Quantitative models in marketing research

Philip Hans Franses · Richard Paap

Quantitative Models in Marketing Research

Philip Hans Franses

and

Richard Paap

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York NY 10011-4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2001

Printed in the United Kingdom at the University Press, Cambridge

Typeface 10½/13pt Times System

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

Library of Congress Cataloguing in Publication data

Franses, Philip Hans, 1963–
Quantitative models in marketing research / by Philip Hans Franses and Richard Paap.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-80166-4
1. Marketing research–Mathematical models. I. Paap, Richard. II. Title.

HF5415.2 .F714 2001
658.8′3′015118–dc21 2001025424

ISBN 0 521 80166 4 hardback

Quantitative Models in Marketing Research

Recent advances in data collection and data storage techniques enable marketing researchers to study the characteristics of a large range of transactions and purchases, in particular the effects of household-specific characteristics and marketing-mix variables.

This book presents the most important and practically relevant quantitative models for marketing research. Each model is presented in detail with a self-contained discussion, which includes: a demonstration of the mechanics of the model, empirical analysis, real-world examples, and interpretation of results and findings. The reader of the book will learn how to apply the techniques, as well as understand the latest methodological developments in the academic literature.

Pathways are offered in the book for students and practitioners with differing statistical and mathematical skill levels, although a basic knowledge of elementary numerical techniques is assumed.

PHILIP HANS FRANSES is Professor of Applied Econometrics affiliated with the Econometric Institute and Professor of Marketing Research affiliated with the Department of Marketing and Organization, both at Erasmus University Rotterdam. He has written successful textbooks in time series analysis.

RICHARD PAAP is Postdoctoral Researcher with the Rotterdam Institute for Business Economic Studies at Erasmus University Rotterdam. His research interests cover applied (macro-)econometrics, Bayesian statistics, time series analysis, and marketing research.

Preface

In marketing research one sometimes considers rather advanced quantitative (econometric) models for revealed preference data (such as sales and brand choice). Owing to an increased availability of empirical data, market researchers can choose between a variety of models, such as regression models, choice models and duration models. In this book we summarize several relevant models and deal with these in a fairly systematic manner. It is our hope that academics and practitioners find this book useful as a reference guide, that students learn to grasp the basic principles, and, in general, that the reader finds it useful for a better understanding of the academic literature on marketing research.

Several individuals affiliated with the marketing group at the School of Economics, Erasmus University Rotterdam, helped us with the data and presentation. We thank Dennis Fok and Jedid-Jah Jonker for making some of the data accessible to us. The original data were provided by ROBECO and the Astmafonds and we thank these institutions for their kind assistance. Paul de Boer, Charles Bos, Bas Donkers, Rutger van Oest, Peter Verhoef and an anonymous reviewer went through the entire manuscript and made many useful comments. Finally, we thank the Rotterdam Institute for Business Economic Studies (RIBES) and the Econometric Institute for providing a stimulating research environment and the Netherlands Organization for Scientific Research (NWO) for its financial support.

PHILIP HANS FRANSES

RICHARD PAAP

Rotterdam, November 2000

Contents

Figures

Tables

1 Introduction and outline of the book

Recent advances in data collection and data storage techniques enable marketing researchers to study the characteristics of many actual transactions and purchases, that is, revealed preference data. Owing to the large number of observations and the wealth of available information, a detailed analysis of these data is possible. This analysis usually addresses the effects of marketing instruments and the effects of household-specific characteristics on the transaction. Quantitative models are useful tools for this analysis. In this book we review several such models for revealed preference data. In this chapter we give a general introduction and provide brief introductions to the various chapters.

1.1 Introduction

It is the aim of this book to present various important and practically relevant quantitative models, which can be used in present-day marketing research. The reader of this book should become able to apply these methods in practice, as we provide the data which we use in the various illustrations and we also add the relevant computer code for EViews if it is not already included in version 3.1. Other statistical packages that include estimation routines for some of the reviewed models are, for example, LIMDEP, SPSS and SAS. Next, the reader should come to understand (the flavor of) the latest methodological developments as these are put forward in articles in, for example, *Marketing Science*, the *Journal of Marketing Research*, the *Journal of Consumer Research* and the *International Journal of Research in Marketing*. For that matter, we also discuss interesting new developments in the relevant sections.

The contents of this book originate from lecture notes prepared for undergraduate and graduate students in Marketing Research and in Econometrics. Indeed, it is our intention that this book can be used at different teaching levels. With that aim, all chapters have the same format, and we indicate which sections correspond with which teaching level. In section 1.2, we will provide more details. For all readers, however, it is necessary to have a basic knowledge of elementary regression techniques and of some matrix algebra. Most introductory texts on quantitative methods include such material, but as a courtesy we bring together some important topics in an Appendix at the end of this book.

There are a few other books dealing with sets of quantitative models similar to the ones we consider. Examples are Maddala (1983), Ben-Akiva and Lerman (1985), Cramer (1991) and Long (1997). The present book differs from these textbooks in at least three respects. The first is that we discuss the models and their specific application in marketing research concerning revealed preference data. Hence, we pay substantial attention to the interpretation and evaluation of the models in the light of the specific applications. The second difference is that we incorporate recent important developments, such as modeling unobserved heterogeneity and sample selection, which have already become quite standard in academic marketing research studies (as may be noticed from many relevant articles in, for example, the *Journal of Marketing Research* and *Marketing Science*). The third difference concerns the presentation of the material, as will become clear in section 1.2 below. At times the technical level is high, but we believe it is needed in order to make the book reasonably self-contained.

1.1.1 On marketing research

A useful definition of marketing research, given in the excellent introductory textbook by Lehmann et al. (1998, p. 1), is that "[m]arketing research is the collection, processing, and analysis of information on topics relevant to marketing. It begins with problem definition and ends with a report and action recommendations." In the present book we focus only on the part that concerns the analysis of information. Additionally, we address only the type of analysis that requires the application of statistical and econometric methods, which we summarize under the heading of quantitative models. The data concern revealed preference data such as sales and brand choice. In other words, we consider models for quantitative data, where we pay specific attention to those models that are useful for marketing research. We do not consider models for stated preference data or other types of survey data, and hence we abstain from, for example, LISREL-type models and various multivariate techniques. For a recent treatment of combining revealed and stated preference data, see Hensher et al. (1999). Finally, we assume that the data have already been collected and that the research question has been clearly defined.

The reasons we focus on revealed preference data, instead of on stated preference data, are as follows. First, there are already several textbooks on LISREL-type models (see, for example, Jöreskog and Sörbom, 1993) and on multivariate statistical techniques (see, for example, Johnson and Wichern, 1998). Second, even though marketing research often involves the collection and analysis of stated preference data, we observe an increasing availability of revealed preference data.

Typical research questions in marketing research concern the effects of marketing instruments and household-specific characteristics on various marketing performance measures. Examples of these measures are sales, market shares, brand choice and interpurchase times. Given knowledge of these effects, one can decide to use marketing instruments in a selective manner and to address only apparently relevant subsamples of the available population of households. The latter is usually called segmentation.

Recent advances in data collection and data storage techniques, which result in large data bases with a substantial amount of information, seem to have changed the nature of marketing research. Using loyalty cards and scanners, supermarket chains can track all purchases by individual households (and even collect information on the brands and products that were not purchased). Insurance companies, investment firms and charity institutions keep track of all observable activities by their clients or donors. These developments have made it possible to analyze not only what individuals themselves state they do or would do (that is, stated preference), but also what individuals actually do (revealed preference). This paves the way for greater insights into what really drives loyalty to an insurance company or into the optimal design for a supermarket, to mention just a few possible issues. In the end, this could strengthen the relationship between firms and customers.

The large amount of accurately measured marketing research data implies that simple graphical tools and elementary modeling techniques in most cases simply do not suffice for dealing with present-day problems in marketing. In general, if one wants to get the most out of the available data bases, one most likely needs to resort to more advanced techniques. An additional reason is that more detailed data allow more detailed questions to be answered. In many cases, more advanced techniques involve quantitative models, which enable the marketing researcher to examine various correlations between marketing response variables and explanatory variables measuring, for example, household-specific characteristics, demographic variables and marketing-mix variables.

In sum, in this book we focus on quantitative models for revealed preference data in marketing research. For conciseness, we do not discuss the various issues related to solving business problems, as this would require an entirely different book. The models we consider are to be viewed as helpful practical tools when analyzing marketing data, and this analysis can be part of a more comprehensive approach to solving business problems.

1.1.2 Data

Marketing performance measures can appear in a variety of formats. And, as we will demonstrate in this book, these differing formats often need different models in order to perform a useful analysis of these measures.

To illustrate varying formats, consider "sales" to be an obvious marketing performance measure. If "sales" concerns the number of items purchased, the resultant observations can amount to a limited range of count data, such as 1, 2, 3, ... However, if "sales" refers to the monetary value in dollars (or cents) of the total number of items purchased, we may consider it as a continuous variable. Because the evaluation of a company's sales may depend on all other sales, one may instead want to consider market shares. These variables are bounded between 0 and 1 by construction.

Sales and market shares concern variables which are observed over time. Typically one analyzes weekly sales and market shares. Many other marketing research data, however, take only discrete (categorical) values or are only partially observed. The individual response to a direct mailing can take a value of 1 if there is a response, and 0 if the individual does not respond. In that case one has encountered a binomial dependent variable. If households can choose between three or more brands, say brands A, B, C and D, one has encountered a multinomial dependent variable. It may then be of interest to examine whether or not marketing-mix instruments have an effect on brand choice. If the brands have a known quality or preference ranking that is common to all households, the multinomial dependent variable is said to be ordered; if not, it is unordered. Another example of an ordered categorical variable concerns questionnaire items, for which individuals indicate to what extent they disagree, are indifferent, or agree with a certain statement.

Marketing research data can also be only partially observed. An example concerns donations to charity, for which individuals have received a direct mailing. Several of these individuals do not respond, and hence donate nothing, while others do respond and at the same time donate some amount. The interest usually lies in investigating the distinguishing characteristics of the individuals who donate a lot and those who donate a lesser amount, while taking into account that individuals with perhaps similar characteristics donate nothing. These data are called censored data. If one knows the amount donated by an individual only if it exceeds, say, $10, the corresponding data are called truncated data.