


Figure 1.3. Sample and Population Regression Function

1.1.4 Example

In order to show how a DGP works, we implement the following experiment. We generate three replicates of sample $n = 10$ of the following data generating process: $y_i = 2 + 0.5x_i + u_i$. X is generated by a uniform distribution as follows $X \sim U[0, 1]$.

 XEGlinreg04.xpl

This code produces the values of X , which are the same for the three samples, and the corresponding values of Y , which of course differ from one sample to the other.

1.2 Estimators and Properties

If we have available a sample of n observations from the population represented by (X, Y) , $(x_1, y_1), \dots, (x_n, y_n)$, and we assume the Population Regression Function is both linear in variables and parameters

$$y_i = E(Y|X = x_i) + u_i = \alpha + \beta x_i + u_i, \quad i = 1, \dots, n, \quad (1.25)$$

we can now face the task of estimating the unknown parameters α and β . Un-

fortunately, the sampling design and the linearity assumption in the PRF, are not sufficient conditions to ensure that there exists a precise statistical relationship between the estimators and its true corresponding values (see section 1.2.6 for more details). In order to do so, we need to know some additional features from the PRF. Since we do not them, we decide to establish some assumptions, making clear that in any case, the statistical properties of the estimators are going to depend crucially on the related assumptions. The basic set of assumptions that comprises the classical linear regression model is as follows:

(A.1) The explanatory variable, X , is fixed.

(A.2) For any $n > 1$,

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 > 0.$$

(A.3)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = m > 0.$$

(A.4) Zero mean disturbances: $E(u) = 0$.

(A.5) Homoscedasticity: $Var(u_i) = \sigma^2 < \infty$, is constant, for all i .

(A.6) Nonautocorrelation: $Cov(u_i, u_j) = 0$ if $i \neq j$.

Finally, an additional assumption that is usually employed to easier the inference is

(A.7) The error term has a gaussian distribution, $u_i \sim N(0, \sigma^2)$

For a more detailed explanation and comments on the different assumption see Gujarati (1995). Assumption (A.1) is quite strong, and it is in fact very difficult to accept when dealing with economic data. However, most part of statistical results obtained under this hypothesis hold as well for weaker such as random X but independent of u (see Amemiya (1985) for the fixed design case, against Newey and McFadden (1994) for the random design).

1.2.1 Regression Parameters and their Estimation

In the univariate linear regression setting that was introduced in the previous section the following parameters need to be estimated


- α - intercept term. It gives us the value of the conditional expectation of Y given $X = x$, for $x = 0$.
- β - linear slope coefficient. It represents the sensitivity of $E(Y|X = x)$ to changes in x .
- σ^2 - Error term measure of dispersion. Large values of the variance mean that the error term u is likely to vary in a large neighborhood around the expected value. Smaller values of the standard deviation indicate that the values of u will be concentrated around the expected value.

Regression Estimation

From a given population described as


$$y = 3 + 2.5x + u \quad (1.26)$$

$X \sim U[0, 1]$ and $u \sim N(0, 1)$, a random sample of $n = 100$ elements is generated.

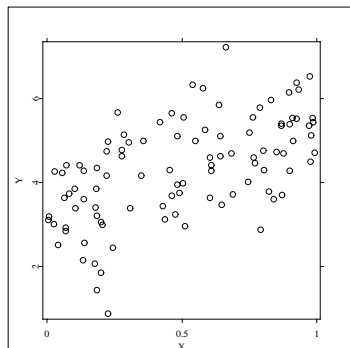
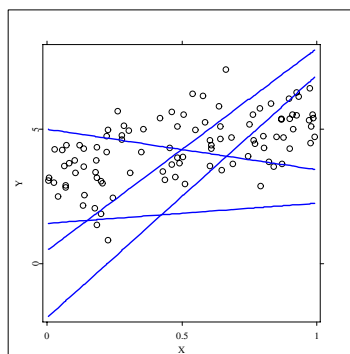
 XELinreg05.xpl

We show the scatter plot in Figure 1.4


Following the same reasoning as in the previous sections, the PRF is unknown for the researcher, and he has only available the data, and some information from the PRF. For example, he may know that the relationship between $E(Y|X = x)$ and x is linear, but he does not know which are the exact parameter values. In Figure 1.5 we represent the sample and several possible values of the regression functions according to different values for α and β .

 XELinreg06.xpl

In order to estimate α and β , many estimation procedures are available. One of the most famous criteria is the one that chooses α and β such that they minimize the sum of the squared deviations of the regression values from their

Figure 1.4. Sample $n = 100$ of (X, Y) Figure 1.5. Sample of X, Y , Possible linear functions

real corresponding values. This is the so called least squares method. Applying this procedure to the previous sample,

 XEGlinreg07.xpl

in Figure 1.6, we show for the sake of comparison the least squares regression curve together with the other sample regression curves.

We describe now in a more precise way how the Least Squares method is implemented, and, under a Population Regression Function that incorporates

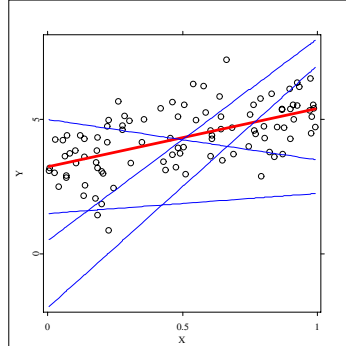


Figure 1.6. Ordinary Least Squares Estimation

assumptions (A.1) to (A.6), which are its statistical properties.

1.2.2 Least Squares Method

We begin by establishing a formal estimation criteria. Let $\hat{\alpha}$ and $\hat{\beta}$ be a possible estimators (some function of the sample observations) of α and β . Then, the fitted value of the endogenous variable is:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i \quad i = 1, \dots, n \quad (1.27)$$

The residual value between the real and the fitted value is given by

$$\hat{u}_i = y_i - \hat{y}_i \quad i = 1, \dots, n \quad (1.28)$$

The least squares method minimizes the sum of squared deviations of regression values ($\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$) from the observed values (y_i), that is, the residual sum of squares—RSS.

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min \quad (1.29)$$

This criterion function has two variables with respect to which we are willing to minimize: $\hat{\alpha}$ and $\hat{\beta}$.

$$S(\hat{\alpha}, \hat{\beta}) = \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2. \quad (1.30)$$

Then, we define as Ordinary Least Squares (OLS) estimators, denoted by $\hat{\alpha}$ and $\hat{\beta}$, the values of α and β that solve the following optimization problem

$$(\hat{\alpha}, \hat{\beta}) = \underset{\hat{\alpha}, \hat{\beta}}{\operatorname{argmin}} S(\hat{\alpha}, \hat{\beta}) \quad (1.31)$$

In order to solve it, that is, to find the minimum values, the first conditions make the first partial derivatives have to be equal to zero.

$$\frac{\partial S(\hat{\alpha}, \hat{\beta})}{\partial \hat{\alpha}} = -2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \quad (1.32)$$

$$\frac{\partial S(\hat{\alpha}, \hat{\beta})}{\partial \hat{\beta}} = -2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)x_i = 0$$

To verify whether the solution is really a minimum, the matrix of second order derivatives of (1.32), the Hessian matrix, must be positive definite. It is easy to show that

$$H(\hat{\alpha}, \hat{\beta}) = 2 \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}, \quad (1.33)$$

and this expression is positive definite if and only if, $\sum_i (x_i - \bar{x})^2 > 0$. But, this is implied by assumption (A.2). Note that this requirement is not strong at all. Without it, we might consider regression problems where no variation at all is considered in the values of X . Then, condition (A.2) rules out this degenerate case.

The first derivatives (equal to zero) lead to the so-called (least squares) normal equations from which the estimated regression parameters can be computed by

solving the equations.

$$n\hat{\alpha} + \hat{\beta} \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad (1.34)$$

$$\hat{\alpha} \sum_{i=1}^n x_i + \hat{\beta} \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \quad (1.35)$$

Dividing the original equations by n , we get a simplified formula suitable for the computation of regression parameters

$$\begin{aligned} \hat{\alpha} + \hat{\beta}\bar{x} &= \bar{y} \\ \hat{\alpha}\bar{x} + \hat{\beta}\frac{1}{n}\sum_{i=1}^n x_i^2 &= \frac{1}{n}\sum_{i=1}^n x_i y_i \end{aligned}$$

For the estimated intercept $\hat{\alpha}$, we get:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad (1.36)$$

For the estimated linear slope coefficient $\hat{\beta}$, we get:

$$\begin{aligned} (\bar{y} - \hat{\beta}\bar{x})\bar{x} + \hat{\beta}\frac{1}{n}\sum_{i=1}^n x_i^2 &= \frac{1}{n}\sum_{i=1}^n x_i y_i \\ \hat{\beta}\frac{1}{n}\sum_{i=1}^n (x_i^2 - \bar{x}^2) &= \frac{1}{n}\sum_{i=1}^n x_i y_i - \bar{x}\bar{y} \\ \hat{\beta}S_X^2 &= S_{XY} \end{aligned}$$

$$\hat{\beta} = \frac{S_{XY}}{S_X^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.37)$$

The ordinary least squares estimator of the parameter σ^2 is based on the following idea: Since σ^2 is the expected value of u_i^2 and \hat{u} is an estimate of u , our initial estimator

$$\widehat{\sigma^{*2}} = \frac{1}{n} \sum_i \hat{u}_i^2 \quad (1.38)$$

would seem to be a natural estimator of σ^2 , but due to the fact that $E(\sum_i \hat{u}_i^2) = (n-2)\sigma^2$, this implies

$$E(\widehat{\sigma^{*2}}) = \frac{n-2}{n} \sigma^2 \neq \sigma^2. \quad (1.39)$$

Therefore, the unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{\sum_i \hat{u}_i^2}{n-2} \quad (1.40)$$

Now, with this expression, we obtain that $E(\hat{\sigma}^2) = \sigma^2$.


In the next section we will introduce an example of the least squares estimation criterion.

1.2.3 Example

We can obtain a graphical representation of the least squares ordinary estimation by using the following Quantlet

```
g1 = grlinreg (x)
```

The regression line computed by the least squares method using the data generated in (1.49)

 XEGlinreg08.xpl

is shown in Figure 1.7 jointly with the data set.

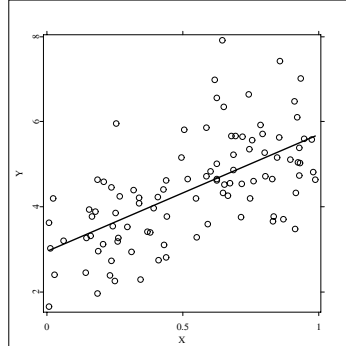


Figure 1.7. Ordinary Least Squares Estimation

1.2.4 Goodness of Fit Measures

Once the regression line is estimated, it is useful to know how well the regression line approximates the data from the sample. A measure that can describe the quality of representation is called the coefficient of determination (either R-Squared or R^2). Its computation is based on a decomposition of the variance of the values of the dependent variable Y .

The smaller is the sum of squared estimated residuals, the better is the quality of the regression line. Since the Least Squares method minimizes the variance of the estimated residuals it also maximizes the R-squared by construction.

$$\sum (y_i - \hat{y}_i)^2 = \sum \hat{u}_i^2 \rightarrow \min. \quad (1.41)$$

The sample variance of the values of Y is:

$$S_Y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} \quad (1.42)$$

The element $\sum_{i=1}^n (y_i - \bar{y})^2$ is known as *Total Sum of Squares* (TSS), it is the total variation of the values of Y from \bar{y} . The deviation of the observed values, y_i , from the arithmetic mean, \bar{y} , can be decomposed into two parts: The deviation of the observed values of Y from the estimated regression values

and the deviation of the estimated regression values from the sample mean. i. e.

$$y_i - \bar{y} = (y_i - \hat{y}_i + \hat{y}_i - \bar{y}) = \hat{u}_i + \hat{y}_i - \bar{y}, \quad i = 1, \dots, n \quad (1.43)$$

where $\hat{u}_i = y_i - \hat{y}_i$ is the error term in this estimate. Note also that considering the properties of the OLS estimators it can be proved that $\bar{y} = \bar{\hat{y}}$. Taking the square of the residuals and summing over all the observations, we obtain the *Residual Sum of Squares*, $RSS = \sum_{i=1}^n \hat{u}_i^2$. As a goodness of fit criterion the RSS is not satisfactory because the standard errors are very sensitive to the unit in which Y is measured. In order to propose a criteria that is not sensitive to the measurement units, let us decompose the sum of the squared deviations of equation (1.43) as

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \end{aligned} \quad (1.44)$$

Now, noting that by the properties of the OLS estimators we have that $\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$, expression (1.44) can be written as

$$TSS = ESS + RSS, \quad (1.45)$$

where $ESS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$, is the so called *Explained Sum of Squares*. Now, dividing both sides of equation (1.45) by n , we obtain

$$\begin{aligned} \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} &= \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} + \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n} \\ &= \frac{\sum_{i=1}^n \hat{u}_i^2}{n} + \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n} \end{aligned} \quad (1.46)$$

and then,

$$S_Y^2 = S_u^2 + S_{\hat{Y}}^2 \quad (1.47)$$

The total variance of Y is equal to the sum of the sample variance of the estimated residuals (the unexplained part of the sampling variance of Y) and the part of the sampling variance of Y that is explained by the regression function (the sampling variance of the regression function).

The larger the portion of the sampling variance of the values of Y is explained by the model, the better is the fit of the regression function.

The Coefficient of Determination

The coefficient of the determination is defined as the ratio between the sampling variance of the values of Y explained by the regression function and the sampling variance of values of Y . That is, it represents the proportion of the sampling variance in the values of Y "explained" by the estimated regression function.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{S_{\hat{Y}}^2}{S_Y^2} \quad (1.48)$$

This expression is unit-free because both the numerator and denominator have the same units. The higher the coefficient of determination is, the better the regression function explains the observed values. Other expressions for the coefficient are

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = \frac{\hat{\beta} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\hat{\beta}^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$


One special feature of this coefficient is that the R-Squared can take values in the following range: $0 \leq R^2 \leq 1$. This is always true if the model includes a constant term in the population regression function. A small value of R^2 implies that a lot of the variation in the values of Y has not been explained by the variation of the values of X .

1.2.5 Example

Ordinary Least Squares estimates of the parameters of interest are given by executing the following quantlet

```
{beta,bse,bstan,bpval}=linreg(x,y)
```

As an example, we use the original data source that was already shown in Figure 1.4

 XEGlinreg09.xpl

1.2.6 Properties of the OLS Estimates of α , β and σ^2

Once the econometric model has been both specified and estimated, we are now interested in analyzing the relationship between the estimators (sample) and their respective parameter values (population). This relationship is going to be of great interest when trying to extend propositions based on econometric models that have been estimated with a unique sample to the whole population. One way to do so, is to obtain the sampling distribution of the different estimators. A sampling distribution describes the behavior of the estimators in repeated applications of the estimating formulae. A given sample yields a specific numerical estimate. Another sample from the same population will yield another numerical estimate. A sampling distribution describes the results that will be obtained for the estimators over the potentially infinite set of samples that may be drawn from the population.

Properties of $\hat{\alpha}$ and $\hat{\beta}$

We start by computing the finite sample distribution of the parameter vector $(\alpha \ \beta)^\top$. In order to do so, note that taking the expression for $\hat{\alpha}$ in (1.36) and $\hat{\beta}$ in (1.37) we can write

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \sum_{i=1}^n \begin{pmatrix} \frac{1}{n} - \bar{x}\omega_i \\ \omega_i \end{pmatrix} y_i, \quad (1.49)$$

where

$$\omega_i = \frac{x_i - \bar{x}}{\sum_{l=1}^n (x_l - \bar{x})^2}. \quad (1.50)$$

If we substitute now the value of y_i by the process that has generated it (equa-

tion (1.22)) we obtain

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \sum_{i=1}^n \begin{pmatrix} \frac{1}{n} - \bar{x}\omega_i \\ \omega_i \end{pmatrix} u_i, \quad (1.51)$$

Equations (1.49) and (1.51) show the first property of the OLS estimators of α and β . They are **linear** with respect to the sampling values of the endogenous variable y_1, \dots, y_n , and they also linear in the error terms u_1, \dots, u_n . This property is crucial to show the finite sample distribution of the vector of parameters $(\hat{\alpha} \ \hat{\beta})$ since then, assuming the values of X are fixed (assumption A.1), and independent gaussian errors (assumptions A.6 and A.7), linear combinations of independent gaussian variables are themselves gaussian and therefore $(\hat{\alpha} \ \hat{\beta})$ follow a bivariate gaussian distribution.

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \sim \mathbf{N} \left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \text{Var}(\hat{\alpha}) & \text{Cov}(\hat{\alpha}, \hat{\beta}) \\ \text{Cov}(\hat{\alpha}, \hat{\beta}) & \text{Var}(\hat{\beta}) \end{pmatrix} \right) \quad (1.52)$$

To fully characterize the whole sampling distribution we need to determine both the mean vector, and the variance-covariance matrix of the OLS estimators. Assumptions (A.1), (A.2) and (A.3) immediately imply that

$$\mathbb{E} \left\{ \begin{pmatrix} \frac{1}{n} - \bar{x}\omega_i \\ \omega_i \end{pmatrix} u_i \right\} = \begin{pmatrix} \frac{1}{n} - \bar{x}\omega_i \\ \omega_i \end{pmatrix} \mathbb{E}(u_i) = 0, \quad \forall i \quad (1.53)$$

and therefore by equation (1.51) we obtain

$$\mathbb{E} \left\{ \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \right\} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (1.54)$$

That is, the OLS estimators of α and β , under assumptions (A.1) to (A.7) are **unbiased**. Now we calculate the variance-covariance matrix. In order to do so, let

$$\begin{pmatrix} \text{Var}(\hat{\alpha}) & \text{Cov}(\hat{\alpha}, \hat{\beta}) \\ \text{Cov}(\hat{\alpha}, \hat{\beta}) & \text{Var}(\hat{\beta}) \end{pmatrix} \equiv \mathbb{E} \left\{ \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{pmatrix} (\hat{\alpha} - \alpha \ \hat{\beta} - \beta) \right\} \quad (1.55)$$

Then, if we substitute $\begin{pmatrix} \hat{\alpha} - \alpha & \hat{\beta} - \beta \end{pmatrix}^\top$ by its definition in equation (1.51), the last expression will be equal to

$$= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \left\{ \begin{pmatrix} (\frac{1}{n} - \bar{x}\omega_i)(\frac{1}{n} - \bar{x}\omega_j) & (\frac{1}{n} - \bar{x}\omega_i)\omega_j \\ \omega_i(\frac{1}{n} - \bar{x}\omega_j) & \omega_i\omega_j \end{pmatrix} u_i u_j \right\} \quad (1.56)$$

Now, assumptions (A.1), (A.5) and (A.6) allow us to simplify expression (1.56) and we obtain

$$= \sigma^2 \sum_{i=1}^n \begin{pmatrix} (\frac{1}{n} - \bar{x}\omega_i)^2 & (\frac{1}{n} - \bar{x}\omega_i)\omega_i \\ \omega_i(\frac{1}{n} - \bar{x}\omega_i) & \omega_i^2 \end{pmatrix} \quad (1.57)$$

Finally, substitute ω_i by its definition in equation (1.50) and we will obtain the following expressions for the variance covariance matrix

$$\begin{pmatrix} \text{Var}(\hat{\alpha}) & \text{Cov}(\hat{\alpha}, \hat{\beta}) \\ \text{Cov}(\hat{\alpha}, \hat{\beta}) & \text{Var}(\hat{\beta}) \end{pmatrix} = \sigma^2 \begin{pmatrix} \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} & -\frac{\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ -\frac{\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} & \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{pmatrix} \quad (1.58)$$

We can say that the OLS method produces BLUE (Best Linear Unbiased Estimator) in the following sense: the OLS estimators are the linear, unbiased estimators which satisfy the Gauss-Markov Theorem. We now give the simplest version of the Gauss-Markov Theorem, that is proved in Johnston and Dinardo (1997), p. 36.

Gauss-Markov Theorem: Consider the regression model (1.22). Under assumptions (A.1) to (A.6) the OLS estimators of α and β are those who have minimum variance among the set of all linear and unbiased estimators of the parameters.

We remark that for the Gauss-Markov theorem to hold we do not need to include assumption (A.7) on the distribution of the error term. Furthermore, the properties of the OLS estimators mentioned above are established for finite samples. That is, the estimator divergence between the estimator and the parameter value is analyzed for a fixed sample size. Other properties of the estimators that are also of interest are the asymptotic properties. In this case, the behavior of the estimators with respect to their true parameter values are

analyzed as the sample size increases. Among the asymptotic properties of the estimators we will study the so called **consistency** property.

We will say that the OLS estimators, $\hat{\alpha}$, $\hat{\beta}$, are consistent if they converge weakly in probability (see Serfling (1984) for a definition) to their respective parameter values, α and β . For weak convergence in probability, a sufficient condition is

$$\lim_{n \rightarrow \infty} E \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad (1.59)$$

and

$$\lim_{n \rightarrow \infty} \begin{pmatrix} \text{Var}(\hat{\alpha}) \\ \text{Var}(\hat{\beta}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (1.60)$$

Condition (1.59) is immediately verified since under conditions (A.1) to (A.6) we have shown that both OLS estimators are unbiased in finite sample sizes. Condition (1.60) is shown as follows:

$$\text{Var}(\hat{\alpha}) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) = \frac{\sigma^2}{n} \left(1 + \frac{\bar{x}^2}{n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

then by the properties of the limits

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{\alpha}) = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} \times \lim_{n \rightarrow \infty} \left(\frac{\frac{1}{n} \sum_{i=1}^n x_i^2}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

Assumption (A.3) ensures that

$$\lim_{n \rightarrow \infty} \left(\frac{\frac{1}{n} \sum_{i=1}^n x_i^2}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right) < \infty$$

and since by assumption (A.5), σ^2 is constant and bounded, then $\lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0$. This proves the first part of condition (1.60). The proof for $\hat{\beta}$ follows the same lines.

Properties of σ^2

For the statistical properties of $\hat{\sigma}^2$, we will just enumerate the different statistical results that will be proved in a more general setting in Chapter 2, Section 2.4.2. of this monograph.

Under assumptions (A.1) to (A.7), the finite sample distribution of this estimator is given by

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-2}^2. \quad (1.61)$$

Then, by the properties of the χ^2 distribution it is easy to show that

$$\text{Var}\left(\frac{(n-2)\hat{\sigma}^2}{\sigma^2}\right) = 2(n-2).$$

This result allows us to calculate the variance of $\hat{\sigma}^2$ as

$$\text{Var}(\hat{\sigma}^2) = \frac{2\sigma^4}{n-2}. \quad (1.62)$$

Note that to calculate this variance, the normality assumption, (A.7), plays a crucial role. In fact, by assuming that $u \sim N(0, \sigma^2)$, then $E(u^3) = 0$, and the fourth order moment is already known and related to σ^2 . These two properties are of great help to simplify the third and fourth order terms in equation (1.62).

Under assumptions (A.1) to (A.7) in Section 1.2 it is possible to show (see Chapter 2, Section 2.4.2 for a proof)

Unbiasedness:

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

Non-efficiency: The OLS estimator of σ^2 is not efficient because it does not achieve the Cramer-Rao lower bound (this bound is $\frac{2\sigma^4}{n}$).

Consistency: The OLS estimator of σ^2 converges weakly in probability to σ^2 , i.e.

$$\hat{\sigma}^2 \rightarrow_p \sigma^2$$

as n tends to infinity.

Asymptotic distribution:

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \rightarrow_d N(0, 2\sigma^4)$$

as n tends to infinity.


From the last result, note finally that although $\hat{\sigma}^2$ is not efficient for finite sample sizes, this estimator achieves asymptotically the Cramer-Rao lower bound.

1.2.7 Examples

To illustrate the different statistical properties given in the previous section, we develop three different simulations. The first Monte Carlo experiment analyzes the finite sample distribution of both $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$. The second study performs a simulation to explain consistency, and finally the third study compares finite sample and asymptotic distribution of the OLS estimator of $\hat{\sigma}^2$.

Example 1

The following program illustrates the statistical properties of the OLS estimators of α and β . We implement the following Monte Carlo experiment. We have generated 500 replications of sample size $n = 20$ of the model $y_i = 1.5 + 2x_i + u_i$ $i = 1, \dots, 20$. The values of X have been generated according to a uniform distribution, $X \sim U[0, 1]$, and the values for the error term have been generated following a normal distribution with zero mean and variance one, $u \sim N(0, 1)$. To fulfil assumption (A.1), the values of X are fixed for the 500 different replications. For each sample (replication) we have estimated the parameters α and β and their respective variances (note that σ^2 has been replaced by $\hat{\sigma}^2$). With the 500 values of the estimators of these parameters, we generate four different histograms

 XEGlinreg10.xpl

The result of this procedure is presented in the Figure 1.8. With a sample size of $n = 20$, the histograms that contain the estimations of $\hat{\beta}$ and $\hat{\alpha}$ in the different replications approximate a gaussian distribution. In the other hand, the histograms for the variance estimates approximate a χ^2 distribution, as expected.

Example 2

This program analyzes by simulation the asymptotic behavior of both $\hat{\alpha}$ and $\hat{\beta}$ when the sample size increases. We generate observations using the model, $y_i = 2 + 0.5x_i + u_i$, $X \sim U[0, 1]$, and $u \sim N(0, 10^2)$. For 200 different sample sizes, ($n = 5, \dots, 1000$), we have generated 50 replications for each sample size.

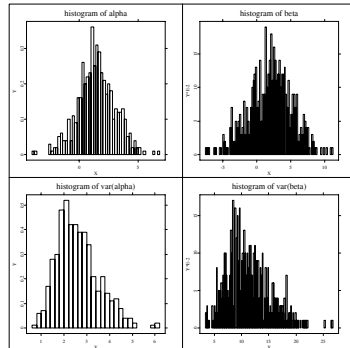



Figure 1.8. Finite sample distribution

For each sample size we estimate 50 estimators of α , β , then, we calculate $E(\hat{\beta})$ and $E(\hat{\alpha})$ conditioning on the sample size.

 XEGLinreg11.xpl

The code gives the output presented in Figure 1.9. As expected, when we increase the sample size $E(\hat{\beta})$ tends to β , in this case $\beta = 0.5$, and $E(\hat{\alpha})$ tends to $\alpha = 2$.

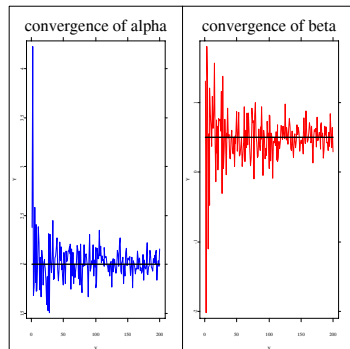



Figure 1.9. Consistency

Example 3

In the model $y_i = 1.5 + 2x_i + u_i$, $X \sim U[0, 1]$, and $u \sim N(0, 16)$. We implement the following Monte Carlo experiment. For two different sample sizes we have generated 500 replications for each sample size. The first 500 replications have a sample size $n = 10$, the second $n = 1000$. In both sample sizes we estimate 500 estimators of σ^2 . Then, we calculate two histograms for the estimates of $\frac{(n-2)\hat{\sigma}^2}{\sigma^2}$, one for $n = 10$, the other for $n = 1000$.

 XELinreg12.xpl

The output of the code is presented in Figure 1.10. As expected, the histogram for $n = 10$ approximates a χ^2 density, whereas for $n = 1000$, the approximated density is the standard normal.

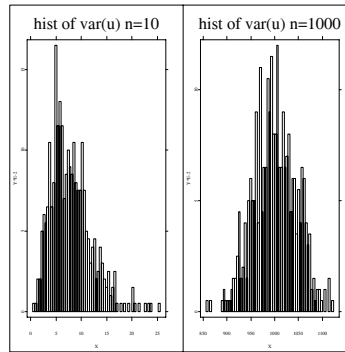


Figure 1.10. Distribution of $\hat{\sigma}^2$

1.3 Inference

In the framework of a univariate linear regression model, one can be interested in testing two different groups of hypotheses about β , α and σ^2 . In the first group, the user has some prior knowledge about the value of β , for example he believes $\beta = \beta_0$, then he is interested in knowing whether this value, β_0 , is compatible with the sample data. In this case the null hypothesis will be $H_0 : \beta = \beta_0$, and the alternative $H_1 : \beta \neq \beta_0$. This is what is called a **two**