ECONOMETRICS

GREGORY C. CHOW

ECONOMETRICS

Gregory C. Chow

Professor of Economics
Princeton University

McGraw-Hill Book Company

New York St. Louis San Francisco Auckland Bogotá Hamburg
Johannesburg London Madrid Mexico Montreal New Delhi
Panama Paris São Paulo Singapore Sydney Tokyo Toronto

This book was set in Times Roman by Santype-Byrd.
The editors were Peter J. Dougherty and Scott Amerman;
the production supervisor was Leroy A. Young.
Halliday Lithograph Corporation was printer and binder.

ECONOMETRICS

234567890 HALHAL 89876543

ISBN 0-07-010847-1

Library of Congress Cataloging in Publication Data

Chow, Gregory C., date
Econometrics.

(Economics Handbook Series)
Includes bibliographies and index.
1. Econometrics. I. Title.
HB139.C483 1983 330'.028 82-20903
ISBN 0-07-010847-1

PREFACE

Since the 1970s, the development of econometrics has been so rapid that it has become difficult to find a textbook that treats the new topics and provides an updated perspective of the field. This text is intended to fill this need.

Most existing texts cover only the topics of Chaps. 1 to 5, the topics of Chaps. 6 to 12 being covered only infrequently. In order to incorporate new developments in a text, as it has been done in field after field for several decades, one must emphasize the basic ideas that underlie the subject. I believe from my teaching experience in Princeton that it is possible to cover the topics of seven chapters in about the same time and with the same effort on the part of students as previously spent in covering the topics of the first five chapters alone. This requires omitting nonessential details in, and improving the exposition of, previous material, as well as exposing the threads that connect all the material, old and new. After the new material has been integrated, the reader will gain a more up-to-date perspective of the subject.

This book is intended mainly for graduate students, although as the curriculum is constantly being upgraded, more and more advanced undergraduates will find the less theoretical parts readable. It assumes familiarity with the basic techniques of statistical inference and some exposure to matrix algebra. Naturally, more preparation in these two areas will make this text easier to read and simpler to comprehend. Chapter 1 can be used to test the reader's preparation. A reader with the minimum required preparation should find Secs. 1.1 to 1.5 easily comprehensible but may find the later starred sections difficult. In this case, I would recommend studying the nonstarred sections of Chap. 2 before returning to the more theoretical starred sections of Chap. 1. In fact, *an applications-*

oriented reader may skip the starred sections or read only the results therein without going through the proofs.

As another means of accommodating students with different backgrounds in statistics and mathematics, I have chosen to weave the required background material into the body of the text rather than presenting appendixes, which are usually dry and formal. This approach has the advantages of motivating the material and helping the student learn it while using it. Material that is used more often will be more firmly absorbed, as it should be.

The level of abstraction at which to present econometrics is a matter of taste. This book has interwoven empirical material to show students how empirical investigations in econometrics are conducted and to motivate the theoretical material. Some readers would prefer more empirical material than is actually presented and some less. Some would prefer having less theorem proving and others would prefer having more and at a higher level of mathematical abstraction. The level selected reflects the author's opinion concerning what the first-year graduate student in economics should be trained to do in the field of econometrics.

Although this text has a broader coverage than existing texts, it cannot possibly cover all important topics, as a textbook is not an encyclopedia. The selection of topics is also a matter of taste. Given the selection of topics, the choice of the tools and concepts to be presented within each topic has been guided by two criteria: the usefulness of the methods in applied work and the usefulness of the theoretical ideas in the further development of econometric methods. For example, the t test is useful in applied work, and the idea of maximum likelihood is useful for developing new estimators. Once the basic tools have been mastered, the student can apply them to study and solve new problems not covered in this text.

How this book should be used depends on the preparation of the students. As mentioned above, the starred sections can be omitted by applications-oriented readers. For the students of applied econometrics who have not had the material in Secs. 1.6 to 1.11, a one-semester course may cover Chaps. 1 to 5, omitting most of the proofs in the starred sections. For the students of theoretical econometrics who have not had the material in Secs. 1.6 to 1.11, a one-semester course may include Chaps. 1 to 5, covering the above sections, and Secs. 2.13, 3.2, and 3.3 *after* the nonstarred sections of Chaps. 1 to 3 have been studied. Less time will be devoted to the empirical studies. Students having had one semester of statistical methods including the materials of Secs. 1.6 to 1.11 may take up additional topics from Chaps. 6, 7, or 9. Chapters 6 to 12 can serve as a text for a second semester in econometrics.

I am indebted to the students at Princeton who helped me learn the material in this text, some using drafts of various chapters and offering comments, including George Mailath, Loretta Mester, and In-Koo Cho, who have also helped prepare the index. Thanks are due to Takeshi Amemiya and John B. Taylor who read Chap. 11, and to David Brownstone who read Chap. 8, and to Adrian

Pagan and Gary Skoog who read many chapters and provided useful comments. It is impossible to express sufficient gratitude to my colleague Richard Quandt, who read the entire manuscript and made many useful suggestions for improvement. Pia Ellen has typed drafts of the manuscript with remarkable efficiency and good spirit. Without her help, the book would not have been completed. I would like to thank all the publishers for granting permission to reprint material written by me and published by them. References are given in the text. Results of research supported by the National Science Foundation through several grants find their way into many sections in this book, especially in Chaps. 7 and 9 to 12.

Gregory C. Chow

CONTENTS

* Starred sections may be omitted without loss of continuity; see Preface.

CHAPTER

ONE

SIMPLE LINEAR REGRESSION

1.1 WHAT IS ECONOMETRICS?

Econometrics is the art and science of using statistical methods for the measurement of economic relations. In the practice of econometrics, economic theory, institutional information, and other assumptions are relied upon to formulate a statistical model, or a set of statistical hypotheses, to explain the phenomena in question. Econometric methods are used to estimate the parameters of the model, to test hypotheses concerning them, and to generate forecasts from the model. The formulation of an econometric model is an art, just as using knowledge of architecture to design a building is an art. A good econometrician uses sound judgment to bring the relevant knowledge in economics to bear in formulating a useful model. The most important variables are selected while the nonessential ones are discarded. The crucial relationships are formulated and incorporated in the model. Care is taken to ensure that the statistical data used actually correspond to the variables to be measured according to theoretical considerations.

Given a set of requirements in the construction of a building, two good architects will come up with two different designs. Both may serve the purposes well. Similarly, given the same objectives, two econometricians are not likely to come up with two identical models, but both may capture the essential elements of the problem sufficiently to be useful. Good econometric models, like good architectural designs, can serve as prototypes to be followed in future investigations. The art of formulating a good econometric model is difficult to learn. One needs a solid command of the tools of economic analysis and sound judgment to select the essential variables of the problem. In a Walrasian system of general equilibrium all economic variables are related, but only a subset of variables will be selected in a particular investigation. One can read the best econo-

metric studies to see how they were done, in the same way that an architect reads the best designs or an artist studies masterpieces of art. Finally, one can practice building one's own models and thus learn by doing.

This book is devoted mainly to an easier aspect of the practice of econometrics, namely, the study of some of the statistical methods most useful for drawing inferences from an econometric model. Knowledge of econometric methods is essential, though by no means sufficient, for the construction of good econometric models, as we have just pointed out. Our presentation presumes that the reader is familiar with the basic ideas of statistical inference. To review some of the basic ideas and to indicate the level of preparation suitable for reading this book, we shall introduce the required tools of statistics in this chapter and apply them to the model of simple linear regression as a warmup exercise. Most of the techniques presented later in this book can be viewed as generalizations, extensions, or modifications of this simple model.

At the beginning of an econometric investigation, an econometrician needs to know clearly what economic phenomena are to be explained, what important factors will contribute to the explanation, how these factors should be measured, what quantitative relationships exist, how such relations can be estimated or tested, and what conclusions can be drawn from the investigation. Consider the example of studying the demand for apples in the United States. The annual consumption of apples through time may be selected as the phenomenon to be explained. The important factors explaining the demand for apples may be the price of apples and income of the consumer. One may choose the per capita annual consumption of apples (in pounds) as the dependent variable. The price of apples per pound, deflated by a consumer price index, and per capita disposable income, also deflated by a consumer price index, may serve as the explanatory variables. The logarithm of per capita apple consumption may be assumed to be a linear function of the log of relative price, the log of per capita real disposable income, and a normally distributed random disturbance summarizing the combined effects of omitted factors. Time-series data may be used to estimate this linear relation. The coefficients of log price and log income will be interpreted respectively as the price and income elasticities of demand for apples. One objective of the investigation may be to estimate these demand elasticities. Another may be to test the hypotheses that both elasticities are less than 1 in absolute value. A third may be to use the estimated relation to forecast the demand for apples 5 years from now.

In the above illustration, many issues will have to be resolved by the econometrician. For example, is deflation of total apple consumption by total population in the United States sufficient to account for the effects of demographic factors on apple consumption? Is the age distribution relevant? Is mean income sufficient to explain apple consumption without taking the distribution of income into account? Is the relationship approximately linear in the logarithms of the variables? Does one have to account for the effects of lagged incomes on consumption? Is it useful to combine cross-section data with time-series data to estimate the price and income elasticities? Does the coefficient of log price in the postulated linear relation measure solely the price elasticity of demand, rather

than the elasticity of supply? Is it necessary to formulate a supply equation for apples and estimate a system of two equations simultaneously?

The study of demand for apples is an easy econometric problem, and yet all these issues and others must be resolved. Consider the more difficult problem of studying the demand for computers in the United States. Just the measurement of the dependent variable requires some serious thought. Since there are large and small computers, or computers with different computing powers and capabilities, how is the total quantity of computers to be measured? Since the quality of the computers, however measured, has improved so rapidly through time, how should the effect of this rapid technological change be incorporated in a study of the demand for computers? Related to the measurement of the quantity of computers is the problem of measuring the price per unit of computers. If the quantity of computers can be measured appropriately, one will know what 1 unit of computers is and the price per unit can be determined accordingly. These probblems are only illustrative of the issues an econometrician faces in the formulation of an econometric model. Other issues have been stated at the beginning of the paragraph before last. The problem of the demand for computers is considered in Sec. 1.12. The issues in formulating an econometric model cannot be systematically discussed in this book but will be illustrated by case studies cited in Chaps. 1, 2, and 4. Before embarking on a statistical analysis, the econometrician should have resolved all these issues and planned ahead, deciding how conclusions should be drawn before performing the statistical computations. Having made these introductory remarks, we begin by describing the model of simple linear regression.

1.2 MODEL OF SIMPLE LINEAR REGRESSION

Let y_i denote the ith observation of the dependent variable and x_i denote the associated explanatory variable. A simple linear regression model can be written as

$$y_i = \alpha + \beta x_i + \epsilon_i \qquad i = 1, \ldots, n \tag{1}$$

where the residual or disturbance terms ϵ_i are assumed to be normal and independent, each having mean zero and variance σ^2. The observations on x_i are treated as fixed numbers, except when specified otherwise, as in Secs. 1.6, 1.7, and 3.8. That is, a probability distribution for x_i is not postulated. The parameters of this model are α, β, and σ^2. The mean of the random variable y_i is $\alpha + \beta x_i$, as we observe by taking the mathematical expectations (or means) of both sides of (1). The mean is thus a linear function of x; the mean is called a *regression function.* As an example, consider n families selected from Princeton, New Jersey. Let y_i denote the logarithm of the quantity of apples consumed by the ith family during a given period and x_i denote the logarithm of its income during the same period. The mean of y_i can be assumed to be a linear function of x_i. Given x_i, that is, for families having log income equal to x_i, the distribution of log apple consumption is assumed to be normal, having mean $\alpha + \beta x_i$ and standard deviation σ.

In classical statistics the two main problems are to estimate the unknown

parameters, α, β, and σ^2 in this case, and to test hypotheses concerning them. These topics will be discussed in turn.

1.3 POINT ESTIMATION

Let point estimates of the parameters α and β be denoted respectively by a and b. Other methods will be presented in due course, but first we discuss a popular method for finding these estimates, namely, the method of least squares. In this method one chooses the values of a and b which minimize the sum of squares of the deviations of y_i from the estimated regression line $a + bx_i$, that is,

$$\sum_{i=1}^{n} [y_i - (a + bx_i)]^2$$

Setting to zero the derivative of this sum with respect to a, we obtain

$$a = \frac{\sum_{i=1}^{n} y_i - b \sum_{i=1}^{n} x_i}{n} \equiv \bar{y} - b\bar{x} \tag{2}$$

Substituting $\bar{y} - b\bar{x}$ for a in this sum and setting to zero its derivative with respect to b, we obtain

$$\begin{aligned} b &= \left[\sum_i x_i(x_i - \bar{x})\right]^{-1} \sum_i x_i(y_i - \bar{y}) \\ &= \left[\sum_i (x_i - \bar{x})^2\right]^{-1} \sum_i (x_i - \bar{x})(y_i - \bar{y}) \end{aligned} \tag{3}$$

where we have used $\sum_i \bar{x}(x_i - \bar{x}) = 0$ and $\sum_i \bar{x}(y_i - \bar{y}) = 0$. Equations (2) and (3) are the *normal equations* for computing the estimates a and b.

To decide whether these formulas or *estimators* are good we consider their sampling distributions. As seen from (2) and (3), the estimators a and b are functions of the random variables $y_1, \ldots, y_n$ and therefore random variables themselves. The term *estimates* refers to the particular numbers representing a and b which we obtain by using a particular sample $(y_1, \ldots, y_n)$. The term *estimators* refers to possible numbers a and b which we could obtain by (hypothetically) drawing repeated samples of $(y_1, \ldots, y_n)$ given the values of $x_1, \ldots, x_n$. An estimator is a random variable; an estimate is a number which is computed by using the data for one particular sample. To evaluate the mean and variance of the distribution of b or of the random variable b we substitute $\alpha + \beta x_i + \epsilon_i$ for y_i in (3) and write b as a function of ϵ_i

$$\begin{aligned} b &= \left[\sum_i (x_i - \bar{x})^2\right]^{-1} \left[\sum_i (x_i - \bar{x}) y_i\right] \\ &= \left[\sum_i (x_i - \bar{x})^2\right]^{-1} \left[\sum_i (x_i - \bar{x})(\alpha + \beta x_i + \epsilon_i)\right] \end{aligned}$$

$$= \left[\sum_i (x_i - \bar{x})x_i\right]^{-1}\left[\sum_i (x_i - \bar{x})\beta x_i + \sum_i (x_i - \bar{x})\epsilon_i\right]$$

$$= \beta + \left[\sum_i (x_i - \bar{x})^2\right]^{-1}\left[\sum_i (x_i - \bar{x})\epsilon_i\right] \tag{4}$$

The mean of the random variable b is obtained by taking the mathematical expectation of (4) yielding

$$E[b] = E[\beta] + \left[\sum_i (x_i - \bar{x})^2\right]^{-1}\left[\sum_i (x_i - \bar{x})E\epsilon_i\right] = \beta \tag{5}$$

where we have treated $\sum_i (x_i - \bar{x})^2$ as a constant and taken the expectation of the sum $\sum_i (x_i - \bar{x})\epsilon_i$ as the sum $\sum_i E[(x_i - \bar{x})\epsilon_i]$ of the expectations, each $E\epsilon_i$ being zero by assumption. Since $E[b] = \beta$, b is an *unbiased estimator* of β.

To find the variance of b we use (4) and (5) to evaluate

$$\text{Var } b \equiv E[(b - Eb)^2] = E\left\{\left[\sum_i (x_i - \bar{x})^2\right]^{-1}\left[\sum_i (x_i - \bar{x})\epsilon_i\right]\right\}^2$$

$$= \left[\sum_i (x_i - \bar{x})^2\right]^{-2} E\left[\sum_i (x_i - \bar{x})\epsilon_i\right]^2$$

$$= \left[\sum_i (x_i - \bar{x})^2\right]^{-2} \sum_i (x_i - \bar{x})^2 E\epsilon_i^2 = \left[\sum_i (x_i - \bar{x})^2\right]^{-1} \sigma^2 \tag{6}$$

where the third line has used the assumption that $E\epsilon_i\epsilon_j = 0$ for $i \neq j$. Thus the variance of the sampling distribution of b is directly proportional to the variance σ^2 of the regression residuals ϵ_i and inversely proportional to the variance of the explanatory variable. Having more variations in x permits a tighter estimate of b.

Note from the first line of (4) that b is a linear function of $y_1, \ldots, y_n$ and is called a *linear* estimator. It is an *unbiased* estimator of β, as shown by (5). In fact, among all linear, unbiased estimators, b has the *smallest variance*, as will be proved in Chap. 2. Thus, the least-squares estimator b is said to be *best linear unbiased*. We leave the evaluation of the mean and variance of the distribution of the estimator a as exercises.

A frequently used point estimate of σ^2 is

$$s^2 = (n - 2)^{-1} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \tag{7}$$

It will be shown in Chap. 2 that $(n - 2)s^2/\sigma^2$ or $\sum_{i=1}^n (y_i - a - bx_i)^2/\sigma^2$ is distributed as χ^2 with $n - 2$ degrees of freedom. Since a $\chi^2(n - 2)$ distribution has a mean equal to $n - 2$, the mean of s^2/σ^2 is 1, or the mean of s^2 is σ^2. In other words, s^2 is an unbiased estimator for σ^2. However, the mean-squared error $E(s^2 - \sigma^2)^2$ can be improved by replacing the denominator $n - 2$ by $n - 1$. This would produce a biased estimator, whose expectation is smaller than σ^2, but the variance of the estimator would also be reduced, leading to a smaller mean-squared error. The estimator s^2 is often used because it is convenient to use standard tables for the χ^2 distribution for its sampling distribution.

1.4 TESTING HYPOTHESES AND INTERVAL ESTIMATION

From (4) we observe that $b - \beta$ is a linear combination of $\epsilon_1, \ldots, \epsilon_n$. Therefore, if the ϵ_i are normal, the linear combination $b - \beta$ will also be normal. Even if ϵ_i are not normal but have a finite variance, the linear combination $b - \beta$ will have a distribution which approaches the normal distribution as the sample size n increases, according to the central limit theorem, discussed in Sec. 1.9. From (6) we obtain the variances of b. Hence

$$\frac{b - \beta}{[\sum_i (x_i - \bar{x})^2]^{-1/2}\sigma} \tag{8}$$

has a standard normal distribution. If σ^2 is known, the statistic (8) can be used to test hypotheses concerning β. Let the null hypothesis be $\beta = \beta_0$ and the alternative hypothesis be $\beta \neq \beta_0$. Let the level of significance be 5 percent. Under the null hypothesis, the probability that the statistic (8) with $\beta = \beta_0$ will be larger than 1.96 or smaller than -1.96 is .05. If this statistic as computed from our sample turns out to exceed 1.96 in absolute value, the null hypothesis $\beta = \beta_0$ will be rejected. If the alternative hypothesis is one-sided, say $\beta < \beta_0$, the null hypothesis will be rejected when the statistic (8) is smaller than -1.64 since the probability for a standard normal random variable to be smaller than this value is .05.

To construct a symmetric interval estimate or confidence interval for β with a confidence coefficient of .95 we use inequalities

$$-1.96 < \frac{b - \beta}{[\sum_i (x_i - \bar{x})^2]^{-1/2}\sigma} < 1.96$$

implying

$$b + 1.96\left[\sum_i (x_i - \bar{x})^2\right]^{-1/2}\sigma > \beta > b - 1.96\left[\sum_i (x_i - \bar{x})^2\right]^{-1/2}\sigma \tag{9}$$

To test the null hypothesis that $\sigma^2 = \sigma_0^2$ we use the fact that under the null hypothesis $(n-2)s^2/\sigma_0^2$ has a $\chi^2(n-2)$ distribution. If this statistic exceeds an upper critical value C_U or falls below a lower critical value C_L according to the $\chi^2(n-2)$ distribution, the null hypothesis will be rejected. Similarly a confidence interval for σ^2 can be constructed by using

$$C_L < \frac{(n-2)s^2}{\sigma^2} < C_U$$

implying

$$\frac{(n-2)s^2}{C_L} > \sigma^2 > \frac{(n-2)s^2}{C_U}$$

To test the null hypothesis $\beta = \beta_0$ when σ^2 is unknown we use the statistic

$$\frac{b - \beta_0}{[\sum_i (x_i - \bar{x})^2]^{-1/2}s} \tag{10}$$

which is the ratio of (8) to s/σ. The numerator (8) is standard normal. The denominator s/σ is the square root of a $\chi^2(n-2)$ variable divided by the degrees of freedom $(n-2)$. As will be shown in Chap. 2, the numerator and the denominator are statistically independent. Therefore, the ratio (10) will have Student's t distribution with $n-2$ degrees of freedom. Hypothesis testing and interval estimation for β can be performed using the $t(n-2)$ distribution for the statistic (10). For example, the .025 upper-tail critical values for $t(10)$ and $t(20)$ are respectively 2.228 and 2.086, compared with 1.960 for the standard normal statistic. A $\chi^2(m)$ variable is the sum of squares of m independent standard normal variables. If $\chi^2(m)$ and $\chi^2(n)$ are independent, the ratio of $\chi^2(m)/m$ to $\chi^2(n)/n$ has an $F(m, n)$ distribution. An $F(1, n)$ variable is the square of a $t(n)$ variable. Tables for the normal, χ^2, t, and F distributions are provided at the end of this book.

1.5 USE OF MATRIX NOTATION

In order to study the model of multiple linear regression, where the mean of y is a linear function of several explanatory variables, it is convenient to use matrix notation. Let the model (1) be written as

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i \qquad i = 1, \ldots, n \tag{11}$$

where we write the original α as β_1 and the original β as β_2, and let the ith observation of the first independent variable x_{i1} be equal to 1 identically so that its coefficient β_1 is the intercept. Denote by y, x_1, x_2, and ϵ, respectively, the column vectors of n observations on the dependent variable, the first and the second explanatory variables, and the random disturbance, namely,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad x_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} \quad x_2 = \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

We can write the n observations of the model (11) as

$$y = \beta_1 x_1 + \beta_2 x_2 + \epsilon \tag{12}$$

The vector y equals the sum of three vectors, $\beta_1 x_1$, $\beta_2 x_2$, and ϵ. When a scalar β_1 multiplies a vector x_1, each element of x_1 is multiplied by the scalar to form a vector $\beta_1 x_1$ as the product. When the three vectors $\beta_1 x_1$, $\beta_2 x_2$, and ϵ are added, their corresponding elements are added to form the elements of the sum, which in our case equals the vector y. The equality sign in (12) signifies that each element of y equals the corresponding element of the vector on the right, obtained as the sum of the three vectors.

The inner product of a row vector and a column vector is the sum of the products of their corresponding elements. For example, let x_1' be the row vector $[x_{11} \quad x_{21} \quad \cdots \quad x_{n1}]$ obtained by taking the transpose (indicated by a prime) of the column vector x_1, and let x_2 be the column vector as defined above. The