4

CLASSICAL NORMAL LINEAR REGRESSION MODEL (CNLRM)

What is known as the **classical theory of statistical inference** consists of two branches, namely, **estimation** and **hypothesis testing.** We have thus far covered the topic of estimation of the parameters of the (two-variable) linear regression model. Using the method of OLS we were able to estimate the parameters β_1, β_2, and σ^2. Under the assumptions of the *classical linear regression model* (CLRM), we were able to show that the estimators of these parameters, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\sigma}^2$, satisfy several desirable statistical properties, such as unbiasedness, minimum variance, etc. (Recall the BLUE property.) Note that, since these are estimators, their values will change from sample to sample. Therefore, these estimators are *random variables*.

But estimation is half the battle. Hypothesis testing is the other half. Recall that in regression analysis our objective is not only to estimate the sample regression function (SRF), but also to use it to draw inferences about the population regression function (PRF), as emphasized in Chapter 2. Thus, we would like to find out how close $\hat{\beta}_1$ is to the true β_1 or how close $\hat{\sigma}^2$ is to the true σ^2. For instance, in Example 3.2, we estimated the SRF as shown in Eq. (3.7.2). But since this regression is based on a sample of 55 families, how do we know that the estimated MPC of 0.4368 represents the (true) MPC in the population as a whole?

Therefore, since $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\sigma}^2$ are random variables, we need to find out their probability distributions, for without that knowledge we will not be able to relate them to their true values.

4.1 THE PROBABILITY DISTRIBUTION OF DISTURBANCES u_i

To find out the probability distributions of the OLS estimators, we proceed as follows. Specifically, consider $\hat{\beta}_2$. As we showed in Appendix 3A.2,

$$\hat{\beta}_2 = \sum k_i Y_i \qquad \textbf{(4.1.1)}$$

where $k_i = x_i / \sum x_i^2$. But since the X's are assumed fixed, or nonstochastic, because ours is conditional regression analysis, conditional on the fixed values of X_i, Eq. (4.1.1) shows that $\hat{\beta}_2$ is a *linear* function of Y_i, which is random by assumption. But since $Y_i = \beta_1 + \beta_2 X_i + u_i$, we can write (4.1.1) as

$$\hat{\beta}_2 = \sum k_i(\beta_1 + \beta_2 X_i + u_i) \qquad \textbf{(4.1.2)}$$

Because k_i, the betas, and X_i are all fixed, $\hat{\beta}_2$ is ultimately a *linear* function of the random variable u_i, which is random by assumption. Therefore, the probability distribution of $\hat{\beta}_2$ (and also of $\hat{\beta}_1$) will depend on the assumption made about the probability distribution of u_i. And since knowledge of the probability distributions of OLS estimators is necessary to draw inferences about their population values, the nature of the probability distribution of u_i assumes an extremely important role in hypothesis testing.

Since the method of OLS does not make any assumption about the probabilistic nature of u_i, it is of little help for the purpose of drawing inferences about the PRF from the SRF, the Gauss–Markov theorem notwithstanding. This void can be filled if we are willing to assume that the u's follow some probability distribution. For reasons to be explained shortly, in the regression context it is usually assumed that the u's follow the normal distribution. Adding the normality assumption for u_i to the assumptions of the classical linear regression model (CLRM) discussed in Chapter 3, we obtain what is known as the **classical normal linear regression model (CNLRM).**

4.2 THE NORMALITY ASSUMPTION FOR u_i

The classical *normal* linear regression model assumes that each u_i is distributed *normally* with

Mean:	$E(u_i) = 0$	**(4.2.1)**
Variance:	$E[u_i - E(u_i)]^2 = E(u_i^2) = \sigma^2$	**(4.2.2)**
cov (u_i, u_j):	$E\{[(u_i - E(u_i)][u_j - E(u_j)]\} = E(u_i u_j) = 0 \quad i \neq j$	**(4.2.3)**

The assumptions given above can be more compactly stated as

$$u_i \sim N(0, \sigma^2) \qquad \textbf{(4.2.4)}$$

where the symbol ∼ means *distributed as* and N stands for the *normal distribution,* the terms in the parentheses representing the two parameters of the normal distribution, namely, the mean and the variance.

As noted in **Appendix A,** for **two normally distributed variables, zero covariance or correlation means independence of the two variables.** Therefore, with the normality assumption, (4.2.4) means that u_i and u_j are not only uncorrelated but are also independently distributed.

Therefore, we can write (4.2.4) as

$$u_i \sim \text{NID}(0, \sigma^2) \qquad \textbf{(4.2.5)}$$

where **NID** stands for *normally and independently distributed.*

Why the Normality Assumption?

Why do we employ the normality assumption? There are several reasons:

1. As pointed out in Section 2.5, u_i represent the combined influence (on the dependent variable) of a large number of independent variables that are not explicitly introduced in the regression model. As noted, we hope that the influence of these omitted or neglected variables is small and at best random. Now by the celebrated **central limit theorem (CLT)** of statistics (see **Appendix A** for details), it can be shown that if there are a large number of independent and identically distributed random variables, then, with a few exceptions, the distribution of their sum tends to a normal distribution as the number of such variables increase indefinitely.[1] It is the CLT that provides a theoretical justification for the assumption of normality of u_i.

2. A variant of the CLT states that, even if the number of variables is not very large or if these variables are not strictly independent, their sum may still be normally distributed.[2]

3. With the normality assumption, the probability distributions of OLS estimators can be easily derived because, as noted in **Appendix A,** one property of the normal distribution is that **any linear function of normally distributed variables is itself normally distributed.** As we discussed earlier, OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are linear functions of u_i. Therefore, if u_i are normally distributed, so are $\hat{\beta}_1$ and $\hat{\beta}_2$, which makes our task of hypothesis testing very straightforward.

4. The normal distribution is a comparatively simple distribution involving only two parameters (mean and variance); it is very well known and

[1]For a relatively simple and straightforward discussion of this theorem, see Sheldon M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists,* 2d ed., Harcourt Academic Press, New York, 2000, pp. 193–194. One exception to the theorem is the Cauchy distribution, which has no mean or higher moments. See M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics,* Charles Griffin & Co., London, 1960, vol. 1, pp. 248–249.

[2]For the various forms of the CLT, see Harald Cramer, *Mathematical Methods of Statistics,* Princeton University Press, Princeton, N.J., 1946, Chap. 17.

its theoretical properties have been extensively studied in mathematical statistics. Besides, many phenomena seem to follow the normal distribution.

5. Finally, if we are dealing with a small, or finite, sample size, say data of less than 100 observations, the normality assumption assumes a critical role. It not only helps us to derive the exact probability distributions of OLS estimators but also enables us to use the t, F, and χ^2 statistical tests for regression models. The statistical properties of t, F, and χ^2 probability distributions are discussed in **Appendix A.** As we will show subsequently, if the sample size is reasonably large, we may be able to relax the normality assumption.

A cautionary note: Since we are "imposing" the normality assumption, it behooves us to find out in practical applications involving small sample size data whether the normality assumption is appropriate. Later, we will develop some tests to do just that. Also, later we will come across situations where the normality assumption may be inappropriate. But until then we will continue with the normality assumption for the reasons discussed previously.

4.3 PROPERTIES OF OLS ESTIMATORS UNDER THE NORMALITY ASSUMPTION

With the assumption that u_i follow the normal distribution as in (4.2.5), the OLS estimators have the following properties; **Appendix A** provides a general discussion of the desirable statistical properties of estimators.

1. They are unbiased.

2. They have minimum variance. Combined with 1, this means that they are **minimum-variance unbiased,** or **efficient estimators.**

3. They have **consistency;** that is, as the sample size increases indefinitely, the estimators converge to their true population values.

4. $\hat{\beta}_1$ (being a linear function of u_i) is *normally distributed* with

$$\text{Mean:} \qquad E(\hat{\beta}_1) = \beta_1 \tag{4.3.1}$$

$$\text{var}\,(\hat{\beta}_1): \qquad \sigma^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 \qquad = (3.3.3) \tag{4.3.2}$$

Or more compactly,

$$\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2_{\hat{\beta}_1}\right)$$

Then by the properties of the normal distribution the variable Z, which is defined as

$$Z = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \tag{4.3.3}$$

follows the **standard normal distribution,** that is, a normal distribution with zero mean and unit (= 1) variance, or

$$Z \sim N(0, 1)$$

5. $\hat{\beta}_2$ (being a linear function of u_i) is *normally* distributed with

$$\text{Mean:} \qquad E(\hat{\beta}_2) = \beta_2 \tag{4.3.4}$$

$$\text{var}\,(\hat{\beta}_2): \qquad \sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum x_i^2} \qquad = \textbf{(3.3.1)} \tag{4.3.5}$$

Or, more compactly,

$$\hat{\beta}_2 \sim N\left(\beta_2, \sigma^2_{\hat{\beta}_2}\right)$$

Then, as in (4.3.3),

$$Z = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \tag{4.3.6}$$

also follows the standard normal distribution.

Geometrically, the probability distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$ are shown in Figure 4.1.

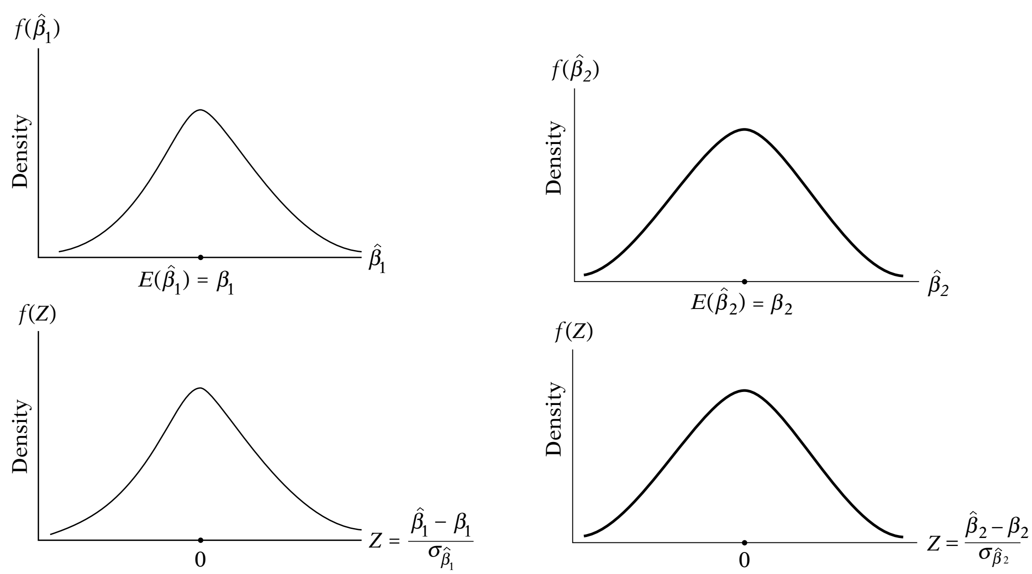

FIGURE 4.1 Probability distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$.

6. $(n-2)(\hat{\sigma}^2/\sigma^2)$ is distributed as the χ^2 (chi-square) distribution with $(n-2)$df.[3] This knowledge will help us to draw inferences about the true σ^2 from the estimated σ^2, as we will show in Chapter 5. (The chi-square distribution and its properties are discussed in **Appendix A.**)

7. $(\hat{\beta}_1, \hat{\beta}_2)$ are distributed independently of $\hat{\sigma}^2$. The importance of this will be explained in the next chapter.

8. *$\hat{\beta}_1$ and $\hat{\beta}_2$ have minimum variance in the entire class of unbiased estimators, whether linear or not.* This result, due to Rao, is very powerful because, unlike the Gauss–Markov theorem, it is not restricted to the class of linear estimators only.[4] Therefore, we can say that the least-squares estimators are **best unbiased estimators (BUE);** that is, they have minimum variance in the entire class of unbiased estimators.

To sum up: The important point to note is that the normality assumption enables us to derive the probability, or sampling, distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$ (both normal) and $\hat{\sigma}^2$ (related to the chi square). As we will see in the next chapter, this simplifies the task of establishing confidence intervals and testing (statistical) hypotheses.

In passing, note that, with the assumption that $u_i \sim N(0, \sigma^2)$, Y_i, being a linear function of u_i, is itself normally distributed with the mean and variance given by

$$E(Y_i) = \beta_1 + \beta_2 X_i \qquad \textbf{(4.3.7)}$$

$$\operatorname{var}(Y_i) = \sigma^2 \qquad \textbf{(4.3.8)}$$

More neatly, we can write

$$Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2) \qquad \textbf{(4.3.9)}$$

4.4 THE METHOD OF MAXIMUM LIKELIHOOD (ML)

A method of point estimation with some stronger theoretical properties than the method of OLS is the method of **maximum likelihood (ML).** Since this method is slightly involved, it is discussed in the appendix to this chapter. For the general reader, it will suffice to note that if u_i are assumed to be normally distributed, as we have done for reasons already discussed, the ML and OLS estimators of the regression coefficients, the β's, are identical, and this is true of simple as well as multiple regressions. The ML estimator of σ^2 is $\sum \hat{u}_i^2/n$. This estimator is biased, whereas the OLS estimator

---

[3]The proof of this statement is slightly involved. An accessible source for the proof is Robert V. Hogg and Allen T. Craig, *Introduction to Mathematical Statistics,* 2d ed., Macmillan, New York, 1965, p. 144.

[4]C. R. Rao, *Linear Statistical Inference and Its Applications,* John Wiley & Sons, New York, 1965, p. 258.

of $\sigma^2 = \sum \hat{u}_i^2/(n-2)$, as we have seen, is unbiased. But comparing these two estimators of σ^2, we see that as the sample size n gets larger the two estimators of σ^2 tend to be equal. Thus, asymptotically (i.e., as n increases indefinitely), the ML estimator of σ^2 is also unbiased.

Since the method of least squares with the added assumption of normality of u_i provides us with all the tools necessary for both estimation and hypothesis testing of the linear regression models, there is no loss for readers who may not want to pursue the maximum likelihood method because of its slight mathematical complexity.

4.5 SUMMARY AND CONCLUSIONS

1. This chapter discussed the classical *normal* linear regression model (CNLRM).

2. This model differs from the classical linear regression model (CLRM) in that it specifically assumes that the disturbance term u_i entering the regression model is normally distributed. The CLRM does not require any assumption about the probability distribution of u_i; it only requires that the mean value of u_i is zero and its variance is a finite constant.

3. The theoretical justification for the normality assumption is the **central limit theorem.**

4. Without the normality assumption, under the other assumptions discussed in Chapter 3, the Gauss–Markov theorem showed that the OLS estimators are BLUE.

5. With the additional assumption of normality, the OLS estimators are not only **best unbiased estimators (BUE)** but also follow well-known probability distributions. The OLS estimators of the intercept and slope are themselves normally distributed and the OLS estimator of the variance of u_i $(= \hat{\sigma}^2)$ is related to the chi-square distribution.

6. In Chapters 5 and 8 we show how this knowledge is useful in drawing inferences about the values of the population parameters.

7. An alternative to the least-squares method is the method of **maximum likelihood (ML).** To use this method, however, one must make an assumption about the probability distribution of the disturbance term u_i. In the regression context, the assumption most popularly made is that u_i follows the normal distribution.

8. Under the normality assumption, the ML and OLS estimators of the intercept and slope parameters of the regression model are identical. However, the OLS and ML estimators of the variance of u_i are different. In large samples, however, these two estimators converge.

9. Thus the ML method is generally called a *large-sample method.* The ML method is of broader application in that it can also be applied to regression models that are nonlinear in the parameters. In the latter case, OLS is generally not used. For more on this, see Chapter 14.

10. In this text, we will largely rely on the OLS method for practical reasons: (*a*) Compared to ML, the OLS is easy to apply; (*b*) the ML and OLS estimators of β_1 and β_2 are identical (which is true of multiple regressions too); and (*c*) even in moderately large samples the OLS and ML estimators of σ^2 do not differ vastly.

However, for the benefit of the mathematically inclined reader, a brief introduction to ML is given in the appendix to this chapter and also in **Appendix A**.

APPENDIX 4A

4A.1 MAXIMUM LIKELIHOOD ESTIMATION OF TWO-VARIABLE REGRESSION MODEL

Assume that in the two-variable model $Y_i = \beta_1 + \beta_2 X_i + u_i$ the Y_i are normally and independently distributed with mean $= \beta_1 + \beta_2 X_i$ and variance $= \sigma^2$. [See Eq. (4.3.9).] As a result, the joint probability density function of Y_1, $Y_2, \ldots, Y_n$, given the preceding mean and variance, can be written as

$$f(Y_1, Y_2, \ldots, Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2)$$

But in view of the independence of the Y's, this joint probability density function can be written as a product of n individual density functions as

$$\begin{aligned} &f(Y_1, Y_2, \ldots, Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2) \\ &\quad = f(Y_1 \mid \beta_1 + \beta_2 X_i, \sigma^2) f(Y_2 \mid \beta_1 + \beta_2 X_i, \sigma^2) \cdots f(Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2) \end{aligned} \tag{1}$$

where

$$f(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\} \tag{2}$$

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

(*Note:* exp means e to the power of the expression indicated by {}.)

Substituting (2) for each Y_i into (1) gives

$$f(Y_i, Y_2, \ldots, Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2) = \frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^n} \exp\left\{-\frac{1}{2}\sum\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\} \tag{3}$$

If $Y_1, Y_2, \ldots, Y_n$ are known or given, but β_1, β_2, and σ^2 are not known, the function in (3) is called a **likelihood function,** denoted by LF($\beta_1, \beta_2, \sigma^2$), and written as[1]

$$\text{LF}(\beta_1, \beta_2, \sigma^2) = \frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^n} \exp\left\{-\frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\} \tag{4}$$

The **method of maximum likelihood,** as the name indicates, consists in estimating the unknown parameters in such a manner that the probability of observing the given *Y*'s is as high (or maximum) as possible. Therefore, we have to find the maximum of the function (4). This is a straightforward exercise in differential calculus. For differentiation it is easier to express (4) in the log term as follows.[2] (*Note:* ln = natural log.)

$$\begin{aligned}\ln \text{LF} &= -n \ln \sigma - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2} \\ &= -\frac{n}{2}\ln \sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\end{aligned} \tag{5}$$

Differentiating (5) partially with respect to β_1, β_2, and σ^2, we obtain

$$\frac{\partial \ln \text{LF}}{\partial \beta_1} = -\frac{1}{\sigma^2}\sum (Y_i - \beta_1 - \beta_2 X_i)(-1) \tag{6}$$

$$\frac{\partial \ln \text{LF}}{\partial \beta_2} = -\frac{1}{\sigma^2}\sum (Y_i - \beta_1 - \beta_2 X_i)(-X_i) \tag{7}$$

$$\frac{\partial \ln \text{LF}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum (Y_i - \beta_1 - \beta_2 X_i)^2 \tag{8}$$

Setting these equations equal to zero (the first-order condition for optimization) and letting $\tilde{\beta}_1$, $\tilde{\beta}_2$, and $\tilde{\sigma}^2$ denote the ML estimators, we obtain[3]

$$\frac{1}{\tilde{\sigma}^2}\sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i) = 0 \tag{9}$$

$$\frac{1}{\tilde{\sigma}^2}\sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i) X_i = 0 \tag{10}$$

$$-\frac{n}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4}\sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i)^2 = 0 \tag{11}$$

---

[1] Of course, if β_1, β_2, and σ^2 are known but the Y_i are not known, (4) represents the joint probability density function—the probability of jointly observing the Y_i.

[2] Since a log function is a monotonic function, ln LF will attain its maximum value at the same point as LF.

[3] We use ˜ (tilde) for ML estimators and ˆ (cap or hat) for OLS estimators.

After simplifying, Eqs. (9) and (10) yield

$$\sum Y_i = n\tilde{\beta}_1 + \tilde{\beta}_2 \sum X_i \tag{12}$$

$$\sum Y_i X_i = \tilde{\beta}_1 \sum X_i + \tilde{\beta}_2 \sum X_i^2 \tag{13}$$

which are precisely the *normal equations* of the least-squares theory obtained in (3.1.4) and (3.1.5). Therefore, the ML estimators, the $\tilde{\beta}$'s, are the same as the OLS estimators, the $\hat{\beta}$'s, given in (3.1.6) and (3.1.7). This equality is not accidental. Examining the likelihood (5), we see that the last term enters with a negative sign. Therefore, maximizing (5) amounts to minimizing this term, which is precisely the least-squares approach, as can be seen from (3.1.2).

Substituting the ML (= OLS) estimators into (11) and simplifying, we obtain the ML estimator of $\tilde{\sigma}^2$ as

$$\begin{aligned}\tilde{\sigma}^2 &= \frac{1}{n}\sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i)^2 \\ &= \frac{1}{n}\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \\ &= \frac{1}{n}\sum \hat{u}_i^2\end{aligned} \tag{14}$$

From (14) it is obvious that the ML estimator $\tilde{\sigma}^2$ differs from the OLS estimator $\hat{\sigma}^2 = [1/(n-2)]\sum \hat{u}_i^2$, which was shown to be an unbiased estimator of σ^2 in Appendix 3A, Section 3A.5. Thus, the ML estimator of σ^2 is biased. The magnitude of this bias can be easily determined as follows.

Taking the mathematical expectation of (14) on both sides, we obtain

$$\begin{aligned}E(\tilde{\sigma}^2) &= \frac{1}{n}E\left(\sum \hat{u}_i^2\right) \\ &= \left(\frac{n-2}{n}\right)\sigma^2 \qquad \text{using Eq. (16) of Appendix 3A, Section 3A.5} \\ &= \sigma^2 - \frac{2}{n}\sigma^2\end{aligned} \tag{15}$$

which shows that $\tilde{\sigma}^2$ is biased downward (i.e., it underestimates the true σ^2) in small samples. But notice that as n, the sample size, increases indefinitely, the second term in (15), the bias factor, tends to be zero. Therefore, *asymptotically* (i.e., in a very large sample), $\tilde{\sigma}^2$ is *unbiased* too, that is, $\lim E(\tilde{\sigma}^2) = \sigma^2$ as $n \to \infty$. It can further be proved that $\tilde{\sigma}^2$ is also a

consistent estimator[4]; that is, as n increases indefinitely $\tilde{\sigma}^2$ converges to its true value σ^2.

4A.2 MAXIMUM LIKELIHOOD ESTIMATION OF FOOD EXPENDITURE IN INDIA

Return to Example 3.2 and regression (3.7.2), which gives the regression of food expenditure on total expenditure for 55 rural households in India. Since under the normality assumption the OLS and ML estimators of the regression coefficients are the same, we obtain the ML estimators as $\tilde{\beta}_1 = \hat{\beta}_1 = 94.2087$ and $\tilde{\beta}_2 = \hat{\beta}_2 = 0.4386$. The OLS estimator of σ^2 is $\hat{\sigma}^2 = 4469.6913$, but the ML estimator is $\tilde{\sigma}^2 = 4407.1563$, which is smaller than the OLS estimator. As noted, in small samples the ML estimator is downward biased; that is, on average it underestimates the true variance σ^2. Of course, as you would expect, as the sample size gets bigger, the difference between the two estimators will narrow. Putting the values of the estimators in the log likelihood function, we obtain the value of −308.1625. If you want the maximum value of the LF, just take the antilog of −308.1625. No other values of the parameters will give you a higher probability of obtaining the sample that you have used in the analysis.

APPENDIX 4A EXERCISES

4.1. "If two random variables are statistically independent, the coefficient of correlation between the two is zero. But the converse is not necessarily true; that is, zero correlation does not imply statistical independence. However, if two variables are normally distributed, zero correlation necessarily implies statistical independence." Verify this statement for the following joint probability density function of two normally distributed variables Y_1 and Y_2 (this joint probability density function is known as the **bivariate normal probability density function**):

$$f(Y_1, Y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\right.$$
$$\left. \times \left[\left(\frac{Y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(Y_1-\mu_1)(Y_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{Y_2-\mu_2}{\sigma^2}\right)^2\right]\right\}$$

[4]See **App. A** for a general discussion of the properties of the maximum likelihood estimators as well as for the distinction between asymptotic unbiasedness and consistency. Roughly speaking, in asymptotic unbiasedness we try to find out the $\lim E(\tilde{\sigma}_n^2)$ as n tends to infinity, where n is the sample size on which the estimator is based, whereas in consistency we try to find out how $\tilde{\sigma}_n^2$ behaves as n increases indefinitely. Notice that the unbiasedness property is a repeated sampling property of an estimator based on a sample of given size, whereas in consistency we are concerned with the behavior of an estimator as the sample size increases indefinitely.

where μ_1 = mean of Y_1
μ_2 = mean of Y_2
σ_1 = standard deviation of Y_1
σ_2 = standard deviation of Y_2
ρ = coefficient of correlation between Y_1 and Y_2

4.2. By applying the second-order conditions for optimization (i.e., second-derivative test), show that the ML estimators of β_1, β_2, and σ^2 obtained by solving Eqs. (9), (10), and (11) do in fact maximize the likelihood function (4).

4.3. A random variable X follows the **exponential distribution** if it has the following probability density function (PDF):

$$f(X) = (1/\theta)e^{-X/\theta} \qquad \text{for } X > 0$$
$$= 0 \qquad \text{elsewhere}$$

where $\theta > 0$ is the parameter of the distribution. Using the ML method, show that the ML estimator of θ is $\hat{\theta} = \sum X_i/n$, where n is the sample size. That is, show that the ML estimator of θ is the sample mean $\bar{X}$.