

**Methods for Business Analysis
and Forecasting: Text and Cases**

METHODS FOR BUSINESS ANALYSIS AND FORECASTING: TEXT AND CASES

PETER TRYFOS

York University



JOHN WILEY & SONS, INC.

New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

EDITOR	<i>Brad Wiley II</i>
PRODUCTION SERVICE	<i>J. Carey Publishing Service</i>
MARKETING MANAGER	<i>Jay Kirsch</i>
DESIGN DIRECTION	<i>Karin Kinchloe</i>
TEXT DESIGNER	<i>Nancy Field</i>
COVER DESIGNER	<i>Carol C. Grobe</i>
ILLUSTRATION COORDINATOR	<i>Sigmund Malinowski</i>

The book was set in Palatino by Eigentype Compositors and printed and bound by Malloy Lithographers. The cover was printed by Lehigh Press.

Recognizing the importance of preserving what has been written, it is a policy of John Wiley & Sons, Inc. to have books of enduring value published in the United States printed on acid-free paper, and we exert our best efforts to that end.

The paper in this book was manufactured by a mill whose forest management programs include sustained yield harvesting of its timberlands. Sustained yield harvesting principles ensure that the number of trees cut each year does not exceed the amount of new growth.

Copyright © 1998 by John Wiley & Sons, Inc. 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 as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (508) 750-8400, fax (508) 750-4470. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ@WILEY.COM.

Library of Congress Cataloging-in-Publication Data

Tryfos, Peter.

Methods for business analysis and forecasting : text and cases / by Peter Tryfos.

p. cm.

Includes bibliographical references (p.) and index.

ISBN 0-471-12384-6 (alk. paper)

1. Commercial statistics. 2. Business forecasting. I. Title.

HF1017.T79 1998

519.5-dc21

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

***METHODS FOR BUSINESS
ANALYSIS AND FORECASTING:
TEXT AND CASES***

PREFACE

This book is intended

- primarily as a main text for a one-semester or two-quarter elective course in methods for analysis and forecasting, for students of business or public administration at the upper undergraduate, M.B.A. or Ph.D. level;
- as a text in a "second business statistics" course;
- as a text with an applied orientation in service courses on regression, the linear model, and econometrics offered by departments of mathematics, statistics, and economics; and
- for self-study by professionals in finance, marketing, planning and other areas of business and government.

The text covers the principal methods for analysis and forecasting, which are traditionally included in separate courses with such titles as "Regression," "Linear Models," "Forecasting," and "Econometrics." There is considerable overlap in the content of such courses and mounting pressure in business curricula (especially at the upper undergraduate and graduate level) to utilize efficiently the time devoted to methods courses. This text deals in a balanced way with models for relationships involving quantitative or qualitative dependent and explanatory variables. Its structure can be described schematically as follows:

<i>Dependent variable(s)</i>	<i>Explanatory variables</i>	<i>Method</i>
Single, quantitative	Quantitative	Regression (Chs. 2-3)
Single, quantitative	Qualitative	Dummy variables and ANOVA (Ch. 4)
Single, quantitative	Quant. and qual.	Regression (Chs. 5-10)
Single, qualitative	Quant. and qual.	Classification (Chs. 11-12)
Many, quantitative	Quant. and qual.	Simultaneous equations (Ch. 13)

VI PREFACE

It is assumed that readers have taken an introductory course in business statistics or general statistics. Since such an introductory course is required by nearly all business schools, this prerequisite ought not be restrictive. Brief reviews of essential statistical concepts are provided in an appendix and in the text proper for readers who do not remember these concepts well.

Some care has been taken, however, to arrange the topics in two "streams" so that most of the text could be used even in a first course in statistical methods or one emphasizing data analysis for exploratory purposes. The recommended sequence for this purpose is described later in this preface.

An interest in quantitative methods is presumed. High school mathematics is adequate for the main text. A second appendix provides selected mathematical derivations for the benefit of students familiar with calculus and linear algebra.

The text is designed to accommodate at least two types of course design. The first (the traditional approach for a technical course) requires students to do exercises and problems and to write exams. Under the second approach (frequently used in business courses), students, perhaps working in groups, read, report, and discuss cases using the text as a resource; they may also do a project and write a paper involving modeling and forecasting in business. The text contains examples, exercises, problems, and small and large cases suitable for either one or for a mixture of these two approaches.

It is generally acknowledged that substantial, real applications of methods are an essential and highly desirable component of a text addressed to an applications-oriented audience. A case is a description of a real situation that lends itself to the application of methods. A case can be thought of as a large problem inviting reflection and providing an opportunity for discussion. Unlike a standard problem, but very much like the real world, the appropriate solution may not be obvious or unique.

Most examples, problems, and cases are based on the author's experience, and their data are real—not contrived. For the purposes of this text, however, actual names, places, and data are sometimes changed, for several reasons. In the first place, and in order to avoid unnecessary obsolescence, dates such as 1992 are written as 19X2, and so on. Secondly, organization names are sometimes changed in order to avoid subjecting the original ones to unnecessary criticism (which students tend to dispense ferociously). Thirdly, the data are occasionally transformed in order to preserve the confidentiality of original sources. Despite these changes, however, the essential integrity of the data, the problem, and the setting has been meticulously preserved.

It is fair to say that none of the methods described in the text (not even simple regression with few observations) is manually executed these days. The computer has made possible not only the implementation of complicated methods, but also the streamlining of the teaching of these methods. In an applied course, there now appears to be little need to dwell on special cases and formulas (for example, analysis of variance) when the general case can be

handled easily by the computer. Nor does it appear essential to describe complicated algorithms in detail if the objective of these algorithms can be stated clearly and tersely. For these reasons, the text emphasizes model formulation and interpretation rather than computation.

The text can be understood and many problems solved without the assistance of a computer program. However, the benefit from most of the cases and challenging problems cannot be realized—and the utility of the methods cannot be fully appreciated—without such assistance. It is, therefore, highly desirable that the student have access to a computer program capable of carrying out at least regression. The basic requirement is adequately fulfilled even by such “mainstream” business programs as Lotus and Excel. Programs such as SAS or SPSS are, of course, more than adequate. These or other programs (e.g., Minitab) form the ideal accompaniment to this text. The number of these programs is rapidly growing and their quality continually improving. The programs are becoming more and more user friendly, offering clear instructions, tutorials, and examples for implementing the methods described in the text either on-line or in their manuals. They can often be purchased at very reasonable retail or academic prices. It would consume too much space, therefore, for this text to provide instruction on the use of some of these programs and unwise to single out any one of them. For these reasons, the text is not designed to be supported by or to support any one particular program. The choice is the instructor’s.

Supplementary Material

The text is accompanied by a *diskette* containing the data (in ASCII form) used in the text and the cases. The file `readme . txt` in this diskette provides additional information on reading the data files.

An *Instructor’s Manual*, available to instructors from the publisher on request, includes solutions to all the problems, teaching notes describing the author’s treatment of each case and possible alternatives, and the programs used by the author for the solution of most problems and cases.

Adopters of this text may download the contents of the data diskette and the *Instructor’s Manual* from the publisher’s Web site at www.wiley.com/college/wave. Also available at this site are two complete *supplementary chapters* on Factor Analysis and Cluster Analysis. These topics round out a survey of the principal methods for business analysis for instructors who share this view. The two chapters are offered as an experiment; if there is sufficient interest, they may be considered for inclusion in any future edition of this text.

A First Course

An introductory course in statistics or business statistics explains probability theory, special distributions, confidence intervals, and statistical tests. This

VIII PREFACE

background is necessary for a good understanding of Chapter 3, Sections 4.5 to 4.8, Chapter 9, Sections 10.6 to 10.8, and Section 13.7. These chapters and sections (marked with a † in the table of contents) form a separate stream that is not required elsewhere in the text. Regression and its extensions can be explained, *if* that is necessary or desired, simply as a method of fitting a mathematical model to a set of observations. Chapter 2 explains the method of least squares, measures of fit and contribution, and the statistics appearing in computer output. The remainder of Chapters 4 to 10 can be clearly understood with just this background. Chapters 11 through 13 are largely self-contained and do not rely heavily on statistical inference; a few references to background concepts in these chapters are explained *in situ*. It is, therefore, possible to use the text even for a first course in statistical methods or one emphasizing data analysis—obviously with a certain sacrifice of rigor but not so much of utility.

ACKNOWLEDGEMENTS

I am indebted to the following organizations and individuals for data and other assistance: NCH Promotional Services, especially Mr. Wayne Moulard, Vice President of Analytical Services; Consumers Gas, especially Ms. Janet Holder, Vice President, Energy Services; Statistics Canada; Decima Research, a division of Hill and Knowlton (Canada) Ltd.; the Insurance Bureau of Canada; and Ontario Hydro. None of the cases, problems or examples in this text, however, should be interpreted as describing the current practices and policies of these organizations.

I have had the pleasure of teaching many good students. Several of their projects became the sources of examples, problems, and cases, as I acknowledge in many places in the text.

I am also grateful to the following people for positive and negative comments on earlier drafts of this book that resulted in significant improvements: Professor Arilee Bagley, SUNY at New Paltz; Professor James Daly, California Polytechnic State University; Professor Marshall Freimer, University of Rochester; Professor Derek Hart, McGill University; Professor George A. Marcoulides, California State University at Fullerton; Professor Kris Moore, Baylor University; Professor Larry Richards, University of Oregon; Professor Gordon S. Roberts, York University; Professor Mack C. Shelley II, Iowa State University; Professor Rafael Solis, California State University at Fresno; Professor Jef Steagall, University of North Florida; Professor Patrick A. Thompson, University of Florida; Professor N. Tryphonopoulos, York University; and Professor Rudy Wulleumier, Eastern Kentucky University.

I owe the greatest debt to my wife Barbara, who read critically several drafts of the manuscript and brought clarity to more places than I care to enumerate.

Any errors that remain despite the help of so many are entirely my own

CONTENTS

PREFACE

v

PART ONE: TEXT

1

CHAPTER 1 INTRODUCTION

3

- 1.1 Analysis and forecasting 3
- 1.2 Case: Life insurance 3
- 1.3 Case: The Metropolitan Transit Commission 4
- 1.4 Case: Salary and experience 5
- 1.5 Case: The City of West York 7
- 1.6 Case: Stock prices 9
- 1.7 Case: Mortgage loans 10
- 1.8 Exact and approximate relationships 11
- 1.9 Case: Gas consumption 13
- 1.10 To carry forward 21
- 1.11 The book in outline 22

CHAPTER 2 REGRESSION

24

- 2.1 Introduction and summary 24
- 2.2 Linear relationships 25
- 2.3 Estimation criteria 27
- 2.4 Least squares estimates 30
- 2.5 Special cases: One or two explanatory variables 31
- 2.6 An interpretation of the OLS estimates 34
- 2.7 Computer programs 35

X CONTENTS

2.8	Measuring the fit of a model	40
2.9	The contribution of explanatory variables	43
	Special case: The contribution of all explanatory variables, R^2	45
	Special case: The contribution of one explanatory variable, t -ratios	47
2.10	More examples	49
2.11	To sum up	55
	<i>Problems</i>	55

CHAPTER 3 INFERENCE IN REGRESSION[†]

65

3.1	Introduction and summary	65
3.2	Additional properties of OLS estimators	65
3.3	A population with special features	69
3.4	A population with different features	74
3.5	Confidence intervals and tests	75
	A confidence interval for a parameter	76
	A test for a parameter	77
	A confidence interval for a forecast	79
	A test for a subset of parameters	80
3.6	Rough checks of the assumptions of the classical linear model	81
3.7	An illustration	82
3.8	To sum up	88
	<i>Problems</i>	89

CHAPTER 4 ATTRIBUTES AS EXPLANATORY VARIABLES

95

4.1	Introduction and summary	95
4.2	Dummy variables	96
4.3	A digression: Extreme collinearity	99
4.4	Dummy variables, continued	102
4.5	One-way analysis of variance [†]	107
4.6	Two-way analysis of variance [†]	114
4.7	Interactions in analysis of variance [†]	122
4.8	Multi-way analysis of variance [†]	122
4.9	To sum up	123
	<i>Problems</i>	123

CHAPTER 5 NONLINEAR RELATIONSHIPS

136

5.1	Introduction and summary	136
5.2	Polynomials	136

5.3	Inverse relationships	144
5.4	Piecewise linear models	146
5.5	Logarithmic models	150
5.6	Nonlinear regression	155
5.7	To sum up	158
	<i>Problems</i>	159

CHAPTER 6 REGRESSION AND TIME SERIES

173

6.1	Introduction and summary	173
6.2	Trend	173
6.3	Trend and seasonality	177
6.4	Seasonal adjustment of time series	185
6.5	Method of moving averages	186
6.6	Time series models in perspective	192
6.7	To sum up	193
	<i>Problems</i>	193

CHAPTER 7 LAGGED VARIABLES

208

7.1	Introduction and summary	208
7.2	Distributed lags	208
7.3	The polynomial lag model	211
7.4	The geometric lag model	216
7.5	The adaptive expectations model	223
7.6	The stock adjustment model	226
7.7	To sum up	228
	<i>Problems</i>	229

CHAPTER 8 REGRESSION MISCELLANEA

235

8.1	Introduction and summary	235
8.2	Regression with zero intercept	236
8.3	Linear equality constraints	238
8.4	Linear equality and inequality constraints	242
8.5	Least absolute deviations	246
8.6	Multicollinearity	250
8.7	Stepwise and all subsets regression	253
8.8	Elements of experimental design	255
8.9	To sum up	261
	<i>Problems</i>	262

XII CONTENTS

CHAPTER 9 MORE ON INFERENCE IN REGRESSION[†] 271

- 9.1 Introduction and summary 271
- 9.2 Nonlinearity 272
- 9.3 Nonconstant variance 274
- 9.4 Nonindependence, serial correlation 280
 - Testing for serial correlation 282
 - Consequences of ignoring serial correlation 283
 - Estimation procedure 284
- 9.5 Specification error 287
 - Inclusion of irrelevant variables 289
 - Omission of relevant variables 289
 - Overlooking a nonlinear relationship 289
- 9.6 Generalized least squares (GLS) 290
- 9.7 Nonnormality, the generalized linear model (GLM) 291
- 9.8 Ridge regression 295
- 9.9 To sum up 299
 - Problems* 300

CHAPTER 10 AUTOREGRESSIVE MODELS 306

- 10.1 Introduction and summary 306
- 10.2 Autoregression 306
- 10.3 Series with trend and no seasonality 310
- 10.4 Series with seasonality 315
- 10.5 Autoregressive models in perspective 320
- 10.6 AR models[†] 324
- 10.7 Nonstationary series, *ARI* models[†] 328
- 10.8 Nonindependence, *ARMA* and *ARIMA* models[†] 329
- 10.9 To sum up 331
 - Problems* 332

CHAPTER 11 THE CLASSIFICATION PROBLEM 338

- 11.1 Introduction and summary 338
- 11.2 An illustration 339
- 11.3 Classification into two categories 343
- 11.4 Measuring the quality of a classification rule 346
- 11.5 More on the conditional probabilities 349
- 11.6 Classification into more than two categories 352
- 11.7 To sum up 359
 - Problems* 360

CHAPTER 12 MORE ON CLASSIFICATION 364

12.1 Introduction and summary	364
12.2 Probability models, two categories	364
12.3 Joint normal distributions, two categories	372
12.4 An alternative justification of the linear discriminant function	381
12.5 A logit model for classification into more than two categories	387
12.6 Multiple categories and normality	390
12.7 An alternative discriminant criterion	396
12.8 To sum up	398
<i>Problems</i>	399

CHAPTER 13 MODELS OF SYSTEMS 409

13.1 Introduction and summary	409
13.2 The Norgas case revisited	409
13.3 First generalization	414
13.4 Systems of equations	416
13.5 Structural or reduced form?	417
13.6 Estimation of reduced form parameters	418
13.7 Estimation of structural form parameters ⁺	423
Ordinary least squares (OLS)	424
Indirect least squares (ILS)	424
Two-stage least squares (2SLS)	426
Three-stage least squares (3SLS)	427
Full information maximum likelihood (FIML)	427
13.8 Case: The Canadian meat processing industry	428
13.9 To sum up	433
<i>Problems</i>	434

PART TWO: CASES 439

The Equal Billing Plan	441
Corporate compensation and discrimination	443
The ABC Manufacturing Company	446
Used car prices	448
The City of West York	454
The Energuide program	461
Carlsen's Brewery	467
Automobile insurance	470
Population projections	480

XIV CONTENTS

Electric power consumption	491
The Emergency Department	496
Northern Hydro	498
ACCS Inc.	507
The Great Wall Bank	517
Johnson & Phillips	524
Norgas Distributing Company	529

PART THREE: APPENDIXES

539

APPENDIX A: STATISTICAL ESSENTIALS

541

A.1 The summation notation	541
A.2 The mean, variance, and standard deviation of a variable	543
A.3 The mean and variance of a linear function of a variable	545
A.4 Frequency and relative frequency distributions	546
A.5 Probability distributions	547
A.6 Covariance and correlation coefficient of two variables	550
A.7 The mean and variance of a linear function of two variables	550
A.8 Joint distributions	551
A.9 Conditional distributions and independence	553
A.10 The covariance matrix of a set of variables	554
A.11 The mean and variance of a linear function of a set of variables	555
A.12 The covariance of linear functions of uncorrelated variables	557

APPENDIX B: MATHEMATICAL NOTES

559

B.1 Derivation of the least squares estimators	559
B.2 The general linear model in matrix notation	560
B.3 Covariance of OLS estimators and forecasts	560
B.4 A regression model with no explanatory variables	561
B.5 Analysis of variance	562
B.6 The OLS estimator of a simple model with zero intercept	563
B.7 The maximum likelihood method of estimation	563
B.8 Generalized least squares (GLS)	565
B.9 Derivation of the classification rule under normality	565
B.10 Structural and reduced form equations	567

SELECTED REFERENCES

569

INDEX

571

PART I

TEXT

INTRODUCTION

1.1 ANALYSIS AND FORECASTING

This, to repeat the title, is a text of methods for business analysis and forecasting. By *analysis* we have in mind the sifting through of available information in order to establish relationships and patterns. By *forecasting* we understand the projection of these relationships and patterns into the future. Analysis and forecasting are essential elements of most business problems.

The methods we shall be dealing with are quantitative, drawn primarily from the field of statistics. These are methods with applications not only in business and economics but also in engineering, the physical and social sciences, medicine, and many other fields.

Before we begin the study of these methods, it is useful to give some simple examples of problems we shall be addressing and to invite the reader to reflect on how he or she would approach these problems. The intention is not that the reader “solve” these problems at this stage; rather, it is hoped that this reflection will bring about some appreciation of the need for methods. The examples themselves are examined at greater length later on in this text.

1.2 CASE: LIFE INSURANCE

The simplest form of life insurance is the one-period term insurance policy: in return for a premium payable in advance, an insurance company agrees to pay the beneficiary of the insured the “face amount” in the event the insured dies during the period (e.g., one year) beginning on the date of issue of the policy. For example, the premium could be \$100 and the face amount \$20,000. In this case, the insurance company receives \$100 on the date the policy is issued and will pay the beneficiary \$20,000 if the insured person dies within one year from this date; if the insured does not die, the company does not pay anything.

The company, of course, cannot predict with certainty when the insured will die. If the insured dies within the year, the company will lose \$19,900; if the insured does not die, the company gains \$100.

An insurance company sells not one but many such policies. Some of its clients will die; some will survive. The company will make a profit if the payments to insureds who die do not exceed the revenue from premiums collected