.0047	$-\hat{\beta}_2\frac{1}{\overline{inc}}$=0.5815	0.5038

The R^2 obtained in the first three models are not comparable with the R^2 obtained in the last three because the functional form of the regressand is different: y in the first three models and $\ln(y)$ in the last three.

Comparing the first three models the best fit is obtained by the linear-log model, if we use the R^2 as goodness of fit measure. Comparing the last three models the best fit is obtained by the log-log model. If we had used the Akaike Information Criterion (AIC), which allows the comparison of models with different functional forms for the regressand, then the-log-log model would have been the best among the six models fitted. The AIC measured will be studied in chapter 3.

Appendixes

Appendix 2.1: Two alternative forms to express $\hat{\beta}_2$

It is easy to see that

$$\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x}) = \sum_{i=1}^{n}(y_i x_i - \bar{x}y_i - \bar{y}x_i + \bar{y}\bar{x}) = \sum_{i=1}^{n} y_i x_i - \bar{x}\sum_{i=1}^{n} y_i - \bar{y}\sum_{i=1}^{n} x_i + n\bar{y}\bar{x}$$

$$= \sum_{i=1}^{n} y_i x_i - n\bar{x}\bar{y} - \bar{y}\sum_{i=1}^{n} x_i + n\bar{y}\bar{x} = \sum_{i=1}^{n} y_i x_i - \bar{y}\sum_{i=1}^{n} x_i$$

On the other hand, we have

$$\sum_{i=1}^{n}(x_i-\overline{x})^2=\sum_{i=1}^{n}(x_i^2-2\overline{x}x_i+\overline{x}\overline{x})^2=\sum_{i=1}^{n}x_i^2-2\overline{x}\sum_{i=1}^{n}x_i+n\overline{x}\overline{x}$$

$$=\sum_{i=1}^{n}x_i^2-2n\overline{x}^2+n\overline{x}^2=\sum_{i=1}^{n}x_i^2-\overline{x}\sum_{i=1}^{n}x_i$$

Therefore, (2-17) can be expressed in the following way:

$$\hat{\beta}_2=\frac{\sum_{i=1}^{n}y_ix_i-\overline{y}\sum_{i=1}^{n}x_i}{\sum_{i=1}^{n}x_i^2-\overline{x}\sum_{i=1}^{n}x_i}=\frac{\sum_{i=1}^{n}(y_i-\overline{y})(x_i-\overline{x})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

Appendix 2.2. Proof: $r_{xy}^2=R^2$

First of all, we are going to see an equivalence that will be used in the proof. By definition,

$$\hat{y}_i=\hat{\beta}_1+\hat{\beta}_2x_i$$

From the first normal equation, we have

$$\overline{y}=\hat{\beta}_1+\hat{\beta}_2\overline{x}$$

Subtracting the second equation from the first one:

$$\hat{y}_i-\overline{y}=\hat{\beta}_2(x_i-\overline{x})$$

Squaring both sides

$$(\hat{y}_i-\overline{y})^2=\hat{\beta}_2^2(x_i-\overline{x})^2$$

and summing for all *i*, we have

$$\sum(\hat{y}_i-\overline{y})^2=\hat{\beta}_2^2\sum(x_i-\overline{x})^2$$

Taking into account the previous equivalence, we have

$$R^2=\frac{\sum_{i=1}^{n}(\hat{y}_i-\overline{\hat{y}})^2}{\sum_{i=1}^{n}(y_i-\overline{y})^2}=\frac{\hat{\beta}_2^2\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sum_{i=1}^{n}(y_i-\overline{y})^2}=\frac{\left[\sum_{i=1}^{n}(y_i-\overline{y})(x_i-\overline{x})\right]^2}{\left[\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^2}\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sum_{i=1}^{n}(y_i-\overline{y})^2}$$

$$=\frac{\left[\sum_{i=1}^{n}(y_i-\overline{y})(x_i-\overline{x})\right]^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\frac{1}{\sum_{i=1}^{n}(y_i-\overline{y})^2}=r_{xy}^2$$

Appendix 2.3. Proportional change *versus* change in logarithms

Change in logarithms is a variation rate, which is used in economics research. The relationship between proportional change and change in logarithms can be seen if we expand (2-45) by Taylor series:

$$\begin{aligned}
\ln(x_1)-\ln(x_0) &= \ln\left[\frac{x_1}{x_0}\right] \\
&= \ln(1)+\left[\frac{x_1}{x_0}-1\right]\left[\frac{1}{\frac{x_1}{x_0}}\right]_{\frac{x_1}{x_0}=1}+\frac{1}{2}\left[\frac{x_1}{x_0}-1\right]^2\left[-\frac{1}{\left[\frac{x_1}{x_0}\right]}\right]_{\frac{x_1}{x_0}=1} \\
&\quad +\frac{1}{3\times 2}\left[\frac{x_1}{x_0}-1\right]^3\left[\frac{2}{\left[\frac{x_1}{x_0}\right]^3}\right]_{\frac{x_1}{x_0}=1}+\cdots\cdots \qquad (2\text{-}100)\\
&= \left[\frac{x_1}{x_0}-1\right]-\frac{1}{2}\left[\frac{x_1}{x_0}-1\right]^2+\frac{1}{3}\left[\frac{x_1}{x_0}-1\right]^3+\cdots \\
&= \frac{\Delta x_1}{x_0}-\frac{1}{2}\left[\frac{\Delta x_1}{x_0}\right]^2+\frac{1}{3}\left[\frac{\Delta x_1}{x_0}\right]^3+\cdots
\end{aligned}$$

Therefore, if we take the linear approximation in this expansion, we have

$$\Delta\ln(x)=\ln(x_1)-\ln(x_0)=\ln\left[\frac{x_1}{x_0}\right]\approx\frac{\Delta x_1}{x_0} \qquad (2\text{-}101)$$

Appendix 2.4. Proof: *OLS* estimators are linear and unbiased

We will only prove the unbiasedness of the estimator $\hat{\beta}_2$, which is the most important. In order to prove this, we need to rewrite our estimator in terms of the population parameter. The formula (2-18) can be written as

$$\hat{\beta}_2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \qquad (2\text{-}102)$$

because $\sum_{i=1}^{n}(x_i-\bar{x})\bar{y}=\bar{y}\sum_{i=1}^{n}(x_i-\bar{x})=\bar{y}\times 0=0$

Now (2-102) will be expressed in the following way:

$$\hat{\beta}_2=\sum_{i=1}^{n}c_iy_i \qquad (2\text{-}103)$$

where

$$c_i = \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2-104}$$

The c_i's have the following properties:

$$\sum_{i=1}^{n} c_i = 0 \tag{2-105}$$

$$\sum_{i=1}^{n} c_i^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2} = \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2-106}$$

$$\sum_{i=1}^{n} c_i x_i = \frac{\sum_{i=1}^{n}(x_i - \bar{x})x_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = 1 \tag{2-107}$$

Now, if we substitute $y = \beta_1 + \beta_2 x + u$ (assumption 1) in (2-102), we have

$$\begin{aligned} \hat{\beta}_2 &= \sum_{i=1}^{n} c_i y_i = \sum_{i=1}^{n} c_i(\beta_1 + \beta_2 x_i + u_i) \\ &= \beta_1 \sum_{i=1}^{n} c_i + \beta_2 \sum_{i=1}^{n} c_i x_i + \sum_{i=1}^{n} c_i u_i = \beta_2 + \sum_{i=1}^{n} c_i u_i \end{aligned} \tag{2-108}$$

Since the regressors are assumed to be nonstochastic (assumption 2), the c_i are nonstochastic too. Therefore, $\hat{\beta}_2$ is an estimator that is a *linear* function of u's.

Taking expectations in (2-108) and taking into account assumption 6, and implicitly assumptions 3 through 5, we obtain

$$E(\hat{\beta}_2) = \beta_2 + \sum_{i=1}^{n} c_i E(u_i) = \beta_2 \tag{2-109}$$

Therefore, $\hat{\beta}_2$ is an unbiased estimator of β_2

Appendix 2.5. Calculation of variance of $\hat{\beta}_2$:

$$\begin{aligned} E\left[\hat{\beta}_2 - \beta_2\right]^2 &= \left[\sum_{i=1}^{n} c_i u_i\right]^2 = \sum_{i=1}^{n} c_i^2 E(u_i^2) + \sum_{i \neq j}\sum_{i=1}^{n} c_i c_j E(u_i u_j) \\ &= \sigma^2 \sum_{i=1}^{n} c_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sigma^2}{nS_X^2} \end{aligned} \tag{2-110}$$

In the above proof, to pass from the second to the third equality, we have taken into account assumptions 6 and 7.

Appendix 2.6. Proof of Gauss-Markov Theorem for the slope in simple regression

The plan for the proof is the following. First, we are going to define an arbitrary estimator $\tilde{\beta}_2$ which is linear in *y*. Second, we will impose restrictions implied by unbiasedness. Third, we will show that the variance of the arbitrary estimator must be larger than, or at least equal to, the variance of $\hat{\beta}_2$.

Let us define an arbitrary estimator $\tilde{\beta}_2$ which is linear in *y*:

$$\tilde{\beta}_2 = \sum_{i=1}^{n} h_i y_i \tag{2-111}$$

Now, we substitute y_i by its value in the population model (assumption 1):

$$\tilde{\beta}_2 = \sum_{i=1}^{n} h_i y_i = \sum_{i=1}^{n} h_i(\beta_1 + \beta_2 x_i + u_i) = \beta_1 \sum_{i=1}^{n} h_i + \beta_2 \sum_{i=1}^{n} h_i x_i + \sum_{i=1}^{n} h_i u_i \tag{2-112}$$

For the estimator $\tilde{\beta}_2$ to be unbiased, the following restrictions must be accomplished:

$$\sum_{i=1}^{n} h_i = 0 \qquad\qquad \sum_{i=1}^{n} h_i x_i = 1 \tag{2-113}$$

Therefore,

$$\tilde{\beta}_2 = \beta_2 + \sum_{i=1}^{n} h_i u_i \tag{2-114}$$

The variance of this estimator is the following:

$$E\left[\tilde{\beta}_2 - \beta_2\right]^2 = \left[\sum_{i=1}^{n} h_i u_i\right]^2 = \sigma^2 \sum_{i=1}^{n} h_i^2 =$$

$$\sigma^2 \sum_{i=1}^{n} \left[h_i - \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} + \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right]^2 = \sigma^2 \sum_{i=1}^{n} \left[h_i - \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right]^2 \tag{2-115}$$

$$+\sigma^2 \sum_{i=1}^{n} \left[\frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right]^2 + 2\sigma^2 \sum_{i=1}^{n} \left[h_i - \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right] \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

The third term of the last equality is 0, as shown below:

$$2\sigma^2\sum_{i=1}^{n}\left[h_i-\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$
$$=2\sigma^2\sum_{i=1}^{n}\left[h_i\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]-2\sigma^2\sum_{i=1}^{n}\left[\frac{(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]=2\sigma^2\times 1-2\sigma^2\times 1=0 \tag{2-116}$$

Therefore, taking into account (2-116) and operating, we have

$$E\left[\tilde{\beta}_2-\beta_2\right]^2=\sigma^2\sum_{i=1}^{n}\left[h_i-c_i\right]^2+\sigma^2\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{2-117}$$

where $c_i=\dfrac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

The second term of the last equality is the variance of $\hat{\beta}_2$, while the first term is always positive because it is a sum of squares, except that $h_i=c_i$, for all i, in which case it is equal to 0, and then $\tilde{\beta}_2=\hat{\beta}_2$. So,

$$E\left[\tilde{\beta}_2-\beta_2\right]^2\geq E\left[\hat{\beta}_2-\beta_2\right]^2 \tag{2-118}$$

Appendix 2.7. Proof: $\hat{\sigma}^2$ is an unbiased estimator of the variance of the disturbance

The population model is by definition:

$$y_i=\beta_1+\beta_2x_i+u_i \tag{2-119}$$

If we sum up both sides of (2-119) for all i and divide by n, we have

$$\bar{y}=\beta_1+\beta_2\bar{x}+\bar{u} \tag{2-120}$$

Subtracting (2-120) from (2-119), we have

$$y_i-\bar{y}=\beta_2\left(x_i-\bar{x}\right)+\left(u_i-\bar{u}\right) \tag{2-121}$$

On the other hand, $\hat{u}_i$ is by definition:

$$\hat{u}_i=y_i-\hat{\beta}_1-\hat{\beta}_2x_i \tag{2-122}$$

If we sum up both sides of (2-122) for all i and divide by n, we have

$$\bar{\hat{u}}=\bar{y}-\hat{\beta}_1-\hat{\beta}_2\bar{x} \tag{2-123}$$

Subtracting (2-123) from (2-122), and taking into account that $\bar{\hat{u}}$ =0,

$$\hat{u}_i=\left(y_i-\bar{y}\right)-\hat{\beta}_2\left(x_1-\bar{x}\right) \tag{2-124}$$

Substituting (2-121) in (2-124), we have

$$\begin{aligned}\hat{u}_i &= \beta_2\left(x_i-\bar{x}\right)+\left(u_i-\bar{u}\right)-\hat{\beta}_2\left(x_1-\bar{x}\right)\\ &=-\left(\hat{\beta}_2-\beta_2\right)\left(x_1-\bar{x}\right)+\left(u_i-\bar{u}\right)\end{aligned} \tag{2-125}$$

Squaring and summing up both sides of (2-125), we have

$$\begin{aligned}\sum_{i=1}^{n}\hat{u}_i^2 = \left[\tilde{\beta}_2-\beta_2\right]^2\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}(u_i-\bar{u})^2\\ -2\left[\tilde{\beta}_2-\beta_2\right]\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\end{aligned} \tag{2-126}$$

Taking expectation in (2-126), we obtain

$$\begin{aligned}E\left[\sum_{i=1}^{n}\hat{u}_i^2\right] &= \sum_{i=1}^{n}(x_i-\bar{x})^2 E\left[\tilde{\beta}_2-\beta_2\right]^2+E\left[\sum_{i=1}^{n}(u_i-\bar{u})^2\right]\\ &\quad -2E\left[\left(\tilde{\beta}_2-\beta_2\right)\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\right]\\ &=\sum_{i=1}^{n}(x_i-\bar{x})^2\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}+\left(n-1\right)\sigma^2-2\sigma^2=\left(n-2\right)\sigma^2\end{aligned} \tag{2-127}$$

To obtain the first term of the last equality of (2-127), we have used (2-64). In (2-128) and (2-129), you can find the developments used to obtain the second and the third term of the last equality of (2-127) respectively. In both cases, assumptions 7 and 8 have been taken into account.

$$\begin{aligned}E\left[\sum_{i=1}^{n}(u_i-\bar{u})^2\right] &= E\left[\sum_{i=1}^{n}u_i^2-n\bar{u}^2\right]=E\left[\sum_{i=1}^{n}u_i^2-n\left(\frac{\sum_{i=1}^{n}u_i}{n}\right)^2\right]\\ &=E\left[\sum_{i=1}^{n}u_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}u_i^2+\sum_{i\neq j}u_iu_j\right)\right]=n\sigma^2-\frac{n}{n}\sigma^2=\left(n-1\right)\sigma^2\end{aligned} \tag{2-128}$$

$$\begin{aligned}E\left[\left(\tilde{\beta}_2-\beta_2\right)\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\right] &= E\left[\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sum_{i=1}^{n}(x_i-\bar{x})u_i\sum_{i=1}^{n}(x_i-\bar{x})u_i\right]\\ &=\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})E\left(u_i\right)\right]^2\\ &=\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2E\left(u_i\right)^2+\sum_{i\neq j}\sum(x_i-\bar{x})(x_i-\bar{x})E\left(u_iu_j\right)\right]=\sigma^2\end{aligned} \tag{2-129}$$

According to (2-127), we have

$$E\left[\sum_{i=1}^{n}\hat{u}_i^2\right]=(n-2)\sigma^2 \qquad (2\text{-}130)$$

Therefore, an unbiased estimator is given by

$$\hat{\sigma}^2=\frac{\sum_{i=1}^{n}\hat{u}_i^2}{n-2} \qquad (2\text{-}131)$$

such as

$$E\left(\hat{\sigma}^2\right)=\frac{1}{n-2}E\left(\sum_{i=1}^{n}\hat{u}_i^2\right)=\sigma^2 \qquad (2\text{-}132)$$

Appendix 2.8. Consistency of the *OLS* estimator

The operator plim has the in variance property (Slutsky property). That is to say, if $\hat{\theta}$ is a consistent estimator of θ and if $g\left(\hat{\theta}\right)$ is any continuous function of $\hat{\theta}$, then

$$\operatorname*{plim}_{n\to\infty} g(\hat{\theta})=g(\theta) \qquad (2\text{-}133)$$

This means is that if $\hat{\theta}$is a consistent estimator of θ, then $1/\hat{\theta}$and ln($\hat{\theta}$) are also consistent estimators of $1/\theta$ and ln(θ) respectively. Note that these properties do not hold true for the expectation operator E; for example, if$\hat{\theta}$is an unbiased estimator of θ [that is to say, $E(\hat{\theta})=\theta$], it is *not true* that $1/\hat{\theta}$is an unbiased estimator of $1/\theta$; that is, $E(1/\hat{\theta})\neq 1/E(\hat{\theta})\neq 1/\theta$. This is due to the fact that the expectation operator can be only applied to *linear* functions of random variables. On the other hand, the plim operator is applicable to any continuous functions.

Under assumptions 1 through 6, the *OLS* estimators, $\hat{\beta}_1$ and $\hat{\beta}_2$, are consistent.

Now we are going to prove that $\hat{\beta}_2$ is a consistent estimator. First, $\hat{\beta}_2$ can be expressed as:

$$\begin{aligned}\hat{\beta}_2&=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(\beta_1+\beta_2x_i+u_i)}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\\&=\frac{\beta_1\sum_{i=1}^{n}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}+\beta_2\frac{\sum_{i=1}^{n}(x_i-\bar{x})x_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}+\frac{\sum_{i=1}^{n}(x_i-\bar{x})u_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\beta_2+\frac{\sum_{i=1}^{n}(x_i-\bar{x})u_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\end{aligned} \qquad (2\text{-}134)$$

In order to prove consistency, we need to take plim´s in (2-134) and apply the *Law of Large Numbers*. This law states that under general conditions, the sample moments converge to their corresponding population moments. Thus, taking plim´s in (2-134):

$$\underset{n\to\infty}{\text{plim}}\,\hat{\beta}_2=\underset{n\to\infty}{\text{plim}}\left[\beta_2+\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)u_i}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\right]=\beta_2+\frac{\underset{n\to\infty}{\text{plim}}\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)u_i}{\underset{n\to\infty}{\text{plim}}\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\tag{2-135}$$

In the last equality we have divided the numerator and denominator by n, because if we do not do so, both summations will go to infinity when n goes to infinity..

If we apply the law of large numbers to the numerator and denominator of (2-135), they will converge in probability to the population moments *cov*(x,u) and *var*(x) respectively. Provided *var*(x)≠0 (assumption 4), we can use the properties of the *probability limits* to obtain

$$\text{plim}\hat{\beta}_2=\beta_2+\frac{cov(x,u)}{var(x)}=\beta_2\tag{2-136}$$

To reach the last equality, using assumptions 2 and 6, we obtain

$$cov(x,u)=E\left[(x-\bar{x})u\right]=(x-\bar{x})E\left[u\right]=(x-\bar{x})\times 0=0\tag{2-137}$$

Therefore, $\hat{\beta}_2$ is a consistent estimator.

Appendix 2.9 Maximum likelihood estimator

Taking into account assumptions 1 through 6 the expectation of y_i is the following:

$$E(y_i)=\beta_1+\beta_2 x_i\tag{2-138}$$

If we take into account assumptions 7, the variance of y_i is equal to

$$\text{var}(y_i)=E\left[y_i-E(y_i)\right]^2=E\left[y_i-\beta_1+\beta_2 x_i\right]^2=E\left[u_i\right]^2=\sigma^2\quad\forall i\tag{2-139}$$

According to assumption 1, y_i is a linear function of u_i, and if u_i has a normal distribution (assumption 9), then y_i will be normally and independently (assumption 8) distributed with mean $\beta_1+\beta_2 x_i$ and variance σ^2.

Then, the joint probability density function of $y_1,y_2,\cdots,y_n$ can be expressed as a product of n individual density functions:

$$\begin{aligned}&f\left(y_1,\,y_2,\,\dots,y_n\,|\,\beta_1+\,\beta_2 x_i,\,\sigma^2\right)\\&=f\left(y_1\,|\,\beta_1+\,\beta_2 x_i,\,\sigma^2\right)f\left(y_2\,|\,\beta_1+\,\beta_2 x_i,\,\sigma^2\right)\cdots f\left(y_n\,|\,\beta_1+\,\beta_2 x_i,\,\sigma^2\right)\end{aligned}\tag{2-140}$$

where

$$f\left(y_i\right)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{\left[y_i-\beta_1-\beta_2 x_i\right]^2}{\sigma^2}\right\}\tag{2-141}$$

which is the density function of a normally distributed variable with the given mean and variance.

Substituting (2-141) into (2-140)for each y_i, we obtain

$$
\begin{aligned}
f\left(y_1, y_2, \cdots, y_n\right) &= f\left(y_1\right) f\left(y_2\right) \cdots f\left(y_n\right) \\
&= \frac{1}{\sigma^n \left(\sqrt{2\pi}\right)^n} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \frac{\left[y_i - \beta_1 - \beta_2 x_i\right]^2}{\sigma^2} \right\}
\end{aligned} \tag{2-142}
$$

If $y_1, y_2, \cdots, y_n$ are known or given, but β_2, β_3, and σ^2 are not known, the function in (2-142) is called a likelihood function, denoted by $L(\beta_2, \beta_3, \sigma^2)$ or simply L. If we take natural logarithms in (2-142), we obtain

$$
\begin{aligned}
\ln L &= -n \ln \sigma - \frac{n}{2} \ln\left(\sqrt{2\pi}\right) - \frac{1}{2} \sum_{i=1}^{n} \frac{\left(y_i - \beta_1 - \beta_2 x_i\right)}{\sigma^2} \\
&= -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln\left(\sqrt{2\pi}\right) - \frac{1}{2} \sum_{i=1}^{n} \frac{\left(y_i - \beta_1 - \beta_2 x_i\right)}{\sigma^2}
\end{aligned} \tag{2-143}
$$

The *maximum likelihood* (*ML*) method, as the name suggests, consists in estimating the unknown parameters in such a manner that the probability of observing the given y_i's is as high (or maximum) as possible. Therefore, we have to find the maximum of the function (2-143). To maximize (2-143) we must differentiate with respect to β_2, β_3, and σ^2 and equal to 0. If $\tilde{\beta}_1$, $\tilde{\beta}_2$ and $\tilde{\sigma}^2$ denote the *ML* estimators, we obtain:

$$
\begin{aligned}
\frac{\partial \ln L}{\partial \tilde{\beta}_1} &= -\frac{1}{\tilde{\sigma}^2} \sum \left(y - \tilde{\beta}_1 - \tilde{\beta}_2 x_i\right)(-1) = 0 \\
\frac{\partial \ln L}{\partial \tilde{\beta}_2} &= -\frac{1}{\tilde{\sigma}^2} \sum \left(y - \tilde{\beta}_1 - \tilde{\beta}_2 x_i\right)(-x_i) = 0 \\
\frac{\partial \ln L}{\partial \tilde{\sigma}^2} &= -\frac{n}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4} \sum \left(y - \tilde{\beta}_1 - \tilde{\beta}_2 x_i\right)^2 = 0
\end{aligned} \tag{2-144}
$$

If we take the first two equations of (2-144) and operate, we have

$$\sum y_i = n\tilde{\beta}_1 + \tilde{\beta}_2 \sum x_i \tag{2-145}$$

$$\sum y_i x_i = \tilde{\beta}_1 \sum x_i + \tilde{\beta}_2 \sum x_i^2 \tag{2-146}$$

As can be seen, (2-145) and (2-146) are equal to (2-13) and (2-14). That is to say, the *ML* estimators, under the *CLM* assumptions, are equal to the *OLS* estimators.

Substituting $\tilde{\beta}_1$ and $\tilde{\beta}_2$, obtained solving (2-145) and (2-146), in the third equation of (2-144), we have

$$\tilde{\sigma}^2 = \frac{1}{n} \sum \left(y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_i\right)^2 = \frac{1}{n} \sum \left(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i\right)^2 = \frac{1}{n} \sum \hat{u}_i^2 \tag{2-147}$$

The *ML* estimator for $\tilde{\sigma}^2$ is biased, since, according to (2-127),

$$E(\tilde{\sigma}^2) = \frac{1}{n} E\left[\sum_{i=1}^{n} \hat{u}_i^2\right] = \frac{n-2}{n} \sigma^2 \tag{2-148}$$

In any case, $\tilde{\sigma}^2$ is a consistent estimator because

$$\lim_{n \to 0} \frac{n-2}{n} = 1 \tag{2-149}$$