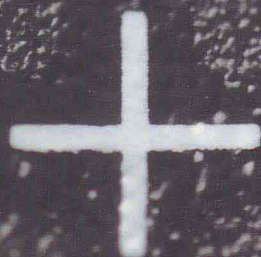


GARY KOOP

ANALYSIS OF ECONOMIC DATA

SECOND EDITION



ANALYSIS OF ECONOMIC DATA SECOND EDITION

by

Gary Koop

University of Leicester



John Wiley & Sons, Ltd

Copyright © 2005 John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England
Telephone: (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page www.wileyeurope.com or www.wiley.com

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770620.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop # 02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Koop, Gary.

Analysis of economic data / by Gary Koop. – 2nd ed.

p. cm.

Includes bibliographical references and index.

ISBN 0-470-02468-2

1. Econometrics. I. Title.

HB141.K644 2004

300'.01'5195-dc22

2004013848

British Library Cataloguing in Publication Data

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

ISBN 0-470-02468-2

Typeset in 11 on 13 pt Garamond Monotype by SNP Best-set Typesetter Ltd., Hong Kong

Printed and bound in Great Britain by T.J. International Ltd, Padstow, Cornwall

This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

ANALYSIS OF ECONOMIC DATA

To Lise

Preface to the first edition

This book aims to teach econometrics to students whose primary interest is not in econometrics. These are the students who simply want to apply econometric techniques sensibly in the context of real-world empirical problems. This book is aimed largely at undergraduates, for whom it can serve either as a stand-alone course in applied data analysis or as an accessible alternative to standard econometric textbooks. However, students in graduate economics and MBA programs requiring a crash-course in the basics of practical econometrics will also benefit from the simplicity of the book and its intuitive bent.

This book grew out of a course I taught at the University of Edinburgh entitled “Analysis of Economic Data”. Before this course was created, all students were required to take a course in probability and statistics in their first or second year. Students specializing in economics were also required to take an econometrics course in their third or fourth year. However, non-specialist students (e.g. Economics and Politics or Economics and Business students) were not required to take econometrics, with the consequence that they entered their senior undergraduate years, and eventually the job market, with only a basic course in probability and statistics. These students were often ill-prepared to analyze sensibly real economic data. Since this is a key skill for undergraduate projects and dissertations, for graduate school, as well as for most careers open to economists, it was felt that a new course was needed to provide a firm practical foundation in the tools of economic data analysis. There was a general consensus in the department that the following principles should be adhered to in designing the new course:

1. It must cover most of the tools and models used in modern econometric research (e.g. correlation, regression and extensions for time series methods).
2. It must be largely non-mathematical, relying on verbal and graphical intuition.

3. It must contain extensive use of real data examples and involve students in hands-on computer work.
4. It must be short. After all, students, especially those in joint degrees (e.g. Economics and Business or Economics and Politics) must master a wide range of material. Students rarely have the time or the inclination to study econometrics in depth.

This book follows these basic principles. It aims to teach students reasonably sophisticated econometric tools, using simple non-mathematical intuition and practical examples. Its unifying themes are the related concepts of regression and correlation. These simple concepts are relatively easy to motivate using verbal and graphical intuition and underlie many of the sophisticated models and techniques (e.g. cointegration and unit roots) in economic research today. If a student understands the concepts of correlation and regression well, then he/she can understand and apply the techniques used in advanced econometrics and statistics.

This book has been designed for use in conjunction with a computer. I am convinced that practical hands-on computer experience, supplemented by formal lectures, is the best way for students to learn practical data analysis skills. Extensive problem sets are accompanied by different data sets in order to encourage students to work as much as possible with real-world data. Every theoretical point in the book is illustrated with practical economic examples that the student can replicate and extend using the computer. It is my strong belief that every hour a student spends in front of the computer is worth several hours spent in a lecture.

This book has been designed to be accessible to a variety of students, and thus, contains minimal mathematical content. Aside from some supplementary material in appendices, it assumes no mathematics beyond the pre-university level. For students unfamiliar with these basics (e.g. the equation of a straight line, the summation operator, logarithms), a wide variety of books are available that provide sufficient background.

I would like to thank my students and colleagues at the University of Edinburgh for their comments and reactions to the lectures that formed the foundation of this book. Many reviewers also offered numerous helpful comments. Most of these were anonymous, but Denise Young, Craig Heinicke, John Hutton, Kai Li and Jean Soper offered numerous invaluable suggestions that were incorporated in the book. I am grateful, in particular, to Steve Hardman at John Wiley for his enthusiasm and the expert editorial advice he gave throughout this project. I would also like to express my deepest gratitude to my wife, Lise, for the support and encouragement she provided while this book was in preparation.

Preface to the second edition

When writing the new edition of my book, I tried to take into account the comments of many colleagues who used the first edition and the reviewers (some anonymous) whom Wiley persuaded to evaluate my proposal for a new edition as well as my personal experience. With regards to the last, I have used the first edition of the book at three different universities (Edinburgh, Glasgow and Leicester) at three different levels. I have used it for a third-year course (for students who were not specialist economists and had little or no background in statistics), for a second-year course (for students with a fair amount of economics training, but little or no training in statistics) and for a first-year course (for students facing economic data analysis for the first time). Based on student performance and feedback, the book can successfully be used at all these levels. Colleagues have told me that the book has also been used successfully with business students and MBAs.

The second edition has not deleted anything from the first edition (other than minor corrections or typos and editorial changes). However, substantial new material has been added. Some of this is to provide details of the (minimal) mathematical background required for the book. Some of this provides more explanation of key concepts such as index numbers. And some provides more description of data sources. Throughout, I have tried to improve the explanation so that the concepts of economic data analysis can be easily understood. In light of the book's use in business courses, I have also added a bit more material relevant for business students, especially those studying finance.

I still believe in all the comments I made in the preface for the first edition, especially those expressing gratitude to all the people who have helped me by offering perceptive comments. To the list of people I thank in that preface, I would like to add the names of Julia Darby, Kristian Skrede Gleditsch and Hilary Lamaison, and all my students from the Universities of Edinburgh, Glasgow and Leicester.

Contents

Preface to the first edition	xi
Preface to the second edition	xiii
Chapter 1 Introduction	1
Organization of the book	3
Useful background	4
Appendix 1.1: Mathematical concepts used in this book	4
Endnote	7
Chapter 2 Basic data handling	9
Types of economic data	9
Obtaining data	14
Working with data: graphical methods	18
Working with data: descriptive statistics	23
Chapter summary	25
Appendix 2.1: Index numbers	25
Appendix 2.2: Advanced descriptive statistics	31
Endnotes	32
Chapter 3 Correlation	35
Understanding correlation	35
Understanding correlation through verbal reasoning	36
Understanding why variables are correlated	39
Understanding correlation through XY -plots	42
Correlation between several variables	45

	Chapter summary	46
	Appendix 3.1: Mathematical details	46
	Endnotes	47
Chapter 4	An introduction to simple regression	49
	Regression as a best fitting line	50
	Interpreting OLS estimates	54
	Fitted values and R^2 : measuring the fit of a regression model	57
	Nonlinearity in regression	61
	Chapter summary	65
	Appendix 4.1: Mathematical details	65
	Endnotes	67
Chapter 5	Statistical aspects of regression	69
	Which factors affect the accuracy of the estimate β ?	70
	Calculating a confidence interval for β	73
	Testing whether $\beta = 0$	79
	Hypothesis testing involving R^2 : the F -statistic	84
	Chapter summary	86
	Appendix 5.1: Using statistical tables for testing whether $\beta = 0$	87
	Endnotes	88
Chapter 6	Multiple regression	91
	Regression as a best fitting line	92
	Ordinary least squares estimation of the multiple regression model	93
	Statistical aspects of multiple regression	93
	Interpreting OLS estimates	94
	Pitfalls of using simple regression in a multiple regression context	97
	Omitted variables bias	99
	Multicollinearity	100
	Chapter summary	106
	Appendix 6.1: Mathematical interpretation of regression coefficients	106
	Endnotes	107
Chapter 7	Regression with dummy variables	109
	Simple regression with a dummy variable	111
	Multiple regression with dummy variables	112
	Multiple regression with both dummy and non-dummy explanatory variables	115

	Interacting dummy and non-dummy variables	117
	What if the dependent variable is a dummy?	119
	Chapter summary	120
	Endnote	120
Chapter 8	Regression with time lags: distributed lag models	121
	Aside on lagged variables	123
	Aside on notation	125
	Selection of lag order	128
	Chapter summary	131
	Appendix 8.1: Other distributed lag models	131
	Endnotes	133
Chapter 9	Univariate time series analysis	135
	The autocorrelation function	138
	The autoregressive model for univariate time series	142
	Nonstationary versus stationary time series	145
	Extensions of the AR(1) model	146
	Testing in the AR(p) with deterministic trend model	151
	Chapter summary	155
	Appendix 9.1: Mathematical intuition for the AR(1) model	156
	Endnotes	157
Chapter 10	Regression with time series variables	159
	Time series regression when X and Y are stationary	160
	Time series regression when Y and X have unit roots: spurious regression	164
	Time series regression when Y and X have unit roots: cointegration	165
	Time series regression when Y and X are cointegrated: the error correction model	171
	Time series regression when Y and X have unit roots but are not cointegrated	175
	Chapter summary	176
	Endnotes	177
Chapter 11	Applications of time series methods in macroeconomics and finance	179
	Volatility in asset prices	179
	Granger causality	186
	Vector autoregressions	193
	Chapter summary	205

Appendix 11.1: Hypothesis tests involving more than one coefficient	205
Endnotes	209
Chapter 12 Limitations and extensions	211
Problems that occur when the dependent variable has particular forms	212
Problems that occur when the errors have particular forms	213
Problems that call for the use of multiple equation models	216
Chapter summary	220
Endnotes	220
Appendix A Writing an empirical project	223
Description of a typical empirical project	223
General considerations	225
Project topics	226
Appendix B Data directory	229
Index	233

Introduction

There are several types of professional economists working in the world today. Academic economists in universities often derive and test theoretical models of various aspects of the economy. Economists in the civil service often study the merits and demerits of policies under consideration by government. Economists employed by a central bank often give advice on whether or not interest rates should be raised, while in the private sector, economists often predict future variables such as exchange rate movements and their effect on company exports.

For all of these economists, the ability to work with data is an important skill. To decide between competing theories, to predict the effect of policy changes, or to forecast what may happen in the future, it is necessary to appeal to facts. In economics, we are fortunate in having at our disposal an enormous amount of facts (in the form of “data”) that we can analyze in various ways to shed light on many economic issues.

The purpose of this book is to present the basics of data analysis in a simple, non-mathematical way, emphasizing graphical and verbal intuition. It focusses on the tools that economists apply in practice (primarily regression) and develops computer skills that are necessary in virtually any career path that the economics student may choose to follow.

To explain further what this book does, it is perhaps useful to begin by discussing what it does **not** do. **Econometrics** is the name given to the study of quantitative tools for analyzing economic data. The field of econometrics is based on probability and statistical theory; it is a fairly mathematical field. This book does not attempt to teach probability and statistical theory. Neither does it contain much mathematical content. In both these respects, it represents a clear departure from traditional econometrics textbooks. Yet, it aims to teach most of the practical tools used by applied econometricians today.

Books that merely teach the student which buttons to press on a computer without providing an understanding of what the computer is doing, are commonly referred to as “cookbooks”. The present book is **not** a cookbook. Some econometricians may interject at this point: “But how can a book teach the student to use the tools of econometrics, without teaching the basics of probability and statistics?” My answer is that much of what the econometrician does in practice can be understood intuitively, without resorting to probability and statistical theory. Indeed, it is a contention of this book that most of the tools econometricians use can be mastered simply through a thorough understanding of the concept of correlation, and its generalization, regression. If a student understands correlation and regression well, then he/she can understand most of what econometricians do. In the vast majority of cases, it can be argued that regression will reveal most of the information in a data set. Furthermore, correlation and regression are fairly simple concepts that can be understood through verbal intuition or graphical methods. They provide the basis of explanation for more difficult concepts, and can be used to analyze many types of economic data.

This book focusses on the **analysis** of economic data; **it is not a book about collecting economic data**. With some exceptions, it treats the data as given, and does not explain how the data is collected or constructed. For instance, it does not explain how national accounts are created or how labor surveys are designed. It simply teaches the reader to make sense out of the data that has been gathered.

Statistical theory usually proceeds from the formal definition of general concepts, followed by a discussion of how these concepts are relevant to particular examples. The present book attempts to do the opposite. That is, **it attempts to motivate general concepts through particular examples**. In some cases formal definitions are not even provided. For instance, P-values and confidence intervals are important statistical concepts, providing measures relating to the accuracy of a fitted regression line (see Chapter 5). The chapter uses examples, graphs and verbal intuition to demonstrate how they might be used in practice. But no formal definition of a P-value nor derivation of a confidence interval is ever given. This would require the introduction of probability and statistical theory, which is not necessary for using these techniques sensibly in practice. For the reader wishing to learn more about the statistical theory underlying the techniques, many books are available; for instance *Introductory Statistics for Business and Economics* by Thomas Wonnacott and Ronald Wonnacott (Fourth edition, John Wiley & Sons, 1990). For those interested in how statistical theory is applied in econometric modeling, *Undergraduate Econometrics* by R. Carter Hill, William E. Griffiths and George G. Judge (Second edition, John Wiley & Sons, 2000) provides a useful introduction.

This book reflects my belief that the use of concrete examples is the best way to teach data analysis. Appropriately, each chapter presents several examples as a means of illustrating key concepts. One risk with such a strategy is that some students might

interpret the presence of so many examples to mean that myriad concepts must be mastered before they can ever hope to become adept at the practice of econometrics. This is not the case. At the heart of this book are only a few basic concepts, and they appear repeatedly in a variety of different problems and data sets. The best approach for teaching introductory econometrics, in other words, is to illustrate its specific concepts over and over again in a variety of contexts.

Organization of the book

In organizing the book, I have attempted to adhere to the general philosophy outlined above. Each chapter covers a topic and includes a general discussion. However, most of the chapter is devoted to empirical examples that illustrate and, in some cases, introduce important concepts. Exercises, which further illustrate these concepts, are included in the text. Data required to work through the empirical examples and exercises can be found in the website which accompanies this book <http://www.wileyeurope.com/go/koopdata2ed>. By including real-world data, it is hoped that students will not only replicate the examples, but will feel comfortable extending and/or experimenting with the data in a variety of ways. Exposure to real-world data sets is essential if students are to master the conceptual material and apply the techniques covered in this book.

The empirical examples in this book are designed for use in conjunction with the computer package Excel. The website associated with this book contains Excel files. Excel is a simple and common software package. It is also one that students are likely to use in their economic careers. However, the data can be analyzed using many other computer software packages, not just Excel. Many of these packages recognize Excel files and the data sets can be imported directly into them. Alternatively, the website also contains all of the data files in ASCII text form. Appendix B at the end of the book provides more detail about the data.

Mathematical material has been kept to a minimum throughout this book. In some cases, a little bit of mathematics will provide additional intuition. For students familiar with mathematical techniques, appendices have been included at the end of some chapters. However, students can choose to omit this material without any detriment to their understanding of the basic concepts.

The content of the book breaks logically into two parts. Chapters 1–7 cover all the basic material relating to graphing, correlation and regression. A very short course would cover only this material. Chapters 8–12 emphasize time series topics and analyze some of the more sophisticated econometric models in use today. The focus on the underlying intuition behind regression means that this material should be easily accessible to students. Nevertheless, students will likely find that these latter chapters are more difficult than Chapters 1–7.

Useful background

As mentioned, this book assumes very little mathematical background beyond the pre-university level. Of particular relevance are:

1. Knowledge of simple equations. For instance, the equation of a straight line is used repeatedly in this book.
2. Knowledge of simple graphical techniques. For instance, this book is full of graphs that plot one variable against another (i.e. standard XY -graphs).
3. Familiarity with the summation operator is useful occasionally.
4. In a few cases, logarithms are used.

For the reader unfamiliar with these topics, the appendix at the end of this chapter provides a short introduction. In addition, these topics are discussed elsewhere, in many introductory mathematical textbooks.

This book also has a large computer component, and much of the computer material is explained in the text. There are myriad computer packages that could be used to implement the procedures described in this book. In the places where I talk directly about computer programs, I will use the language of spreadsheets and, particularly, that most common of spreadsheets, Excel. I do this largely because the average student is more likely to have knowledge of and access to a spreadsheet rather than a specialized statistics or econometrics package such as E-views, Stata or MicroFit.¹ I assume that the student knows the basics of Excel (or whatever computer software package he/she is using). In other words, students should understand the basics of spreadsheet terminology, be able to open data sets, cut, copy and paste data, etc. If this material is unfamiliar to the student, simple instructions can be found in Excel's on-line documentation. For computer novices (and those who simply want to learn more about the computing side of data analysis) *Computing Skills for Economists* by Guy Judge (John Wiley & Sons, 2000) is an excellent place to start.

Appendix 1.1: Mathematical concepts used in this book

This book uses very little mathematics, relying instead on intuition and graphs to develop an understanding of key concepts (including understanding how to interpret the numbers produced by computer programs such as Excel). For most students, previous study of mathematics at the pre-university level should give you all the background knowledge you need. However, here is a list of the concepts used in this book along with a brief description of each.

The equation of a straight line

Economists are often interested in the relationship between two (or more) variables. Examples of variables include house prices, gross domestic product (GDP), interest rates, etc. In our context a variable is something the economist is interested in and can collect data on. I use capital letters (e.g. Y or X) to denote variables. A very general way of denoting a relationship is through the concept of a function. A common mathematical notation for a function of X is $f(X)$. So, for instance, if the economist is interested in the factors that explain why some houses are worth more than others, he/she may think that the price of a house depends on the size of the house. In mathematical terms, he/she would then let Y denote the variable “price of the house” and X denote the variable “size of the house” and the fact that Y depends on X is written using the notation:

$$Y = f(X)$$

This notation should be read “ Y is a function of X ” and captures the idea that the value for Y depends on the value of X . There are many functions that one could use, but in this book I will usually focus on linear functions. Hence, I will not use this general “ $f(X)$ ” notation in this book.

The equation of a straight line (what was called a “linear function” above) is used throughout this book. Any straight line can be written in terms of an equation:

$$Y = \alpha + \beta X$$

where α and β are **coefficients**, which determine a particular line. So, for instance, setting $\alpha = 1$ and $\beta = 2$ defines one particular line while $\alpha = 4$ and $\beta = -5$ defines a different line.

It is probably easiest to understand straight lines by using a graph (and it might be worthwhile for you to sketch one at this stage). In terms of an XY graph (i.e. one which measures Y on the vertical axis and X on the horizontal axis) any line can be defined by its intercept and slope. In terms of the equation of a straight line, α is the intercept and β the slope. The intercept is the value of Y when $X = 0$ (i.e. point at which the line cuts the Y -axis). The slope is a measure of how much Y changes when X is changed. Formally, it is the amount Y changes when X changes by one unit. For the student with a knowledge of calculus, the slope is the first derivative, $\frac{dY}{dX}$.

Summation notation

At several points in this book, subscripts are used to denote different observations from a variable. For instance, a labor economist might be interested in the wage of every one of 100 people in a certain industry. If the economist uses Y to denote this variable, then he/she will have a value of Y for the first individual, a value of Y for the second individual, etc. A compact notation for this is to use subscripts so that Y_1