Probability Theory and Statistical Inference

This major new textbook from a distinguished econometrician is intended for students taking introductory courses in probability theory and statistical inference. No prior knowledge other than a basic familiarity with descriptive statistics is assumed.

The primary objective of this book is to establish the framework for the empirical modeling of observational (non-experimental) data. This framework is formulated with a view to accommodating the peculiarities of observational (as opposed to experimental) data in a unifying and logically coherent way. *Probability Theory and Statistical Inference* differs from traditional textbooks in so far as it emphasizes concepts, ideas, notions, and procedures which are appropriate for modeling observational data. Special emphasis is placed on relating probabilistic concepts to chance regularity patterns exhibited by observed data.

Aimed primarily at students at second-year undergraduate level and above studying econometrics and economics, this textbook will also be useful for students in other disciplines which make extensive use of observational data, including finance, biology, sociology, education, psychology, and climatology.

ARIS SPANOS is a leading researcher and teacher in econometrics. He is currently based at Virginia Polytechnic Institute and State University (Virginia Tech) and has taught previously at institutions including, Birkbeck College, the University of Cambridge, and the University of California. His previous books include *Statistical Foundations of Econometric Modelling*, published by Cambridge University Press in 1986.

Probability Theory and Statistical Inference

Econometric Modeling with Observational Data

Aris Spanos

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521424080

First published 1999
Third printing 2007

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

Spanos, Aris, 1952–
Probability Theory and Statistical Inference: econometric modeling with observational data / Aris Spanos
p. cm.
Includes bibliographical references (p.) and index.
ISBN 0 521 41354 0
1. Econometrics. 2. Probabilities. I.Title.
HB139.S62 1998
330′.01′5195–dc21

ISBN 978-0-521-41354-1 Hardback
ISBN 978-0-521-42408-0 Paperback

Cambridge University Press
978-0-521-42408-0 - Probability Theory and Statistical Inference: Econometric Modeling with Observational Data
Aris Spanos
Frontmatter
More information

Contents

Cambridge University Press
978-0-521-42408-0 - Probability Theory and Statistical Inference: Econometric Modeling with Observational Data
Aris Spanos
Frontmatter
More information

Preface

1 Intended audience and distinguishing features

This is a textbook intended for an introductory course in *probability theory* and *statistical inference*, written for students who have had at least a semester course in calculus. The additional mathematics needed are coalesced into the discussion to make it self-contained, paying particular attention to the intuitive understanding of the mathematical concepts. *No prerequisites in probability and statistical inference are required but some familiarity with descriptive statistics will be of value*.

The primary objective of this book is to lay the foundations and assemble the overarching framework for the empirical modeling of **observational** (non-experimental) data. This framework, known as *probabilistic reduction*, is formulated with a view to accommodating the peculiarities of observational (as opposed to **experimental**) data in a unifying and logically coherent way. It differs from traditional textbooks in so far as it emphasizes concepts, ideas, notions, and procedures which are appropriate for modeling observational data.

The primary intended audience of this book includes interested undergraduate and graduate students of econometrics as well as practicing econometricians who have been trained in the traditional textbook approach. Special consideration has been given to the needs of those using the textbook for self-study. This text can also be used by students of other disciplines, such as biology, sociology, education, psychology, and climatology, where the analysis of observational data is of interest.

The traditional statistical literature over the last 50 years or so, has focused, almost exclusively, on procedures and methods appropriate for the analysis of **experimental-type** (experimental and sample survey) **data**. The traditional modeling framework has been that of the *experimental design* tradition, as molded by Fisher (1935) (and formulated by the sample survey literature of the 1940s and 1950s), and the "curve fitting" perspective of the *least-squares* tradition. Both of these traditions presume a modeling framework in the context of which the observed data are interpreted as a realization of an observable phenomenon which can be realistically viewed as a *nearly isolated* (by divine or human intervention) system; see Spanos (1995b). This book purports to redress (somewhat) the balance by expounding a modeling framework appropriate for observational data. This modeling framework can be viewed in the spirit of the *conditioning* perspective of the Biometric school (Galton, Karl Pearson, Yule) formulated in the late 19th and early 20th century. How conditioning can be used as the cornerstone of a framework for modeling observational data, will be elaborated upon in what follows. The alternative framework is deemed necessary because observational data constitute the rule and not the exception in fields such as econometrics. Indeed, *econometrics* will largely provide both the motivation as well as the empirical examples throughout the discussion that follows.

The most important **distinguishing features** of this text are the following:

(1) The discussion revolves around the central pivot of *empirical modeling* and the order of introduction of the various concepts, ideas, and notions is largely determined by the logical coherence and completeness of the unfolding story. Probability theory and statistical inference are interweaved into empirical modeling by emphasizing the view of the former as a modeling framework; this is in contrast to probability theory as part of mathematics proper or as a rational decision making framework under uncertainty.

(2) Special emphasis is placed on the notion of *conditioning* and related concepts because they provide the key to a modeling framework for observational data. The notion of *regression* arises naturally out of the modeler's attempt to capture *dependence* and *heterogeneity*, when modeling observational data. The discussion does not neglect the importance of the historical development of these ideas and concepts.

(3) The interplay between abstract concepts of probability theory and the corresponding chance regularity patterns exhibited by the data is extensively utilized using a variety of graphical techniques. Special emphasis is placed on mastering the creative skill of "reading" data plots and discerning a number of different chance regularity patterns as well as relating the latter to the corresponding mathematical concepts.

(4) A clear separating line is drawn between *statistical* and (economic) *theory information*, with a statistical model being specified exclusively in terms of statistical information (probabilistic assumptions relating to observable random variables); in contrast to specifying statistical models by attaching autonomous *error terms* (carrying the probabilistic structure) to theoretical relationships.

(5) The discussion emphasizes certain neglected aspects of empirical modeling which are crucially important in the case of observational data. These facets include: *specification* (the choice of a statistical model), *misspecification testing* (assessing the validity of its assumptions), and *respecification;* the discussion puts the accent on the *empirical adequacy* of a statistical model.

(6) A statistical model is viewed not as a set of probabilistic assumptions in the middle of nowhere, but contemplated in the context of the all-embracing perspective of the totality of possible statistical models. This view is adopted in an attempt to systematize the neglected facets of modeling by charting the territory beyond the postulated statistical model in broad terms using an assemblage of restrictions (*reduction assumptions*) on the set of all possible statistical models; hence the term probabilistic reduction.

REMARK: in view of the proposed bridge between probabilistic assumptions and observed data patterns, the neglected facets of modeling are thus transformed into informed procedures and not *hit-or-miss* attempts in connection with the probabilistic structure of some *unobservable* error term(s).

(7) The traditional textbook hypothesis testing is reinterpreted in an attempt to bring out the distinct nature of *misspecification testing*. The proposed interpretation calls into question the traditional view that the Neyman–Pearson approach to testing has (largely) superseded that of Fisher. It is argued that they constitute two distinct (although related) approaches to testing with very different objectives which can be used as complements to one another, not as substitutes.

2 Origins and pedigree*[1]

The present textbook has its roots in the author's book *Statistical Foundations of Econometric Modelling*, published in 1986 by CUP, and has been growing in the form of lecture notes ever since. The *Statistical Foundations* book was my first attempt to put forward an alternative comprehensive methodology to the traditional textbook approach to econometric modeling. This was motivated by the state of econometrics after its failure to fulfill the expectations fomented in the 1960s and 1970s; a failure which led to a re-examination of its foundations and elicited a number of different diagnoses from the critics of the traditional textbook approach in the early 1980s (Hendry, Sims, and Leamer; see Granger (1990)). Naturally, the seeds of the proposed methodology can be traced back to my graduate studies at the London School of Economics (LSE) in the late 1970s; see Hendry (1993).

The primary objective of the *Statistical Foundations* book was to put forward a **logically coherent methodological framework** by *entwining* probability theory and statistical inference into empirical modeling. The *modus operandi* of this modeling approach was the distinction between statistical and theory information and the related recasting of statistical models exclusively in terms of statistical information: probabilistic assumptions relating to the observable random variables underlying the data; the error was reinterpreted as the residuum (unmodeled part). In the context of this framework:

(a) not only the theory, but the nature and the (probabilistic) structure of the observed data is thoughtfully taken into consideration, and
(b) not only estimation and hypothesis testing, but also specification, misspecification testing and respecification, are explicitly recognized as both legitimate and necessary facets of modeling.

The ultimate objective of that book was to propose a framework in the context of which some of the problems that have troubled the traditional textbook approach since

[1] All sections marked with an asterisk (*) can be skipped at first reading without any serious interruption in the flow of the discussion.

the 1920s, including the problems raised by some critics in the 1970s and 1980s, can be resolved. These problems include general methodological issues in modeling such as the nature and structure of theories and the role of observed data, as well as statistical issues such as misspecification testing and respecification, the omitted variables bias, pre-test bias, identification, and exogeneity.

Looking back, the *Statistical Foundations* book was a bold, and some would say audacious, attempt to influence the way econometrics is taught and practiced. Despite its controversial nature and its mathematical level, its success was a pleasant surprise. A purely subjective assessment of its success has been based on:

(a) Its favorable reception in the market place.
(b) Its likely influence on several textbooks in econometrics published after 1986, which took several forms, the most important being:

    (i) A more substantive treatment of probability theory (not an appendage of definitions).
    (ii) A shift of emphasis from unobservable error terms to observable random variables.
    (iii) A historically more accurate interpretation of the regression model (as opposed to the Gauss linear model), proposed as being better suited for the analysis of observational (non-experimental) data.
    (iv) The introduction of the notion of a *statistical generating mechanism* (GM) as an orthogonal decomposition of a random variable (or vector), given an information set; as opposed to a functional relationship among theoretical variables with an error attached.
    (v) A more systematic treatment of misspecification testing and respecification.

    In addition to some explicitly acknowledged influences (see *inter alia* Cuthbertson *et al.* (1992), Mills (1993), Davidson and MacKinnon (1993), Hendry (1995), Poirier (1995)), I would like to think that there has also been some indirect influence in relation to:

    (vi) Heightening the level and broadening the role of probability theory in econometrics (see Dhrymes (1989), Goldberger (1991), Amemiya (1994), Davidson (1994)).
    (vii) Helping to focus attention on the issue of misspecification testing (see Godfrey (1988)).

(c) Its influence on current practice in empirical modeling with regard to misspecification testing. Both, textbooks as well as econometric computer packages, nowadays, take a more serious view of misspecification testing; regrettably misspecification testing for systems of equations (see Spanos (1986) chapter 24) is yet to be implemented.[2] This should be contrasted with the pre-1980s treatment of misspecification testing and respecification, which amounted to little more than looking at the

[2] PcGive (see Doornik and Hendry (1997)) is the exception.

Durbin-Watson statistic and when the null was rejected, palliating the problem by modeling the error.

The *Statistical Foundations* book, however, was a work in progress with a number of crucial limitations arising from externally imposed restrictions (mainly the prohibitive cost of typesetting graphs at the time) as well as some initial hesitation on my behalf to focus my research exclusively on such a long term project. The advances in personal computing technology and the success of the *Statistical Foundations* book provided the spur to take the story a step further, unimpeded by the initial constraints.

This textbook represents a more ripened elucidation of the approach proposed in the *Statistical Foundations* book with particular attention paid to the exposition of concepts and ideas, and the logical coherence, consistency and completeness of the approach. The primary objective of the *Statistical Foundations* book is pursued further in a number of different ways:

(1) Ameliorating the interweaving of probability theory and statistical inference into empirical modeling by presenting statistical models as constructs based on three basic forms of (statistical) information:

(D) Distribution, (M) Dependence, (H) Heterogeneity.

This information can be related to observed data (chance regularity) patterns using a variety of graphical techniques, rendering modeling an informed procedure.

(2) Strengthening the logical coherence of the proposed methodology by stressing the distinction between theoretical and statistical information in terms of which the respective *theory* and *statistical models* are defined. *Statistical information* is codified exclusively in terms of probabilistic concepts and *theory information* in terms of economic agents' behavior.

(3) Enhancing the probabilistic reduction framework by viewing statistical models in the context of a broader framework where the neglected facets of empirical modeling (specification, misspecification, respecification), can be implemented in a more systematic and informed fashion, in conjunction with a variety of graphical techniques.

(4) Enriching the probabilistic reduction framework by extending empirical modeling beyond the Normal/linear/homoskedastic territory in all three dimensions of statistical information; the emphasis in the *Statistical Foundations* book was placed on modeling *dependence*, with only brief remarks in relation to exploring the other two dimensions of modeling; *distribution* (beyond Normality) and *heterogeneity*.

The ultimate objective of this textbook remains the same as that of the *Statistical Foundations*. In the context of the proposed methodology the nature and structure of theories as well as the role of the data in the assessment of theories is addressed without shying away from any difficult methodological issues as they arise naturally during the discussion. In the context of the proposed framework a purely statistical viewing angle (as opposed to the traditional theory viewing angle) is put forward, in an attempt to elucidate some statistical issues, such as the *omitted variables*, *pre-test bias*, *multicollinearity*, and *exogeneity*, which are often (misleadingly) viewed from the theoretical viewing angle creating confusion; see Spanos (1997a).

3 As a teaching and learning device

Empirical modeling is a difficult subject to master primarily because it inherits the sum (and then some) of the inherent difficulties in learning probability theory and statistical inference. Some of these inherent difficulties are:

(a) The requirement to master too many concepts, ideas, and notions, too quickly.
(b) The abstract nature of these concepts without any apparent connection to tangible counterparts.
(c) The problem of terminology in so far as the (historically) established terminology is often inept as well as misleading because the same term is often used to denote several different notions.
(d) The extensive utilization of mathematical symbols.

In a purposeful attempt to deal with (a), the most troublesome difficulty, the discussion that follows utilizes several *learning techniques* such as:

(i) *repetition* of crucial concepts and notions,
(ii) *story so far* abridgements,
(iii) *bird's eye view of the chapter* outlines,
(iv) *looking ahead* summaries, and
(v) recurrent references to *what is important*.

In addition, the discussion utilizes *short historical excursions* in order to dilute the high concentration of concepts, ideas and notions.

The extensive utilization of graphical techniques enables the reader to build direct connections between abstract probability concepts, such as Normality, Independence and Identical Distribution, and the corresponding chance regularity patterns, rendering the task of mastering these concepts easier.

The problem of inept terminology is brought out at the appropriate places during the discussion and the reader is warned about possible confusions. For example, the term *mean* is used in the traditional literature to denote at least four different (but historically related) notions:

A Probability theory
  (i) the mean of a random variable X: $E(X)=\int_{x\in\mathbb{R}_X} xf(x)dx$,

B Statistical inference
  (ii) the sample mean: $\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$,
  (iii) the value of the sample mean: $\overline{x}=\frac{1}{n}\sum_{k=1}^{n}x_k$,

C Descriptive statistics
  (iv) the arithmetic average of a data set $(x_1, x_2, \ldots, x_n)$: $\overline{x}=\frac{1}{n}\sum_{k=1}^{n}x_k$.

The teacher is obliged to spend a sizeable part of his/her limited time explaining to the students firstly, the (often subtle) differences between the different notions represented by the same term and secondly, how these different notions constitute relics of the historical development of the subject. In the case of the mean, up until the 1920s (see Fisher (1922)), the whole literature on statistics conflated probabilities and relative frequencies, leading to numerous befuddlements including that between $E(X)$, $\overline{X}$ and $\overline{x}$. What is more, very few statisticians of the 1920s and 1930s noticed the transition from descriptive statistics to statistical inference proper; Karl Pearson died in 1936 without realizing that Fisher had turned the tables on him; see chapters 11–13. In an attempt to ameliorate the inept terminology problem, new terms are introduced whenever possible. This is often impossible, however, because the terminology has been entrenched over several decades.

Students with a somewhat weaker background in mathematics are often intimidated by *mathematical terminology* and *symbolism*. On the issue of mathematical symbols every effort has been made to use uniform symbolism throughout this book. The student should know, however, that systematic thinking is made more accessible by the utilization of heedful symbolism. Symbols are essential for systematic thinking because they economize on unnecessary and nebulous descriptions! Hence, the proper utilization of symbols is considered a blessing not a hindrance. Unfortunately, undergraduates often confuse symbolism and haughty-sounding terminology with mathematical sophistication and develop a phobia to any hint of the Greek alphabet; fraternity parties not withstanding. The only known cure for such phobias is the reverse cold turkey treatment: face your phobia head on until it becomes a friend not a foe.

Another important feature of the book is that a lot of emphasis is placed on providing the reader with an opportunity to learn things properly, as opposed to pretend learning, so eloquently articulated by Binmore (1992, p. xxvi):

> Much of what passes for an undergraduate education, both in the United States and in Europe, seems to me little more than an unwitting conspiracy between the teacher and the student to defraud whoever is paying the fees. The teacher pretends to teach, and the student pretends to learn, material that both know in their hearts is so emasculated that it cannot be properly understood in the form in which it is presented. Even the weaker students grow tired of such a diet of predigested pap. They understand perfectly well that 'appreciating the concepts' is getting them nowhere except nearer to a piece of paper that entitles them to write letters after their names. But most students want more than this. They want to learn things properly so that they are in a position to feel that they can defend what they have been taught without having to resort to the authority of their teachers or the textbook. Of course, learning things properly can be hard work. But my experience is that students seldom protest at being worked hard provided that their program of study is organized so that they quickly see that their efforts are producing tangible dividends.

I have instinctively subscribed to the creed of *learning things properly or don't even bother* since I can remember. As a student I refused to learn things by heart because I knew *in my heart* that it was (largely) a waste of time and intellectual energy; I even caught myself refusing to answer exam questions which were designed to test my ability to memorize in a parrot-like fashion. As a teacher of econometrics for over 18 years both in Europe and the United States, I have tried very hard to implement the creed of learning things properly and **think, things, through, systematically** with encouraging success.

After a few years of teaching in the United States, however, I found myself on the slippery declivity of unconsciously emasculating probability theory and statistical inference down to maize-porridge. After the first two years, I reluctantly reached the conclusion that fighting the system would be tantamount to jousting with windmills but at the same time I could not indulge in pretend teaching. The only strategy with a fighting chance to circumvent this problem seemed to be to *start* with a diet of predigested pap and gradually work my way up to something more solid. The reality was that the digestive system of the majority of students, after years on such a diet, would not accept anything but predigested pap. It goes without saying, that I consider the students as being the *real victims* of the "pretend teaching conspiracy" and in no circumstance share in the blame for this state of affairs. The educational system itself puts the overwhelming majority of the students on this diet from a very early age, encouraging them to think that the ultimate objective of learning is the exam which leads to the piece of paper and what that entitles them to; learning for exams fosters pretend learning. In view of this situation, I resigned myself to the idea of trying to reach everybody along the lines of learning things properly, but also to satisfy myself with galvanizing a minority of students who would be prepared to think along during the lectures. Inevitably, the predigested pap component of the course tended to expand slowly but surely. Fortunately, I stumbled across the above passage from Binmore just in time to salvage myself from sliding down this slippery declivity even further. The undergraduates of a different educational system, at the University of Cyprus, helped restore my faith in learning things properly and the somewhat emasculated lecture notes of the period 1989–1992 have been transformed into the current textbook after five years of teaching undergraduate econometrics.

I sympathize with the teachers of econometrics who find themselves in a situation where they have to teach something their students do not *really* want to learn and are forced to indulge in pretend teaching, but I offer no apologies for the level of difficulty and the choice of material in the current textbook. I know very well that this book could never qualify to be included in any of the popular series with titles such as "Econometrics for dummies," "An idiot's guide to econometrics" and "A *complete* idiot's guide to econometrics". Personally, I could not water-down the material any further without emasculating it; I leave that to better teachers.

This is a textbook designed for interested students who want to learn things properly and are willing to put in the necessary effort to master *a systematic way of thinking*. With that in mind, no effort has been spared in trying to explain the crucial concepts and ideas at both an intuitive as well as a more formal level. Special attention is given to Binmore's advice that "the program of study is organized so that they [the students] quickly see that their efforts are producing tangible dividends." Early in the discussion of probability theory, the book introduces a variety of graphical techniques (see chapters 5–6) in an attempt to enable the reader to relate the abstract probability concepts to discernible observed data patterns and thus develop an intuitive understanding of these concepts. Moreover, *the story so far* abridgements and *looking ahead* summaries in almost every chapter are designed to give the student a bird's eye view of the forest and encourage learning by presenting the material as an unfolding story with a single unifying plot. At the same time these abridgements and summaries can be used by the readers to assess their understanding of the crucial steps and the important concepts in the discussion.

Another dimension of learning things properly is related to the way the material is presented. I consider university teaching not as an attempt to force some predigested pap down students' throats but as an attempt to initiate the students into thinking, things, through, systematically in the context of a coherent framework; the teacher offers both the systematic thinking and the coherent framework. In class I try to induce the students to think along with me in a conscious attempt to coach them in a particular way of thinking. Needless to say that the present book largely reflects this attitude to teaching and learning and I have no delusions with regard to its suitability for students who prefer the force-feeding and pretend teaching and learning methods. The book does not subscribe to pretend teaching even when discussing some of the most difficult concepts and ideas in probability theory and statistical inference, such as *σ-fields*, *stochastic conditioning*, *limit theorems*, *the functional limit theorem*, *stochastic processes*, *maximum likelihood*, and *testing*. Instead, when the material is deemed necessary for a coherent and proper understanding of empirical modeling, every effort is made to **demystify** the concepts and ideas involved by ensuring that the discussion is both comprehensive and systematic. Special emphasis is placed on motivating the need for these concepts and on the intuitive understanding of their essence.

To those teachers who are not convinced that this material can be taught to interested undergraduates, I can only offer my experience at the University of Cyprus for a period of five years. The overwhelming majority of undergraduates in economics could assimilate the bulk of the material in a two semester course and over a third of these students, having toiled through this textbook, would then *choose* to specialize in econometrics and proceed to struggle through Spanos (1986, forthcoming). Moreover, when I returned to the United States for a semester, I was able to use the book in undergraduate courses by utilizing a number of shortcuts which affect only the depth of the discussion but not its coherence.

The creed of learning things properly entails a logically coherent, complete (as far as possible) and in depth discussion which goes beyond skimming the surface of the subject. Compromising the logical coherence and the completeness of the discussion will often be counterproductive. The book, however, is written in such a way so as to provide the teacher and student with options to decide the depth of the analysis by taking shortcuts when the going gets tough. As a general rule, all sections marked with an asterisk (*) can be skipped at first reading without any serious interruption in the flow of the discussion. This textbook includes enough material for two semester courses, but with some judicious choices the material can be shortened to design several one semester courses. Using my experience as a guide, I make several suggestions for one semester courses at different levels of difficulty.

1 **One semester undergraduate course in probability theory and statistical inference**: **Chapter 1**(1.1–1.5), **chapter 2**(2.1–2.2), **chapter 3**(3.4.2–3.7.1, 3.7.4), **chapter 4**(4.1–4.2.3, 4.3–4.4.4, 4.5.7, 4.7.1, 4.7.4), **chapter 5**(5.1–5.6), **chapter 6**(6.1–6.2, 6.4.1–6.4.2, 6.7.2), **chapter 7**(7.1–7.2), **chapter 8**(8.1–8.3, 8.10.1–8.10.3), **chapter 9**(9.1–9.2, 9.3.1–9.3.3, 9.4.1, 9.5.1, 9.9), **chapter 11**(11.1–11.2, 11.5–11.7.1), **chapter 12**(12.1–12.4), **chapter 13**(13.1–13.3), **chapter 14**(14.1–14.3).

2 **One semester undergraduate (intermediate) course in probability theory**: **Chapter 1**(1.1–1.6), **chapter 2**(2.1–2.9), **chapter 3**(3.1–3.7), **chapter 4**(4.1–4.7), **chapter 5**(5.1–5.6), **chapter 6**(6.1–6.7), **chapter 7**(7.1–7.2,), **chapter 8**(8.1–8.5, 8.9–10), **chapter 9**(9.1–9.5).

3 **One semester undergraduate (intermediate) course in statistical inference**: **Chapter 1**(1.1–1.6), **chapter 5**(5.1–5.6), **chapter 7**(7.1–7.6), **chapter 9**(9.1–9.5), **chapter 11**(11.1–11.6), **chapter 12**(12.1–12.5), **chapter 13**(13.1–13.5), **chapter 14**(14.1–14.6), **chapter 15**(15.1–15.5).

4 **One semester graduate course in probability theory and statistical inference**: **Chapter 1**(1.1–1.6), **chapter 2**(2.1–2.9), **chapter 3**(3.1–3.7), **chapter 4**(4.1–4.7), **chapter 5**(5.1–5.6), **chapter 6**(6.1–6.7), **chapter 7**(7.1–7.2), **chapter 8**(8.1–8.5, 8.9–8.10), **chapter 9**(9.1–9.5), **chapter 11**(11.1–11.6), **chapter 12**(12.1–12.5), **chapter 13**(13.1–13.4), **chapter 14**(14.1–14.6), **chapter 15**(15.1–15.5).

5 **One semester graduate course in probability theory:** **Chapters 1–9**; all sections marked with an asterisk (*) are optional.

6 **One semester graduate course in statistical inference**: **Chapters 1, 5, 7, 9, 10–15**; all sections marked with an asterisk (*) are optional.

In view of the fact that the book is written with a variety of audiences in mind, it should be apparent that another learning attitude that I do not subscribe to is that textbooks should include only the material everybody is supposed to toil through and any additional material can only confuse the helpless student. It is true that when a student wants to learn things by heart, additional material can only blur the part to be memorized, rendering such a task more difficult, but when the aim is to learn things properly, no such danger exists; seeing more of the jigsaw puzzle can only illuminate the part the student is trying to master. Indeed, some care has been taken to cater for the inquisitive student who wants to learn things properly and pursue certain lines of thought further.

4 Looking ahead: a bird's eye view

The text begins with an introductory chapter which demarcates the intended scope of empirical modeling and sets the scene for what is to follow by elaborating on the distinguishing features of modeling with observational data and summarizing the main methodological *prejudices* of the present author. The primary objective of the chapters that follow is to transform these prejudices into *theses*: contentions supported by coherent (and hopefully persuasive) arguments.

Probability theory

The journey through the probability theory forest, which commences in chapter 2 and stretches all the way to chapter 10, will take the reader along a specific pathway we call the **modeling trail**. This pathway has been designed so that the hiker reaches the other side (statistical inference) in the most direct way possible, avoiding, wherever possible, treacherous terrain and muddy valleys (however interesting) where the visibility is limited and the traveler is likely to get bogged down. The construction of the modeling trail sometimes led the author into uncharted territory which has been cleared of the undergrowth for that purpose. The pathway has been drawn mostly along mountain ridges in order to enable the hiker to:

(a) have a broader view of the forest and
(b) to catch periodic glimpses of the other side (data analysis), not to get disheartened along the trail.

The choice of concepts and ideas and the order in which they are introduced at different points of the trail are determined by the requirement to provide a coherent account of the modeling aspect of the problem, and might seem unorthodox when compared with the traditional treatment of these topics. Concepts and ideas are not introduced as required by strict adherence to the mathematical principles of consistency and completeness, but in the order needed for modeling purposes with emphasis placed not on mathematical rigor and sleekness, but on the intuitive understanding and the *coherence* of the discussion along this modeling trail.

The key to reaching the other side in a creative mood (mastering the material) is to keep one eye on the forest (the unfolding story along the modeling trail) and avoid getting distracted by the beauty (or ugliness!) of the trees along the chosen path or venturing into dense areas of the forest where visibility is limited. Moreover, any attempt to apply superficial memorization of the concepts and ideas along the pathway is doomed to failure. Anybody inclined to use learning by heart (as opposed to proper learning) is cautioned that the sheer number of different notions, concepts and ideas renders this a hopeless task.

In an attempt to help the reader make it to the other side, the discussion, as mentioned above, utilizes several *learning techniques* including:

(a) regular stops along the way at key points with exceptional visibility, and
(b) short historical excursions.

The regular stops are designed to give the hiker:

(i) an opportunity to take a breather,
(ii) a break to reflect and take stock of what has been encountered along the way and
(iii) a chance to look back and master concepts missed during the first passing.

Such stops are signposted from afar and the reader is advised to take advantage of them; sometimes even retrace his/her steps in order to have a more coherent picture of the view from the modeling trail. The trail is also interspersed with short historical excursions designed to:

(1) offer a better perspective for certain crucial concepts and ideas,
(2) provide some initial coherence to the view from certain vantage points and, most importantly,
(3) present the forest as a living organism which grows and changes ceaselessly and not as an artificially-put-together amusement (or torture!) park.

The discussion in chapters 2–4 has one primary objective: to motivate probability theory as a modeling framework by introducing the idea of *a simple statistical model* as the mathematization of the simplest form of a stochastic phenomenon we call a random experiment. In chapter 5, we relate the basic probabilistic concepts defining a simple statistical model with observed data patterns using a variety of graphical techniques. This builds a bridge linking certain key abstract probability concepts with the reality we call observational data. In chapters 6–7, the discussion purports to show that more realistic statistical models, which can account for both dependence and heterogeneity, can be viewed as natural extensions of the simple statistical model. The key concept in this extension is that of *conditioning* which leads naturally to regression models viewed as models based on conditional distributions and the associated moments. The formal concept which is needed to make the transition from a simple statistical to a regression-type model is that of a stochastic process which is discussed in some detail in chapter 8. In anticipation of several important results in statistical inference, chapter 9 discusses limit theorems. Chapter 10 constitutes the bridge between probability theory and statistical inference and purports to solidify the notion of chance regularity introduced in chapters 1–2.

Statistical inference

Having made it to the other side (chapter 10) the hiker will soon realize that the worst is over! After an introductory chapter (see chapter 10), which purports to provide a more solid bridge between probability theory and statistical inference (and can be avoided at first reading), we proceed to provide an elementary overview of statistical inference in chapter 11. The discussion on estimation commences with the optimal properties of estimators which revolve around the notion of the *ideal estimator* (see chapter 12). In chapter 13 we proceed to discuss several estimation methods: the moment matching principle, the method of moments (Pearson's and the parametric versions), the least-squares and the maximum likelihood methods. A notable departure from the traditional treatment is the calling into question of the conventional wisdom on comparing the method of moments and the maximum likelihood method in terms of the efficiency of the resulting estimators. In a nutshell the argument is that the former method has little (if anything) to do with what Karl Pearson had in mind and any comparison constitutes an anachronism. This is because the nature of statistics has been changed drastically by Fisher but Pearson died before he had a chance to realize that a change has taken place. Using a crude analogy, Pearson was playing checkers but Fisher changed the game to chess and demonstrated that the former's strategies [designed for checkers] do not work as well [when playing chess] as his own strategies [designed for playing chess], leaving out the details in square brackets! The discussion on testing (see chapter 13) differs greatly from traditional treatments in several important respects, the most important of which is the comparison and the contrasting of two alternative paradigms, Fisher's pure significance testing and Neyman–Pearson's optimal testing. We call into question the conventional wisdom that the Fisher approach has been largely superseded by the Neyman–Pearson approach. In chapter 14, it is argued that the two approaches have very different objectives and are largely complementary. Fisher's approach is better suited for *testing without* and the Neyman–Pearson approach is more appropriate for *testing within* the boundaries of the postulated statistical model. In chapter 15 we discuss the problem of misspecification testing and argue that the Fisher approach is the procedure of choice with the Neyman–Pearson approach requiring certain crucial modifications to be used for such a purpose. When testing theoretical restrictions within a statistically adequate model, however, the Neyman–Pearson approach becomes the procedure of choice.

Enjoy the journey!

Acknowledgments

In a long-term project such as the one on which this textbook is based, it is only natural that one accumulates a huge debt to numerous colleagues and students. Foremost on this list is the intellectual debt I owe to my teachers at the London School of Economics in the late 1970s; grateful thanks go to David Hendry, Denis Sargan, Alan Stuart, Jim Durbin, Grayham Mizon, and Ken Binmore. In the same group I should include Jean-François Richard who brought a crucial dimension to the LSE tradition during my graduate studies (see Hendry (1993)). Indeed, thinking back, his (1980) paper was instrumental in setting me off in this great adventure!

I'm also greatly indebted to several colleagues, graduate students and research assistants who actively participated in the exploring of this uncharted territory. Special thanks go to my colleagues Anya McGuirk and John Robertson whose help in writing most of the computer programs behind the empirical analysis and the graphical techniques in this book, and their insightful comments during our long coffee sessions in my office at Virginia Tech, I gratefully acknowledge. Proof-reading, the most tedious stage in writing a book, became tolerable thanks to the invaluable help from Anya McGuirk. Many thanks also go to Elena Andreou and Christina Christou for their help in increasing my productivity over the last few years.

Several colleagues and students contributed significantly to the improvement of the discussion in this book by making detailed comments and suggestions on several chapters and sometimes several versions of the same chapter. I take pleasure in thanking my colleagues Louis Christofides, Yannis Ioannides, Julio Lopez, Nikitas Pittis, Djavad Salehi-Isfahani, and my students Elena Andreou, Andros Kourtellos, Maria Loizou, and Mohammad Arzaghi. I would like to single out Thanasis Stengos for a special thanks because his insightful comments and encouragement helped to shape several chapters of this text.

Over the years, many colleagues and students have been enthusiastic sounding-boards, helpful critics and discussion partners . They are of course too numerous to list, but certain ones truly deserve special mention. In particular, I would like to thank Rick Ashley, Paul Driscoll, Anna Hardman, Essie Maasoumi, Michalis Magdalinos, and Panos Pashardes.

I would also like to thank several generations of students who toiled through earlier versions of this text during the period 1989–1997, both at Virginia Tech and the University of Cyprus and whose questions helped to clarify several aspects of the approach. Many thanks are also due to the Economics Editor of CUP Ashwin Rattan and his creative team, Elinor Cole (Design), Jayne Aldhouse (Production) and Robert George (Marketing), for doing such a great job.

It is inevitable that when one embarks on a long-term project whose primary objective is to put forward an alternative to the traditional approach, the road would often be arduous and the going would get very toilsome and solitary at times. During periods of forlorn hope a pat on the back or a vote of confidence from colleagues whose views one values, is of paramount importance in keeping one's enthusiasm going. It gives me great pleasure to thank David Hendry and Yannis Ioannides whose encouragement kept me going during times that I felt like throwing in the towel. I would also like to record my thanks to Peter Phillips and Essie Maasoumi whose open-minded editorship of *Econometric Theory* and *Econometric Reviews*, respectively, gives dissenting views a chance to be heard.

My sincerest apologies to any colleagues and students whose help I have inadvertently failed to acknowledge. I hope to have an opportunity to correct the omission in a second edition.

Finally, I owe a great debt to my wife Evie and daughters Stella, Marina and Alexia for their patience and support during the long period spent on writing this book.

Cambridge University Press
978-0-521-42408-0 - Probability Theory and Statistical Inference: Econometric Modeling with Observational Data
Aris Spanos
Frontmatter
More information

Symbols

$\mathbb{T}$	index set; often $\mathbb{T}=\{1, 2, \ldots, \mathbb{T}, ..\}$
$\mathbb{Z}$	set of integers
$\mathbb{N}$	set of natural numbers
$\mathbb{Q}$	set of rational numbers
$\mathbb{R}$	the set of real numbers; real line $(-\infty, \infty)$
$\mathbb{R}^n$	$\overbrace{\mathbb{R}\times\mathbb{R}\times\cdots\times\mathbb{R}}^{n \text{ times}}$
$\mathbb{R}_+$	the set of positive real numbers; real line $(0, \infty)$
$\mathcal{B}(\mathbb{R})$	Borel $\sigma-$ field generated by $\mathbb{R}$
α_3	skewness coefficient
α_4	kurtosis coefficient
$f(x; \boldsymbol{\theta})$	density function of X with parameters $\boldsymbol{\theta}$
$F(x; \boldsymbol{\theta})$	cumulative distribution function
μ_r	central moment of order r
μ_r'	raw moment of order r
$\mathsf{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
$\mathsf{U}(a, b)$	Uniform distribution over the interval $[a, b]$
$\mathsf{F}(m, n)$	F distribution with parameters m and n
$\chi^2(n)$	chi-square distribution with n degrees of freedom
$\mathsf{St}(\nu)$	Student's t distribution with ν degrees of freedom
S	outcomes set
$\mathfrak{F}$	event space (a σ-field)
$\mathbb{P}(.)$	probability set function
ε	random experiment
$\sigma(X)$	minimal sigma-field generated by X
$\sim\mathsf{D}(.)$	distributed as $\mathsf{D}(.)$
$\overset{H}{\sim}$	distributed under H
$\sim\mathsf{D}(.)$	asymptotically distributed as $\mathsf{D}(.)$
$\overset{\alpha}{\propto}$	proportional to

$E(.)$	expected value
$Var(X)$	variance of X
$Corr(X, Y)$	correlation between X and Y
$Cov(X, Y)$	covariance between X and Y
$\stackrel{a.s}{\to}$	almost sure convergence
$\stackrel{c}{\to}$	complete convergence
$\stackrel{\mathcal{D}}{\to}$	convergence in distribution
$\stackrel{\mathbb{P}}{\to}$	convergence in probability
$\stackrel{r}{\to}$	convergence in rth moment
$\mathbf{I}_T(\boldsymbol{\theta})$	Fisher's information matrix for T observations
$\mathbf{I}(\boldsymbol{\theta})$	Fisher's information matrix for 1 observation
$\mathbf{I}_\infty(\boldsymbol{\theta})$	Fisher's asymptotic information matrix
$\xi_n(\boldsymbol{\theta}; \mathrm{X})$	conditional information process
$:=$	definitional identity

Acronyms

AR(p)	Autoregressive model of order *p*
ARMA(p,q)	Autoregressive-Moving Average model of order *p*,*q*
CAN	Consistent, Asymptotically Normal
cdf	cumulative distribution function
CLT	Central Limit Theorem
DGP	Data Generating Process
ecdf	empirical cumulative distribution function
FCLT	Functional Central Limit Theorem
GM	Generating Mechanism
ID	Identically Distributed
IID	Independent and Identically Distributed
LIL	Law of Iterated Logarithm
LS	Least-Squares
mgf	moment generating function
MLE	Maximum Likelihood Estimator
MSE	Mean Square Error
NIID	Normal, Independent and Identically Distributed
PMM	Parametric Method of Moments
PR	Probabilistic Reduction
SLLN	Strong Law Large Numbers
UMP	Uniformly Most Powerful
WLLN	Weak Law Large Numbers