themes in modern econometrics

Statistics and Econometric Models

VOLUME ONE

Christian Gourieroux
and Alain Monfort

Translated by Quang Vuong

Statistics and Econometric Models

VOLUME 1

General Concepts, Estimation, Prediction, and Algorithms

Christian Gourieroux
CREST and CEPREMAP, Paris
Alain Monfort
CREST-INSEE, Paris

Translated by Quang Vuong

CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

Originally published in French as
Statistique et modèles économétriques Vol 1 by Économica 1989

First published in English by Cambridge University Press 1995
as *Statistics and econometric models Volume 1*

Printed in Great Britain at the University Press, Cambridge

A catalogue record of this book is available from the British Library

Library of Congress cataloguing in publication data applied for

ISBN 0 521 40551 3 (V. 1) 0 521 47162 1 (V. 2)
ISBN 0 521 47744 1 (V. 1 pb) 0 521 47755 X (V. 2 pb)
ISBN 0 521 47837 5 (two volume paperback set)

ISBN 0 521 40551 3 (hardback)
ISBN 0 521 47744 1 (paperback)

TAG

Preface

The present textbook was motivated by two principles drawn from our teaching experience. First, the application of statistical methods to economic modelling has led to the development of some concepts and tools that are frequently ignored in traditional statistical textbooks. Second, it is possible to introduce these various tools, including the most sophisticated ones, without having recourse to mathematical or probabilistic concepts that are too advanced. Thus the main goal of this book is to introduce these statistical methods by taking into account the specificity of economic modelling and by avoiding a too abstract presentation.

Our first principle has various consequences. Though we discuss problems, concepts, and methods of mathematical statistics in detail, we rely on examples that essentially come from economic situations. Moreover, we analyse the characteristics and contributions of economic methods from a statistical point of view. This leads to some developments in model building but also in concepts and methods. With respect to model building, we address issues such as modelling disequilibrium, agents' expectations, dynamic phenomena, etc. On conceptual grounds, we consider issued such as identification, coherency, exogeneity, causality, simultaneity, latent models, structural forms, reduced forms, residuals, generalized residuals, mixed hypotheses, etc. It is, however, in the field of methods that the economic specificity brings the most important consequences. For instance, this leads to focus on predictions problems, maximum likelihood or pseudo conditional maximum likelihood type estimation methods, M-estimators, moment type estimators, bayesian and recursive methods, specification tests, nonnested tests, model selection criteria, etc.

Our second principle concerns our presentation. In general, we have tried to avoid proofs that are too technical. Instead, we have attempted to emphasize the intuition behind the results, which is a condition necessary to a real understanding. In particular, we have not introduced the concept of σ-algebra and we have systematically left aside issues concerning measurability and negligible sets. We have also tried to strengthen the intuitive understanding of the results by multiplying examples. For the reader interested in technical problems, however, a special chapter (Chapter 24, Volume II) collects rigorous proofs of various asymptotic results. This allows us to lower significantly the mathematical level required for the reading of our book. Lastly, we have included two appendices reviewing basic elements of linear algebra and probability theory. These appendices (A and B, found in Volume II) should provide the reader with a self-contained textbook.

Our textbook can be used in various ways according to the topics covered and levels of difficulty desired. Below, we suggest a three-course sequence. For each course we have defined three sections called "Basic Concepts and Tools," "Estimation and Prediction," and "Tests and Confidence Regions."

Course I

Sections	Chapters
Basic Concepts and Tools	1, 2, 3, and Appendices A, B
Estimation and Prediction	5, 6, 7.1–7.3, 11.1–11.2
Tests and Confidence Regions	14, 15, 16, 20.1–20.5

Course II

Sections	Chapters
Basic Concepts and Tools	4,13.1–13.3
Estimation and Prediction	7.4–7.5, 8.1–8.4, 9, 10, 11.3, 12
Tests and Confidence Regions	17, 18, 20.6–20.8

Course III

Sections	Chapters
Basic Concepts and Tools	13.4–13.6, 22.1, 24
Estimation and Prediction	8.5, 21.1–21.2, 23.1, 23.3
Tests and Confidence Regions	19, 21.3–21.4, 22.2–22.3, 23.2

The first course corresponds to a first-year graduate course in statistical methods for econometrics. The topics covered are indeed the basic ones in statistics. An econometric aspect, however, is provided through the weighting of various topics, the introduction of specific notions, and the choice of examples.

The second course completes the first course by covering asymptotic results. The first two courses should constitute the basic background for a statistician/econometrician wishing to apply the most recent econometric tools.

The third course is the most difficult one on technical grounds. It is also a collection of more advanced topics. A good understanding of the contents of the third level should allow the reading of specialized literature and the beginning of methodological research in good conditions.

We owe a special debt to Martine Germond and Beatrice Lejeune who typed respectively the French and English versions of this book. They both performed promptly and accurately a difficult task. We are also grateful to Quang Vuong for his painstaking translation of our work. Financial support for the publication and translation of this book was provided by the French Ministère de la Recherche et de la Technologie and by the French Ministère de la Culture.

C. Gourieroux
A. Monfort

Contents

CHAPTER 1

Models

1.1 Modelling

1.1.1 Principles

The problem of modelling arises as soon as one wants to describe and analyze a real phenomenon. Because reality is often complex, the human mind is unable to comprehend it in its entirety. Thus it is necessary to construct a simplification of reality (or *model*) allowing one to study it partially. This simplification cannot take into account all the characteristics of reality, but only those that seem related to the object of study and that are of sufficient importance. A model suited to a particular purpose often becomes inadequate when the object of study changes (even if the study concerns the same phenomenon) or when there is a need for greater accuracy.

Example 1.1: To predict the result of an election when there are two candidates A and B, one draws with equal probability and replacement a sample of n voters and observes the number Y of individuals in this sample intending to vote for A. The variable Y can be thought of as a random variable distributed as a binomial $B(n, p_A)$, where p_A is the unknown proportion of individuals voting for A on the day of the election.

In this example, the model is given by the family

$$\{B(n, p_A),\ p_A \in [0, 1]\}.$$

Even if this model seems quite natural here, it is only an approximation of reality for it makes several simplifications:

a) For instance, the sample is drawn with a computer using a program to generate random numbers. Since these programs are based on a deterministic algorithm, one only has an approximation of sampling from a uniform distribution.

b) The proportion of individuals voting for A may change between the date of the survey and the date of the election.

c) Some sampled individuals may not divulge their true voting intentions, etc.

Example 1.2: To evaluate the effect on consumption of a change in income, one may propose a model to describe the relationship between these two variables. For instance, one may suppose that they are related through an equality of the form

$$\log C = a \log R + b, \qquad a, b \in I\!R.$$

The parameter a, called the *consumption elasticity with respect to income*, is equal to the logarithmic derivative $d \log C / d \log R$. If income changes by 1 percent, consumption changes approximately by a percent. The parameter a provides a natural measure of the effect on consumption of a change in income and the model seems appropriate to the study of this effect. The model is clearly an approximation of reality. Indeed time series data (C_t, R_t), $t = 1, \ldots, T$ on consumption and income will not in general be related by an exact equality such as the one proposed.

Example 1.3: A model frequently used to analyze the date of an event is based on the theory of Poisson processes. See Section B.5 in Appendix B, Volume II. The model is more or less suited to the study of unemployment spells (here the event is to find and accept a job). It can be improved and made more realistic by:

a) no longer assuming independence of the past from the present since the probability of an unemployed person finding a job may depend on the length of the unemployment spell,

b) introducing various factors affecting an individual's chances of finding a job. These include general factors such as current economic conditions, and individual factors such as the amount of unemployment benefits received by the individual, whether their spouse is employed, etc.

1.1.2 Stochastic Models

Some of the previous examples have a stochastic character (see Examples 1.1 and 1.3). On the other hand, the model of consumption behavior (1.2) is purely deterministic, which makes it incompatible with the data. One way to solve this difficulty consists in making the model stochastic. Hence the approximate deterministic relation

$$\log C_t = a \log R_t + b, \qquad t = 1, \dots, T,$$

is replaced by

$$\log C_t = a \log R_t + b + u_t, \qquad t = 1, \dots, T,$$

where $u_t, t = 1, \dots, T$, are random variables with zero means called *disturbances* or *error terms.*

A disturbance measures the discrepancy between the observation $\log C$ and the approximate mean $a \log R + b$ proposed by the model. The disturbance may be due to:

(i) the fact that the relation between $\log R$ and $\log C$ is not linear,

(ii) the fact that the coefficients a and b vary over time,

(iii) the omission (voluntary or not) of secondary variables,

(iv) measurement errors on the variables C and R ...

The disturbance is therefore interpreted as a summary of various kinds of ignorance. In fact, the interpretation of the disturbance is important only when hypotheses are made about its distribution. The form of this distribution depends on the kind of ignorance that the disturbance is supposed to represent.

Apart from its interpretation, the introduction of a disturbance has another purpose. As seen later, in a stochastic model it will be possible to construct measures of the error associated with the use of the model. Indeed an evaluation of the accuracy of the results becomes quite essential.

1.1.3 Formulation of a Model

In general, a model has a deterministic component and a random component. How can one obtain an appropriate specification for each of the components? Specifications can rely on observations of the variables or