Repeat-Buying

Facts, Theory And Applications

New Edition

A.S.C. EHRENBERG
London Business School

CHARLES GRIFFIN & COMPANY LTD
London
OXFORD UNIVERSITY PRESS
New York

CHARLES GRIFFIN & COMPANY LIMITED
16 Pembridge Road, London W11 3HL, U.K.

Published in the U.S.A. by Oxford University Press, 200 Madison Avenue, New York, N.Y. 10016

Oxford University Press ISBN 0-19-520634 7

First published North Holland Publishing Company 1972
This edition first published by Charles Griffin & Company, 1988

British Library Cataloguing in Publication Data

Ehrenberg, A.S.C.
Repeat-buying: facts, theory and applications. — New ed., rev.
I. Brand choice
I. Title
658.8'343 HF5415.3

ISBN 0-85264-287-3

Printed & bound in Great Britain by
Redwood Burn Limited, Trowbridge, Wilts

Repeat-Buying

Facts, Theory and Applications

Of the thousand and one variables which might affect buyer behaviour, it is found that nine hundred and ninety-nine usually do not matter. Many aspects of buyer behaviour can be predicted simply from the penetration and the average purchase frequency of the item, and even these two variables are interrelated.

CONTENTS

*Part IV and Chapter 13 are for the more technical reader.

FOREWORD

The buying and consumption of goods plays a major role both in the economy and in private life. The purpose of this book is to bring together various empirical and theoretical results concerning consumers' *repeat-buying* behaviour. This is a relatively narrow but important topic since the continuing consumption (or sales) of many goods depends on repeat-buying – for example, on the extent to which different consumers buy an item more than once in a given time-period, and on the extent to which the same consumers buy the item again in the *next* time-period. The empirical results discussed here relate to frequently-bought branded consumer goods and can largely be summarised by the "NBD/LSD" theory of repeat-buying (now usually called the NBD Model). This book is essentially a monograph on the present state of this theory and on some of its initial practical applications.

The book describes what a certain aspect of consumer behaviour – repeat-buying – is generally like. It does not deal directly with questions of how to *change* consumer behaviour. It therefore deals with a part of the *context* of marketing management rather than with the *execution* of marketing management. (It is not so much a book on how to build aeroplanes, as it were, but on certain elementary strands in aerodynamics.)

The results described here will be of relevance to economists and sociologists dealing with the consumer, to marketing practitioners and market researchers, and to students in these areas. The results also have some potential *methodological* interest to those engaged in model-building in the social or management sciences generally.

The approach centres on first of all establishing empirical generalisations. The fundamental finding is that despite all the apparent complexity of the buying situation, there are simple results. Thus the same empirical patterns hold for different brands and product-fields. In general, the observed patterns of repeat-buying do not depend on the brand or product itself, nor on what else buyers of the brand buy as well, nor yet on external factors such as advertising, pricing, distribution, etc. Instead, repeat-buying patterns depend only on buyer behaviour characteristics as such (e.g. how many people buy the brand and how often). All this applies in the "stationary" situation where external marketing influences are effectively in equilibrium so that there is little or no trend in aggregate sales – the kind of situation which in practice occurs in most markets most of the time.

A natural reaction to such practical findings is to wonder how they work *in theory*. Here there have also been successful results in terms of modelling the various empirical patterns in mathematical form and combining them into a coherent theory. This can be summarised by the expression $\{1 + a\, \Sigma T_i\, (1 - u_i)\}^{-k}$, (the probability-generating function of the multivariate Negative Binomial Distribution or NBD, as is discussed in Chapter 7). This one expression essentially covers all that is contained in this book about repeat-buying. In itself the expression is not of practical value to the general practitioner, but there are numerous much simpler formulae for specific purposes and for practical use.

A good deal of effort has in fact gone into developing *applications* of the theoretical results, e.g. to the understanding of consumer behaviour and to the elucidation of practical marketing problems. Such applications are illustrated in this book by a varied selection of published and previously unpublished material. The theory for example provides interpretative norms for evaluating situations where some trend in sales has occurred – i.e. "non-stationary" situations. In general, the theory has already been applied in thousands of cases, and continues to be in daily use.

The repeat-buying theory developed here is *descriptive*. It describes *how* (rather than perhaps *why*) consumers behave as they do, and on what factors this does (or does not) depend. Before one can explain the individual consumer's decision-processes and behaviour, one needs to know and understand the overt behaviour that has to be explained – what generalisable regularities there are and what apparent inconsistencies. And knowing the factors from which one can successfully predict consumer behaviour (and especially also the factors which do *not* matter in this respect) does in fact already provide major insights into its nature.

The formulation of the theory is mathematical but the mathematics are treated at different levels. The heavier mathematics have been restricted to Part IV, for the more technical reader. The details of the working formulae which are given in the rest of the text can however readily be passed over on a first or second reading, or even altogether by the reader who is not going to be personally involved in carrying through the numerical calculations: He needs to know that formulae exist to perform certain functions, but what really matters is understanding the empirical facts and concepts of buyer behaviour, and recognising the potentialities for practical applications of the theory.

The theoretical model in this book generally works well, but there are exceptions. In particular, there are certain "boundary" situations where the theory does not work at all, such as for very short time-periods. Indeed, the existing theory seems to be conceptually wrong and it should be possible to develop a repeat-buying theory based on a different analysis-concept (perhaps the "purchase week" rather than the presently-used "purchase occasion"). This would then deal with these discrepancy problems whilst still giving virtually the same results as the present NBD/LSD theory in all those areas where the latter already successfully models real buyer behaviour.

A similar reformulation of repeat-buying theory already occurred some years ago (in moving from a model expressed in terms of the "amount bought" to the "purchase occasion"). Various previously intractable problems as well as the initial results could then be dealt with by the reformulated theory. The same process of "wrong but right" seems also to apply to the present state of the NBD/LSD theory. The justification for the approach outlined in this book is that in a very wide range of practical conditions it *does* work, not always perfectly, but quite well enough to give both practitioners and theoreticians practical tools and greater insights into important aspects of consumer behaviour.

CONTENTS

The book brings together a variety of results obtained stage-by-stage in an extensive sequence of studies carried out over more than 10 years *. It is structured in six main parts, each of two chapters. The first four parts deal with repeat-buying as such, as follows:

Part I is of an introductory nature. The general nature of buyer behaviour is outlined in Chapter 1 **. Chapter 2 introduces the existence of empirical regularities in *repeat-buying*, and goes on to illustrate how these regularities can be successfully modelled by theoretical formulae, and be of practical value.

Part II gives the basic findings on repeat-buying for the general reader. Chapter 3 first illustrates the repeat-buying structure of a typical product-field in numerical form. Chapter 4 then describes the

* Published papers are listed in the bibliography but are referred to in the main text only when there is a special historical point to be made or where there are fuller details in the original.

** Previous *research* into buyer behaviour is reviewed briefly at the beginning of Chapter 11 in Part VI.

basic repeat-buying relationships in conceptual and (simple) mathematical form (a worked numerical example being given in Appendix A).

Part III covers a selection of the practical applications of the repeat-buying theory that have been developed so far, both under "stationary" and under more dynamic conditions.

Part IV describes the more mathematical aspects of the NBD (Chapter 7) and LSD (Chapter 8) models. *This part is for the more technical reader.*

The various repeat-buying results in Parts I to IV are put into a wider context in Parts V and VI. Part V gives some basic results of *multi-brand* buying behaviour, i.e. the extent to which buyers of one brand also buy other brands. Here too a great many simple regularities have been established in recent years, but the theoretical foundations are as yet far less well developed than in repeat-buying.* Part VI briefly discusses methodological implications and future trends. Thus

Part V shows in Chapter 9 the structuring of multi-brand buyer behaviour for the same product-field as was covered in terms of repeat-buying in Chapter 3. Chapter 10 summarises the main theoretical results in multi-brand buying that are as yet available.

Part VI briefly describes in Chapter 11 the general background of other research into buyer behaviour and some of the theoretical inter-relationships and explanations that are now beginning to appear in the present work, and discusses in Chapter 12 the need for deliberately investing in the development of proto-type applications of the results.

A worked numerical example of the NBD/LSD calculations and a number of tables to facilitate applications of the theory are given in the Appendices.

* The Dirichlet model in the new Chapter 13 now provides this.

ACKNOWLEDGEMENTS

The preparation of this book has been facilitated through a visiting appointment as Albert Wesley Frey Professor of Marketing at the University of Pittsburgh during 1970/1. The aim of the book has been to bring together a variety of published as well as unpublished results, and in this I have been much helped by Mr. G.C. Naylor. I am also indebted to various colleagues for their helpful comments on the draft of this book – and in particular to Mr. P. Charlton (who also assisted with the Appendices), Prof. J.U. Farley and Mr. G.J. Goodhardt – and to Mr. M. Collins, a director of Aske Research Ltd., for work on the coded form of the "Loyalty Reports" on which Chapters 3 and 9 are based.

Most of the results described in this book were obtained by Aske Research Ltd. in work carried out in the U.K. and U.S. since 1963, major clients including Cadbury-Schweppes, Esso, Gallaher, ICI, J. Walter Thompson, Pillsbury, Shell-Mex & B.P., and Unilever. Specific examples and background information have also been provided by work for Allied Breweries, Beecham, H.J. Heinz, Lyons Maid and Nabisco. The earliest work in the late 1950's was carried out at Attwood Statistics Ltd.

The data analysed come mainly from various consumer panels such as the Television Consumer Audit (as operated for the owners by Audits of Great Britain Ltd,) and the panels run by Attwood Statistics Ltd., Research Bureau Ltd. (Unilever) and the Market Research Corporation of America. I am indebted to these and other data-collection firms for permission in the past to quote data in coded form.

Much of the work on the LSD model (Chapter 8) was carried out by Dr. C. Chatfield (now of the University of Bath) as consultant to Aske Research Ltd. The formulation of the multivariate negative binomial model and of the "conditional" form of analysis (Chapter 7) was developed by Mr. G.J. Goodhardt, a director of Aske Research Ltd., who has also more generally played a very major role in providing a deeper understanding of the theoretical results and in developing their practical applications, as is evidenced by the references to our published work.

March 1971

A.S.C. Ehrenberg
London Business School, and
Graduate School of Business Studies,
University of Pittsburgh

PREFACE

The 1972 edition of *Repeat-Buying* has been unavailable for many years. Yet the empirical patterns and theoretical superstructure described there still hold. Furthermore, interest in consumers' buying behaviour has grown steadily since those early days.

It is therefore good that the book is being re-issued by such a well-established publisher of statistical texts as Griffins. No changes to the text had to be made other than a small number of minor clarifications or corrections in the wording. Chapter 13 on the Dirichlet Model is however new.

The Dirichlet Model integrates all the repeat-buying results in Parts I to IV of the original book and all the multi-brand ones in Part V into one single model.

The development of such a parsimonious model by my long-standing colleagues Professor Gerald Goodhardt and Dr. Christopher Chatfield has been a major advance. The new chapter is an edited version of a paper on the Dirichlet read by us to the Royal Statistical Society in 1984. The technical calculations for the Dirichlet are described in the new Appendix C, prepared by my colleague Dr. Mark Uncles.

A Floppy Disc

A floppy disc is now also available at low cost, giving IBM PC or compatible software for the theoretical NBD and Dirichlet calculations discussed in this book and also for the relevant tabulations of consumer panel data. Enquiries should be sent to Dr. Mark Uncles at the London Business School, Sussex Place, London NW1 4SA, England (Phone: 01-262-5050).

Other Developments Since 1972

For the more knowledgeable reader, I can also add that a number of the gaps such as those that were listed in Section 11.7 originally have now been closed.

Thus on the theoretical side I noted in 1972 (footnote p. 63) that there was at that stage no strong reason for one of the two assumptions (the "Gamma") leading to the NBD model, except that the model worked. The theoretical gap was closed already a few months later in 1972 by a very elegant result by Gerald Goodhardt and Christopher Chatfield which is now noted in Section 13.2. It marked the first step towards developing the Dirichlet model.

On the empirical side, the brand-choice results have now been extended to the choice of distribution channels, as is noted on p. 241. In brief, it has been shown that consumers' store-choice is like brand-choice, both in the UK and the US. More generally, the wider range of conditions under which the main repeat-buying and/or multi-brand buying results have been found to hold is summarised in Table 1. They now also include one or two industrial products or services like Aviation Fuel and Management courses, and other countries like Japan.

A less technical exposition, *Understanding Buyer Behaviour*, has also been developed (Ehrenberg and Goodhardt 1979a).

Table 1. Conditions Under Which the Empirical Patterns and Theories Hold

— Aviation fuel, Biscuits, Breakfast cereals, Butter, Canned vegetables, Cat and dog foods, Cocoa, Coffee, Confectionery, Convenience foods, Cooking fats, Cosmetics, Detergents, Disinfectants, Flour, Food drinks, Gasoline, Household soaps, Household cleaners, Instant potatoes, Jams and jellies, Margarine, Motor Oil, Polishes, Processed cheese, Refrigerated dough, Sausages, Shampoos, Soft drinks, Soup, Take-home beer, Toilet paper, Toilet soap, TV programmes.
— The leading brands in each product-field; Large, medium and small pack-sizes.
— Retail chains; Individual stores; Brands within chains.
— Great Britain, Continental Europe, USA, Japan; Demographic subgroups; 1950–1985.
— Analysis periods ranging from 1 week to 12 months.

A different extension is that broadly similar results to those in this book have also emerged from the study of viewers' choice of *television programmes* (e.g. Goodhardt et al. 1975, 1987; Barwise and Ehrenberg 1988).

Practical applications of the buyer behaviour results (Chapters 5, 6 and 12) have also grown since 1972. Two examples are the development of the ATR theory of consumer behaviour (e.g. Ehrenberg 1974; Barwise and Ehrenberg 1988), and applications to new product development (e.g. Ehrenberg 1987).

Consumer Dynamics

Finally, a start has been made in tackling *consumer dynamics*. The present book describes how all the different aspects of buying behaviour in a stationary no-trend situation can generally be predicted just from a brand's market-share (e.g. §12.4). The question however remains why one brand sells more than another. What determines market-share?

To start to tackle this, one or two real-life cases of changing sales have now been examined, such as a certain new brand launch or seasonal ups and downs (e.g. Wellan and Ehrenberg 1987a,b). Many more such case studies are however still needed, before any real progress can be achieved. But cases of large trends tend to be rare.

Our main emphasis in studying consumer dynamics in recent years has therefore been on *experimental work*. Here there has been some very promising early progress, particularly on pricing for example (e.g. Ehrenberg 1986, Ehrenberg and England 1987). But once again, the vast bulk of such work still needs to be tackled.

Acknowledgements

I am exceptionally glad to be able to record my debt to Gerald Goodhardt for more than thirty years' collaboration in this area, culminating at least momentarily in the vastly successful and elegant Dirichlet theory, developed with Christopher Chatfield. I also acknowledge permission from them both and from the Royal Statistical Society to reprint the edited version of our joint paper as my new Chapter 13.

In addition I am grateful to my colleague Mark Uncles at the LBS Centre for Marketing and Communication for preparing the new Appendix C and the floppy disc. The programming of the Dirichlet on the latter is due to

Dr. Richard Dunn of Bristol University, developed in early applications of the model in the Geography Department there. We are grateful to him for making this now more widely available.

The new results incorporated or referred to in the re-issue of the book largely stem from work at the LBS's Centre for Marketing and Communication (CMaC) which is supported by over thirty leading British companies. Preparation of the re-issue itself has been part of a programme of work supported by 35 companies in the UK, and by Colgate Palmolive, General Foods, General Mills, M & M Mars, the Ogilvy Center, and Procter and Gamble in the USA.

A.S.C. Ehrenberg
Centre for Marketing and Communication,
London Business School.

DETAILED CONTENTS

* The two chapters in Part IV are for the more technical reader.

PART I

INTRODUCTION

CHAPTER 1

BUYER BEHAVIOUR

1.1. The Role of Repeat-Buying

This book deals with the repeat-buying of frequently-bought branded goods. In this chapter, we first outline some general aspects of buyer behaviour in § 1.2 and then consider the data-collection or measurement procedures most commonly used in § 1.3, the general choice of analysis-variables in § 1.4, and some specific indices of repeat-buying in § 1.5. The crucial point that regularities occur in consumers' observed repeat-buying is introduced in Chapter 2, together with theoretical formulations to summarise and interrelate these regularities, and a first example of a practical application of the results. Both theory and applications will be more fully developed in Parts II to IV.

Consumer behaviour is mainly studied by manufacturers of the goods that are consumed. Initially, a manufacturer tends to be production-orientated and thinks mainly in terms of his sales: so many tons of X, so many millions of cases of Y, so many pounds or dollars worth of Z – and sales perhaps going up or going down. Then comes some realisation that individual and highly varied people are involved who buy or use his product (the marketing or consumer-orientated approach). Data are therefore collected which allow him to separate sales into two components – numbers of buyers and how much they buy. There comes also the discovery that some consumers buy the brand or product in question far more frequently than do others (with a small proportion of buyers usually accounting for the bulk of sales), and that consumers may vary in their needs and habits (consumer profiles and market segmentation).

The data also show that buyers of a brand do not necessarily satisfy all their needs through this one brand, but tend to buy other brands as well – one of several points first documented in some quantitative form in Brown's pioneering articles on brand-loyalty many years ago [Brown 1951]. And it also appears that a brand does not always keep all its buyers: perhaps only 60% of the people who bought it last month buy it again this month. Such a loss of repeat-buyers might seem to imply that only 36% of last month's buyers would buy it again *next* month (60% of 60%), and only about 20% the month after that, and so on.

The failure to achieve 100% repeat-buying might indicate a loss of consumer demand or acceptance and presage a catastrophic decline in sales. On the other hand, it might be that a large turnover of customers is quite normal under steady sales conditions, with the lost buyers constantly being replaced by new buyers (the "leaky bucket" theory). Or perhaps the "lost" buyers merely lapse for a time and normally come back again later.

But what really happens in terms of repeat-buying, and what its implications are, is broadly speaking the topic of this book. Is the month-by-month incidence of repeat-buying for Brand X really 60%? And if so, is 60% in fact "high" or "low" – as *many* as 60% repeat-buyers, or *only* 60% repeat-buyers? What does this depend on? Are the implications of 60% "good" or "bad" or "normal"? Is there additional erosion of a brand's repeat-buying franchise in time-periods further apart? And what are the underlying mechanics of brand-choice and repeat-buying generally? Are there generalisable norms of behaviour?

Some interpretative background of knowledge is clearly needed. Repeat-buying is any situation where a person buys the item in question more than once. It can be studied in different ways, but these must ultimately lead to the same insights and the same practical conclusions. The justification of the particular analytic approach taken in this book – which turns on the NBD/LSD theory – is that it works in practice in giving simple and interrelated results which hold under a wide range of conditions.

The main result is that repeat-buying of any item from any frequently-bought branded product-field tends, within certain broad limits, to follow a common pattern and can be dealt with by one single theory, irrespective of what the brand or product is and irrespective of what other brands its buyers may or may not have bought as well. This simple result is noteworthy, given the large variety of different conditions under which buyers make their purchasing decisions. It makes it possible to establish an empirically-based theory and to discern simple interpretative norms to assess whether any particular observation (e.g. the 60% incidence of repeat-buyers month-by-month) is *high* or *low*, or perhaps merely *normal*.

1.2. Consumer Behaviour

Consumer behaviour is complex. There are pre-purchase needs and attitudes, the experience of previous usage, and external influences such as advertising and promotion, retail availability, personal selling and word-of-mouth effects, and differences in product-formulation, packaging and pricing. Decisions have to be made about whether and what to buy, how much and at what cost, and when and where. Then there are various usage or consumption activities and responses which generate post-usage feelings of satisfaction and changes in attitude. These various aspects of consumer behaviour are briefly touched on in Chapter 11 (§11.2), but in the bulk of this book we consider only buyer behaviour in the narrow sense of the buying act itself, and repeat-buying in particular.

Repeat-buying is one aspect of the way in which consumers buy fast-moving goods. These are the kinds of products which are bought fairly frequently, like the various lines of food and drink, of soap and toiletries, of cigarettes and petrol, and so on, which tend to be generally available from grocery and other retail outlets. In as far as the same person buys any particular item more than once within a relatively short time-period, such as a week, a month, or a year, the notion of *repeat-buying* becomes particularly relevant.

The factors involved in the buying situation are highly varied. The kinds of products dealt with here are typically sold in pre-packaged branded form, but they may be available in different types of packages, in different pack-sizes, in different varieties (e.g. flavours), at different levels of quality and price (e.g. grades of petrol) and under different manufacturers' brand-names. The items are generally low-priced (although a purchase of several gallons of petrol, say, may be relatively costly). Some products are bought more or less as necessities or staple commodities. (Most people tend to buy *some* bread, potatoes, meat, vegetables, soap, petrol – if they have a car – and so on). Others are bought more for variety with an element of luxury (different types of breakfast cereals, or the modern "convenience" foods, say). These various distinctions might affect the regularity with which the items were bought, i.e. their repeat-buying patterns.

Further variations in the buying situation are that some products like tinned vegetables have a long storage life and can be stocked up, others (like frozen vegetables) can be stored for a relatively short time only (other than in a deep-freeze), and some can only be bought and stored in limited amounts (like petrol). Some products are generally used up in one go once the package is opened (e.g. frozen or tinned vegetables), others tend to be used in small amounts but may have a relatively short storage life once opened (e.g. breakfast cereals or butter) or quite a long one (e.g. toilet soap or toothpaste).

For some products the different brands available are virtually indistinguishable apart from their brand-names and possibly their packaging (e.g. petrol, or some brands of washing powder or of cigarettes), in others the "brands" are clearly different (e.g. breakfast cereals) or there are different "varieties" of the same brand (e.g. different flavours), while in yet other cases there may be two or more distinct *types* of product competing directly (soups in cans and soups in foil packets).

Some products have a single end-use and others have a variety of different end-uses (e.g. butter is used for cooking, frying, baking, and to put on bread). Some end-uses can be satisfied by only one product (e.g. petrol), while others are open to a variety of products (e.g. eggs and bacon and/or cereals for breakfast, or neither).

Shopping habits for different products and retail availability also greatly vary. Some may be bought nearly every day as demand requires (break, milk, or cigarettes), some are mainly bought at most once a week, often as part of a general weekly purchase trip for household needs. Some are always kept in stock in the household and some are only bought again some time after the initial purchase has run out.

Grocery outlets usually sell several different brands of the same product (including possibly the retailer's own "private label" version), but may not stock the particular brand required. In contrast, a given brand of petrol is usually only available from a solus-site outlet but is then always "in stock". The extent of retail availability is often correlated with total sales or market-share, and here different brands differ widely, with the market-leader often having 30% or more of the market, whilst small brands may account for only 1% or less.

Promotional support (advertising, special offers, etc.) tends to vary greatly by product and brand. It is usually much heavier for market-leaders than for smaller-selling brands, and differs also in type and content. Consumer attitudes to different products and brands may therefore also differ.

Individual consumers differ greatly in their consumption levels of particular products or brands, quite apart from obvious factors like household size. Some households consume several tins of soup per week every week, and many others only buy a few tins in a year. Some people mostly buy one single brand, pack-size, variety or whatever, while others switch around a great deal.

The buyer's role tends also to vary. Purchases made by the housewife, for example, may be made primarily for her own usage (e.g. laundry products), or for the family as a whole (many food products), or with individual family-member's tastes predominating (as perhaps with types of breakfast cereals or toothpaste), or for someone else's usage altogether (as with pet-foods).

It is clear from this brief review that there are many different buying situations and a possibly almost bewildering set of choices and decisions which the buyer has to make. But in as far as the consumer is dealing with frequently-bought and low-priced items, the amount of risk involved in the individual purchasing transaction is low and there is ample opportunity to develop *habits*, to simplify the repetitive choice-situation.

Logically, when a buyer is choosing between different manufacturers' brands which are of more or less identical product-formulation, pricing and availability, it might appear equally "rational" either to buy the same brand as before or to buy a different brand. Empirically, the finding is that most people tend to develop habits of buying one or some small number of brands, each fairly regularly.

A simplifying tendency towards brand-loyalty and repeat-buying appears to exist in practice. What its origins and causes are in psychological terms (e.g. "risk-reduction", "brand-image differentiation", "advertising", "segmentation", "learning", "cognitive dissonance", "reinforcement", etc.) is beyond the intended scope of this book, and is in any case still largely unclear*. First we need to understand rather precisely what it is that we would want to explain – i.e. *how* people buy, before we can succesfully consider *why*. A theory is needed at this stage which describes and interrelates, rather than one which aims to provide instant explanation.

The results reported here still do not say why consumers buy a particular product, or why they choose one brand rather than another. Instead, if in a given time-period a certain number of people buy a particular brand, the results show *how* they do this, and with what other facts this ties in. For example, the number of people who buy a particular brand at all in a given period is related to how *often* they buy it, to how many additional buyers of the brand there will be in some longer period, to how often these additional buyers will buy the brand, to what *other* brands any of these people also buy, and to how often they do so. And we note how these patterns are interrelated for different brands, for different product-fields, and for different lengths of time-period.

The unifying tendency towards some more or less "habitual" buying behaviour, instead of the variety of buying situations, is therefore what

* Reference to some leading accounts of the broader aspects of consumer behaviour is made in §11.2 of Chapter 11.

seems to dominate actual buyer behaviour. It underlies the simple empirical generalisations which are discussed in this book. The theory described here embraces many variations in behaviour – some more easily than others – and is therefore of significance in explaining actual buyer behaviour for frequently-bought branded consumer goods.

1.3. The Data Analysed

The main data analysed in studying buyer behaviour (in the literal sense of "buying") are records of what people buy, day-by-day or week-by-week. Such records are primarily obtained from so-called "consumer panels". These are market research operations where all purchases in a specified range of product-classes are continuously measured for a sample of potential consumers.

Informants may be *motorists* when measuring petrol and oil purchasing, or *adults* (including teenagers) in measuring "personal" purchases such as cigarettes or confectionery or, most often, *housewives* (reporting on behalf of all household members) when measuring purchases of food and other household products (soap, polishes, etc.).

The measurement procedures used can vary in detail. Some consumer panels involve weekly or fortnightly calls on each panel-member to record the previous week's purchases, using specially developed measurement procedures plus questioning. More often, each panel-member is required to return by post a weekly record or "diary" in which all relevant purchases are noted by him, together with the size and number of units bought, the price paid, the type of retail outlet, the day of week of purchase and any special "offers" (price cuts, etc.). At their best, consumer panels are sophisticated and well-controlled data-collection procedures. Sample sizes range up to about 5,000 in the U.K. and 8,000 in the U.S. The sample composition of a panel is usually periodically up-dated.*

Continuous consumer panels are mainly run for commercial market research purposes and are open to subscription by manufacturing companies and other interested parties, many millions a year being spent in total on this type of data collection. There appear to be no very detailed descriptions of consumer panel operations in the literature, but a number of writers have given relatively brief accounts [e.g. Ferber 1953, E.I.U. 1963, Day 1970, Moser and Kalton 1971]. Data sources in

*Nowadays there are also so-called "scanner panels", using computerised recording of Universal Product Codes at suitably equipped supermarket check-outs.

the work underlying the present book – and used in the illustrative examples – include the Television Consumer Audit (operated by Audits of Great Britain), and panels run by Attwood Statistics, Research Bureau, and Sales Research Services in the United Kingdom, and by the Chicago Tribune and the Market Research Corporation of America in the United States.

The crucial feature of consumer panel data is that it consists of the continuous purchasing records of the same people or households over extensive periods of time of up to one year or more. Table 1.1 gives a simple illustration of the variable nature of individual purchasing behaviour. It shows the purchases of two brands A and B made by four households over successive weeks. That is the form in which the raw data would be used for analysis – together usually with information about the type of shop visited, the price paid, any special offer involved, any special variation in the flavour or variety or packaging and so on, plus the actual day (rather than merely the week) of purchase.

Table 1.1. An Illustration of Purchasing of Two Brands A and B in Successive Weeks

	Purchases in Week:												
	1	2	3	4	5	6	7	8	9	10	11	12	etc.
1st household	–	–	–	–	A	–	A	A	A	B	–	A	.
2nd household	A	–	A	A	–	–	–	–	–	–	–	–	.
3rd household	–	B	–	–	–	–	–	B	A	–	–	–	.
4th household	–	–	–	–	–	–	–	–	–	–	–	–	.
etc.	.	.	.	.	.	.	.	.	.	.	.	.	.

Collapsing such daily or weekly records into 4-week "months" as in Table 1.2 simplifies the results for visual inspection here. Thus we see that the first household bought nothing in the first month, made three purchases of Brand A in the second month, and two purchases of Brand A and one of B in the third month. The second household made three purchases in the first month and nothing in the next two. And so on.

Table 1.2 still typifies the apparent irregularity of purchasing behaviour at the individual level. It illustrates two further features of such data:

Table 1.2. The Purchasing Data in Table 1.1 Aggregated in 4-Week Periods

	Number of Purchases in Month:				
	I	II	III	etc.	Total
1st household	–	3A	2A, 1B	.	5A, 1B
2nd household	3A	–	–	.	3A
3rd household	1B	1B	1A	.	1A, 2B
4th household	–	–	–	.	–
etc.	.	.	.	.	.
Total	3A, 1B	3A, 1B	3A, 1B	.	9A, 3B

(i) There are large differences between different consumers in their average (or total) purchasing frequency in any given period of time (the first household made a total of 6 purchases in the three months covered, the next two households bought 3 times each, the fourth bought nothing).

(ii) Despite such variations, the *aggregate* levels of purchasing can be more or less steady or "stationary" from period to period. Thus there is no change from month to month, with a total of three purchases of Brand A and one of B each month.

Errors in the Data

In practice, not all panel-members return a completed diary in every single week and such informants have then to be excluded from most kinds of repeat-buying analyses. This is one of several possible sources of statistical error or bias in panel data. There is however evidence that co-operators and non-co-operators do not differ systematically in their purchasing behaviour (cf. §6.4 in Chapter 6), and that increasing length of panel-membership does not produce any changes in purchasing claims [e.g. Ehrenberg 1960; for related technical evidence on long-term panels in television audience measurement see Ehrenberg and Twyman 1966].

Errors of *measurement* also occur, such as incorrect recording or omission of information, or over-reporting [cf. Sudman 1964]. But these errors tend to be less than might be thought. Panel-members become experienced in keeping their records if they continue to co-operate, and it seems that the efficient and satisfying way of filling out a diary each week is to try and do so more or less correctly.

The aggregate data from consumer panels tend to be widely checked by users of the data for systematic bias against other information (e.g.

retail audits, factory shipments, and data from *ad hoc* sample surveys). When bias occurs, it generally affects the aggregate macro-totals (e.g. sales estimates or brand-shares) rather than the internal relationships between different aspects of buying at the micro-level [cf. Kosobud and Morgan 1964]. Errors and discrepancies can certainly occur, but the better-run consumer panels tend to give reasonably accurate representations of real-life purchasing patterns. The data are in fact amongst the most fully checked and reproducible that are available in the social sciences.

One is of course dealing with *reported* purchases, or *claims* to have bought something, and considerable care and vigilance is necessary, as with any massive data-handling exercise. But there is little need here to use the possible occurrence of measurement or statistical errors in the data to explain away *major* discrepancies in the repeat-buying results, as major discrepancies do not generally occur. Instead, since a great variety of purchasing data from many different sources tend to show the same regularities (as is discussed in this book), the role of error or bias in the data can be summarised as follows [Chatfield et al. 1966]:

> "If the purchasing claims more or less represent *actual* purchasing behaviour, we are discussing some empirical regualrities in sociology and marketing, but if they refer to *imaginary* purchases, then we are discussing regularities in psychology and as such they might be deemed the more remarkable."

Apart from data from panel operations which are fully continuous in the sense of aiming to measure all purchases made by a sample of informants over some extensive period of time, repeat-buying can also be studied with data from "dip-stick" surveys of the same informants made at intervals, or even with single interviews in which the informant is asked to recall his past purchasing behaviour over some specified period of time. The danger of measurement error with these techniques is usually much greater than with full-scale panels, but their flexibility and relative inexpensiveness will make them an increasingly important source of data as repeat-buying patterns become better understood. Some early examples are discussed in Chapters 5 and 6.

1.4. The Choice of Analysis-Variables

Two of the basic decisions the consumer may be thought to make are (i) whether or when to buy the given product-class at all, and (ii) if so, what brand (or variety, etc.) to buy. It may therefore seem natural to analyse buyer behaviour likewise – first to analyse repeat-buying for the total product-class and then to follow on with the question of brand-choice – and this has been quite a common approach [for a recent discussion, see for example Massy et al. 1970].

An alternative approach, again a two-step one, is to consider first the purchasing of any particular brand on its own – the subject-matter of Parts I to IV of this book – and only then to start dealing with the question of brand-choice and the relation between buying the brand and the total product-class – where early results are outlined in Part V.

This approach of first looking to the individual brand and only later integrating results for different brands – rather than examining total product-class buying and then breaking it down by brand-choice – has not been taken for *a priori* or intuitive reasons, but because it has worked in the sense of giving the simple and generalisable results set out in this book. The basic empirical finding has been that one can successfully examine repeat-buying of one particular brand (or of one pack-size or variety, etc.) without having to take into account what other brands people may or may not be buying as well. It is by no means obvious that one should be able to do this usefully. The justification is in fact *empirical* – the facts have shown that it works. This is one of the most fundamental empirical discoveries treated here, and one whose explanation or *theoretical* justification (§11.4, Chapter 11) arose only quite recently, ten years afterwards.

Next we need to consider the choice of analysis unit. The most useful unit in which to work has been found to be neither the amount of money paid nor the weight or volume of units bought, but the "purchase occasion". In other words, we have to concentrate our attention on the frequency of purchase. This has turned out to be useful, again in the sense that it has led to a wide range of simple and coherent results. The initial work on repeat-buying was however all carried out in terms of the number of units bought, which related directly to sales volume. This worked well when dealing with a single pack-size of any particular brand. But aggregating different pack-sizes posed major problems, as did products where more than one unit was generally bought on a single shopping trip. These problems were partly dealt with by various "fudges" [e.g. Ehrenberg 1959], but largely by-passed altogether. Gradually there came the realisation that the number-of-units-bought formulation was really inappropriate and that a "purchase occasion" approach would work much better. In many product-fields, a single pack or other "unit" tends to be bought at each purchase, so that the distinction between "units" and "purchase occasions" is trivial anyway. But the purchase occasion is more generally useful because it also allows multi-unit purchases (as in buying several tins of dog food or varying numbers of gallons of petrol) and the aggregation of different pack-sizes all to be dealt with by the same theory, and it also leads to simple results in dealing with multi-brand buying or brand-switching (see also §11.4 in Chapter 11). Sales volumes can still be estimated, by multiplying (at the *end* of an analysis) the number of purchase occasions by the average number of units bought per purchase and by the weight or price per unit.

A third basic orientation in determining an analysis approach is to work in terms of distinct time-periods (such as 1-week, 4-week, or quarterly periods). An alternative is to follow each individual consumer's sequence of purchases, e.g. for the two brands A and B in Table 1.1, the sequence AAAABA, say, for the first household, AAA for the second household, and BBA for the third household, and so on. This purchase-sequence approach has been a very popular one to try out [Lawrence 1966, Sheth 1967, and Massy et al. 1970 have reviewed most of the work here].

One consumer's purchases however quickly get "out of phase" with the next consumers's purchases, and no generalisable results have in fact yet been reported. The justification for using fixed time-periods instead is two-fold, firstly that it has given a wide range of generalisable results and secondly that these are easy to tie in with other marketing data which are measured on a time-period basis (such as sales figures, promotional activity, retail availability, seasonalities, etc.) One particular simplification resulting from this time-period orientation is that most repeat-buying results for any given item (e.g. a brand, a pack-size, a variety, etc.) can be expressed in terms of just two main variables, namely:

Penetration: the proportion of people who buy an item *at all* in a given period, denoted by b.

Purchase Frequency: the average number of times these buyers buy the item in the period, denoted by w.

with w, the average frequency of purchase per buyer, itself being the most basic measure of repeat-buying in the theory*. Sales therefore equal the number of buyers times the average number of purchases per buyer times (as noted earlier) the average number of packs per purchase and the average price or size of the pack.

* The distribution of individual readings about this average is generally of the same form (the LSD – see Chapter 4) and hence this is a statistically good measure to use.

A major finding in the more general study of buyer behaviour (§ 10.2 of Chapter 10) about these two variables b and w is that the penetration of different brands can differ greatly one from the other, whilst their average purchase frequencies tend to be more or less the same. The two variables tend therefore to be relatively independent or uncorrelated. And this applies not only for *all* buyers of a brand, but also to a great many sub-groupings, both in the analysis of repeat-buying and in that of multi-brand buying (Part V). Many analysis problems can be tackled simply by considering the proportions of various relevant sub-groups who buy the item, ignoring (as it were) how often they buy – and this is possible just because the average frequency of purchase per buyer varies little from sub-group to sub-group. This is the kind of simplifying breakthrough which is crucial for *analysis* purposes and which at the same time is of direct *marketing* significance†.

A fourth and final basic orientation in developing an analysis approach is that the theoretical model-building work here generally relates to the "stationary" or "equilibrium" condition. This is defined as the situation where there is no short-term change in the aggregate sales or penetration level of the brand or item in question**. Such stability on the surface however covers highly variable and quite complex patterns of *individual* purchasing behaviour (as was illustrated by Table 1.1), and the focus of this book is on the analysis and suitable aggregation of these individual purchasing patterns.

"Stationarity" in the sense used here – absence of any marked short-term sales trend for the item in question – does not necessarily mean a lack of changing conditions in the market-place (or an absence of trends for other brands), but merely that the sum total of all the varying and dynamic marketing inputs – advertising, pricing, distribution, etc. – has had no overall effect on the sales of the item in question during the relevant time-period. In practice, the degree of stationarity in even the most stationary observed data is usually approximate rather than exact (e.g. a few percent up or down from one period to the next is typical). This should lead to discrepancies between the data and the theoretical models, but the stationary models to be discussed here tend in fact to give a good fit even in such situations where the stationarity is only approximate.

Doubts have often been expressed about the general restriction to the stationary or near-stationary situation. The question raised is

† If the average purchase frequency of different items does not vary greatly, this imposes a major constraint on marketing action.

** The term "stationary" is used here in the specific sense defined, and does not carry overtones from its uses in economics, etc.

whether there is much point in examining "stationary" conditions when good marketing is thought to mean trying to *change* the status quo anyway. However, to dismiss a theory as being purely "academic" because it deals with the stationary purchasing situation would be naive, for two distinct reasons.

Firstly, if one wants to create change, it is as well to understand the stationary no-trend situation from which one wants to depart. And to evaluate what change has been achieved, one must compare the results with what would have happened in the *absence* of change. Many applications of the theory to non-stationary situations have in fact been made and some examples will be given here (starting in §2.4 of Chapter 2).

Secondly, any examination of actual data for frequently-bought consumer goods will show that large trends or big variations in sales are the exception. The sales of most established brands or products are in fact approximately stationary most of the time.

The theory discussed here therefore does not account in direct terms for any dynamic effects of advertising, pricing etc. in changing the level of sales – it is not that sort of theory. But nobody yet knows much about the effects of such varying marketing inputs, and it would be premature to try and "model" what we do not know. Put in another way, the repeat-buying theory tells us more or less all about stationary buyer behaviour except for one thing, namely why one brand has more buyers than another. That is therefore where past and current marketing activities come in and further research is needed, as is discussed in Part VI.

1.5. Some Repeat-Buying Indices

As an introduction to repeat-buying behaviour as such, we now describe several aspects of repeat-buying in terms of a small illustrative example. Table 1.3 sets out the reported purchases of the standard pack-size of Lux by a sample of 983 housewives in a certain 12-week period*. There were 22 buyers of the standard size of Lux in this

* An impression of competitive anonymity is maintained in this case-history by suppressing the time and place, and the precise definition of a "standard" pack-size. "Lux" here is a brand of Soap Flakes, but it could also be liquid detergent, a toilet soap, or another product altogether. In any case, all the main companies who market the product in question obtained the data in Table 1.3 long ago; indeed, anybody could have had the information by suitably measuring a sample of 983 households. In general however, it is necessary in reporting the data to which we have had access to "code" the product-fields and brand-names fully.

sample in the 12-week period, of whom 17 made 1 purchase, 3 made 2 purchases, and 2 made 3 purchases, accounting for 17, 6 and 6 purchases respectively and totalling 29 purchases in all, as is shown in the first ("All buyers") section of the table.

Table 1.3. Buying of Standard Lux in a 12-Week Period

A Worked Example of various Repeat-Buying Indices for a Panel of 983 Households

Standard Size Lux in 12 Weeks		Number of Purchase Occasions in in the period 1 2 3 4+	Total		Average Purchases per Buyer
All Buyers in the given period	No. of Buyers: Total Purchases:	17 3 2 – 17 6 6 –	22 100% 29	 100%	1.3
Those who HAD bought in the previous period	No. of Repeat-Buyers: Total Purchases:	6 2 1 – 6 4 3 –	9 41% 13	 45%	1.4
Those who had NOT bought in the previous period	No. of "New" buyers: Total Purchases:	11 1 1 – 11 2 3 –	13 59% 16	 55%	1.2

The table also shows how these purchases in the given 12-week period break down by whether or not the buyer had bought standard Lux in the preceding 12-week period. Thus the 17 households who made precisely one purchase in the given period are made up of 6 who *had* bought in the previous period and 11 who had *not* bought in the previous period. Similarly, of the 3 people who made two purchases in the given period, 2 were "repeat-buyers" and 1 was a "new" buyer ("new" only in the limited sense of not having bought in the previous period). And so on.

† Multiplying these two measures 2.2% and 1.3 together gives the number of purchases (or sales) in the 12-week period as about 0.03 purchases *per informant*, or 3 purchases per 100 households in the sample.

Such buying behaviour of individual consumers can be summarised by various indices or statistical measures. For a single time-period, two measures are the penetration and the average frequency of buying. The penetration is the percentage of informants who made at least one purchase in the period, i.e. 22 out of the sample of 983 housewives in Table 1.3, or just over 2.2%. The average frequency of purchase per buyer is 29/22 = 1.3†.

Next, the "Totals" column shows that in terms of repeat-buying from the preceding time-period, 41% of the buyers in the period (9 out of 22) had bought the item in the preceding period, and that these repeat-buyers accounted for 45% of total sales in the given period. The last column shows that they had bought in the given period at an average rate per (repeat-) buyer of 1.4 purchases (i.e. 13/9). Similarly, 59% of all buyers in the 12 weeks were "new" buyers, and they accounted for 55% of all sales in the period, at an average rate of 1.2 purchases per "new" buyer.

Another simple index arises if we want to assess the relative importance of the lighter and heavier buyers in the given period. For example, we may ask what percentage of total purchases (or sales) are accounted for by consumers who made at least r purchases, where r may be any whole number greater than or equal to 1. The calculations are illustrated in Table 1.4 and show that households making at least 2 purchases of standard Lux (i.e. two or more) accounted for 41% of total sales (12/29) in the 12 weeks, and that those making 3 (or more) purchases accounted for 21%. It also follows that 59% of sales are taken up by once-only buyers in the period.

Table 1.4. The Cumulative Number of Purchases by Housewives who made at least r Purchases (r = 1, 2, 3 etc.)

Standard Size Lux in 12 weeks	Number of Purchase Occasions in the Period				Total
	1	2	3	4+	
Cumulative Purchases:	29	12	6	–	29
Share of Total Purchases:	100%	41%	21%	–	100%

In the next chapter we start to examine patterns in such data. The question is whether there are in fact any common patterns for different brands, different products, and different lengths of time-periods.

1.6. Summary

Consumer goods are bought for a great variety of needs and under varying conditions. For frequently-bought goods, the act of repeat-buying is a central part of the consumer's response to the product.

Data on individual consumer's purchasing behaviour are mostly obtained from continuous "consumer panels". When well-run, these tend to be amongst the most fully checked and reliable data-sources that are available in the social sciences.

Such data allow one to tabulate repeat-buying results, (e.g. that 60% of the people who bought the item one month bought it again the next month, and how often they bought it then). The approach used is in fact to analyse the individual consumer's frequency of purchasing a given item in specific time-periods. The main model-building emphasis is on the *stationary* situation, where there is no change in the aggregated purchasing-levels of the brand from one time-period to the next. The way in which this leads to simple and generalisable results is discussed in the next chapter.

CHAPTER 2

REGULARITIES OF BEHAVIOUR

2.1. The Fundamental Finding

The fundamental finding in the study of buyer behaviour is that there are simple and highly generalisable patterns. This is by no means an obvious result, given the complexities of the buying situation, and it is illustrated by some examples in §2.2 of the present chapter and in more depth in Chapter 3.

Simple formulae have been developed which successfully summarise or "model" these observed regularities. Such formulae are introduced in cook-book fashion in §2.3, the general theory from which they stem being given in outline in Chapter 4 and in more mathematical detail in Part IV of the book.

The theory applies to the "stationary" situation, defined as there having been no change in the sales-level of the item being analysed. This is the most usual situation for most brands in most markets. Any change in marketing inputs which may have occurred – e.g. a change in price, advertising, distribution, etc. – therefore did not have any *net* effect on the sales of the item and does not enter into the analysis. This stationary theory can however also be used to interpret *non-stationary* situations. An example of such an application is given in §2.4, further practical applications of the theory being described in Chapters 5 and 6 in Part III.

2.2. Regular Patterns

In §1.5 of Chapter 1 we described several indices of repeat-buying for a particular product. Examination of similar data for a wide range of cases has shown that such indices tend to follow regular patterns which are the same for different brands and products and different marketing conditions*.

To illustrate the nature of these generalisations and the range of conditions under which they tend to hold, a varied assortment of 20

* It is primarily by starting from such empirical regularities that the work described here differs from other attempts at building models of buyer behaviour (as are briefly reviewed in Chapter 11).

case-histories is set out in Table 2.1. The cases cover various food and non-food products at different points in time from 1951 to 1963, from various parts of the U.K. or the U.S., and for different length time-periods ranging from one week to half a year (as is shown in the third column from the right). Penetration levels varied from 1% to 50% of the sample buying the brand in question at least once in the analysis-period, and sales-levels varied from 3 purchases to as many as 200 purchases per 100 informants. In all cases sales had been more or less stationary from the preceding period.

The first case in the table – Soap Flakes, Brand A – is in fact that of the standard pack-size of "Lux" in the 12-week period which was already discussed in § 1.5*. The table here shows that just over 2% of the sample bought standard Lux at an average rate of 1.3 purchases each in the 12 weeks, and since buying was "stationary", about 2% of the sample (but not necessarily the same buyers) had also bought standard Lux in the preceding 12-week period, again at an average rate of about 1.3 purchases each.

There is nothing special about the selection of the 20 cases in the table. Many thousands of similar cases have been analysed in more recent years (see for example Chapters 3, 5 and 6) and the major manufacturers of non-durable and semi-durable consumer goods in Western Europe and the United States have spent something like fifty million pounds in the last ten or twenty years on collecting consumer purchasing information and therefore have the same kind of data available on almost innumerable similar cases.

Three of the indices of buyer behaviour which were described in § 1.5 of Chapter 1 are now set out in the central section of Table 2.1, namely:

Sales Accounted for by Repeat-Buyers: The percentage of total sales in the analysis-period accounted for by those buyers who also bought the item in the preceding period (e.g. 45% in the 12 weeks for Soap Flakes Brand A in the first line),

Sales Accounted for by Heavier Buyers: The percentage of total sales accounted for by buyers making at least r purchases in the period, for selected values of r (e.g. 41% and 21% for $r = 2$ and 3 in the first line).

The Average Purchasing Rate per "New" Buyer: The average number of purchases made in the time-period by "new" buyers, i.e. those who had *not* bought the item in the preceding period (e.g. 1.2 in the first line).

Certain patterns can now be seen by visual inspection of the table. This inspection has been facilitated by arranging the 20 cases according

* The fifth case in Table 2.1 is also standard Lux, but for a *24-week* period.

Table 2.1. The Percentage of Total Sales accounted for by Repeat-Buyers from the Preceding Period and by Consumers making r Purchases in the given Period, and Other Statistics

(Approximately stationary sales over two equal time-periods: twenty varied case-histories)

Product	Brand	Place & Year	Average Number of Purchases per Buyer: w	Percentage of Sales accounted for: by repeat-buyers	by buyers of at least r purchases, for r = 2	3	4	6	8	12	16	20	Average Number of Purchases per "new" buyer	Buyers (as % of sample)	Sales: Purchases per 100 Informants	Analysis-Period (in weeks)	Buying of other brands	Marketing activities
Soap Flakes	A	Midl. '65	1.3	45	41	21							1.2	2	3	12		
Clothing	B	U.K. '65	1.4	55	54	17	8	3					1.5	45	63	1		
Flour	C	U.S.'51	1.8	65	67	41	41	13					1.6	21	36	13		
Detergent	D	Lancs. '63	1.8	69	72	39	25						1.3	8	14	4	NOT SPECIFIED	NOT SPECIFIED**
Soap Flakes	A	Midl. '65	1.9	*	67	53	29	16					*	4	8	24		
Detergent	E	Lancs. '63	2.2	83	81	59	40	(6)					1.3	21	45	4		
Soap Flakes	F	Midl. '65	2.5	82	78	70	58	38	19				1.3	4	10	12		
Drink	G	U.K. '55	2.6	76	81	68	55	39	29	19	2		1.6	15	38	13		
Detergent	H	Lancs. '63	2.9	84	88	80	54	(27)	(27)				1.6	14	40	4		
Fuel	I	U.K. '66	3.1	85	89	78	63	(36)	(21)	(6)	(4)		1.8	27	85	4		
Drink	G	U.K. '53	3.3	*	87	76	65	49	37	22	14	11	*	19	63	26		
Detergent	D	Lancs. '63	3.3	78	88	74	67	48	38	(18)			1.7	13	43	12		
Soap Flakes	F	Midl. '65	3.8	*	88	82	74	56	46	32	17		*	5	19	24		
Soap Flakes	J	U.S. '51	3.9	89	90	82	67	62	46	33			1.5	16	65	13		
Margarine	K	U.S. '51	3.9	87	91	84	72	72	53	33	33	15	1.8	26	100	13		
Soup	L	U.K. '58	4.7	*	94	87	82	71	63	50	38	30	*	14	66	26		
Detergent	D	Lancs. '63	4.7	93	94	89	84	67	55	(20)	(3)		1.6	30	140	12		
Fuel	I	U.K. '66	4.8	*	94	88	84	69	58	(32)	(15)	(8)	*	46	221	8		
Detergent	H	Lancs. '63	6.2	94	96	93	90	86	75	47	(26)	(26)	1.7	18	108	12		
Detergent	H	Lancs. '63	10.1	*	98	95	93	91	87	77	64	49	*	22	219	24		

* No information on previous period

** No net effect on sales level, i.e. stationary.

to the size of the average rate of buying per buyer (the first column of figures), this being generally denoted by *w*. Such inspection of the table leads to six main regularities:

(i) Repeat-buyers account for a percentage of sales which increases more or less steadily from 45% to 94% as we go down the table. It therefore increases with increasing values of *w*: the higher the average frequency of purchases per buyer, the higher the percentage in question. (The relationship is non-linear, and its quantitative form will be described later.)

(ii) The percentage of sales accounted for by buyers making at least 2 purchases in the time-period ($r = 2$) also increases steadily, as we go down the table, from 41% to 98%. It is therefore also directly related to *w*.

(iii) Each pair of percentages just discussed, i.e. the percentage of sales accounted for by buyers of at least 2 purchases and that accounted for by repeat-buyers (the two columns of ***bold*** figures), are numerically almost identical, to within an average of 3%, i.e. 45% and 41%, 55% and 54%, 65% and 67%, and so on. This is therefore another very simple regularity in the data*.

(iv) The percentage of sales accounted for by those buyers who made at least 3 purchases in the analysis period also tends to increase with the increasing value of *w*, i.e. 21, 17 (a slight inversion), 41, 39, 53, 59, 70, and so on. Similar relationships with *w* occur for the percentages of sales accounted for by still heavier buyers, e.g. those making at least 4 purchases, and so on.

(v) For relatively high values of *r* – ones equal or greater than the number of weeks in the analysis-period – a departure from the pattern in (iv) sometimes occurs. The observed values in question are shown in brackets and are always *lower* than expected. For example, for the $r = 6$ column, we have the usual pattern of increasing numbers, broken by the bracketed exceptions, i.e. 3, 13, 16, (6), 38, 39, (27), (36), 49, etc. There are therefore *two* empirical patterns, each regular in its own way.

* The near-equality of these two percentages would be trivial if virtually all people who made at least 2 purchases in a given time-period were also repeat-buyers from the previous period (and if none of the once-only buyers in the period had bought the item in the previous one). This however is not so, as was illustrated by the real-life case-history of standard Lux in Table 1.3: of the five people who had made at least 2 purchases of Lux in the 12-week period in question, 2 had *not* bought it in the previous period. (It would have been a very striking regularity if virtually all buyers making more than one purchase in each single period were also repeat-buyers from period to period. But of course this is not so. Only the total numbers of purchases made by these two groups tend to be equal.)

(vi) Turning to the "new" buyers in the analysis-period, we note that the average number of purchases per "new" buyer in Table 2.1 is roughly 1.5 units, within about ± 0.2 on average. This variation is small both in absolute terms and compared with the variation in the average purchase frequency w by *all* buyers (which varies from about 1.3 to 10). The average purchase frequency for "new" buyers in Table 2.1 is in fact virtually unrelated either to w, or to the proportion of informants who bought (i.e. the penetration), or to the length of analysis-period, or to any other specific characteristic of the brand or product-field. In other words, this particular buying rate appears to be more or less constant, a very simple result.

2.3. Repeat-Buying Theory

The natural question raised by the patterns illustrated in Table 2.1 is why they occur. For example, what factors lead to the percentage of sales accounted for by repeat-buyers being 45% in the case of Brand A Soap Flakes (the first line in Table 2.1), and 94% for Brand H Detergents (the last-line-but-one)? Do Soap Flakes generally have fewer repeat-buyers than powdered Detergents, or is there some specific difference between Brands A and H as such, between the two regions of the U.K. involved (the Midlands and Lancashire), or between the two lengths of analysis-period (12 weeks and 24 weeks), or is the difference due to some marketing variable (such as heavier advertising leading either to higher loyalty or to more brand-switching)?

It might be thought that the incidence of repeat-buying will in fact depend on a large variety of factors, such as:

> current or past marketing activities (such as advertising, consumer promotions, distribution, pricing, etc.), the nature of the brand and the product-field in question, the brand's sales level or share of the market, the brand's penetration level, the average rate of buying per buyer, the length of the time-period analysed, the purchasing pattern for competitive brands, the general degree of brand-switching in the product-class, consumer attitudes toward the brand, usage habits, and indeed specific factors such as the particular time and country or region to which the data refer.

Table 2.1 however illustrated not only that regular patterns of stationary purchasing behaviour exist, but also that the variation in the values in question can largely be explained by, or predicted from, one single measure, the average rate w of buying per buyer. This was made self-evident to the eye by the layout of the table: all the repeat-buying indices tended to increase in reading down the figures in any one column, just as did w. And this pattern was not some coincidence for the particular selection of data in the table, but has also generally been found in many thousands of other cases.

It follows that the other factors mentioned above – the brand or product as such, the length of the analysis-period, and so on – have little or no apparent bearing on the results. This is not a consequence of any theoretical considerations or hypothetical assumptions, but a sheer matter of empirical observation and analysis, as is illustrated by Table 2.1. No mathematical analysis has so far been needed – these patterns simply exist, and they can be observed as such. The importance of this finding lies in the wide range of different conditions under which this has been established empirically (as is discussed further in Part II).

To describe the quantitative detail of these relationships it is necessary to model each regularity in some convenient mathematical form. Here we therefore set out some simple formulae in "cook-book" fashion, because all that needs to be noted at this stage is the way in which quite simple formulae do what is required. The formulae all derive from a particular version of the general repeat-buying theory which involves only w, the average frequency of buying per buyer*.

Table 2.2. Values of the LSD Parameter q for Selected Values of the Average Purchase Frequency w

w	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
q	.00	.17	.30	.40	.47	.53	.58	.62	.66	.69
w	2.0	2.2	2.4	2.6	2.8	3.0	4.0	5.0	6.0	8.0
q	.72	.76	.79	.81	.83	.85	.90	.93	.95	.96

To apply this theory we have to start with one relatively complex step, namely the need to calculate a new quantity q. This is directly related to w, the average frequency of purchase per buyer. Table 2.2

* This is the Logarithmic Series Distribution (or LSD) model which is discussed in §4.4 and Chapter 8. It is a simpler version of the more general Negative Binomial Distribution (or NBD) theory which is discussed in §4.2 and Chapter 7, and in which the proportion of buyers of the item also enters into the calculations.

gives the numerical value of q for a selection of values of w. Thus for $w = 1.3$ (as observed for Brand A in the first line of Table 2.1), q is about .40*.

Given the value of q corresponding to each observed w, there are then six theoretical formulae corresponding to the six empirical results (i) to (vi) which were noted for Table 2.1 in §2.2. The background to these formulae is discussed in later chapters as already mentioned, but the formulae themselves are very simple:

(i) Firstly, a formula for the proportion of total sales accounted for by repeat-buyers. This is simply

$$q,$$

(or $100q$ in percentage terms). Thus for an item like Brand A in Table 2.1 with $w = 1.3$ and hence $q = .4$, repeat-buyers from the previous period should theoretically account for .4 (or 40%) of total sales in the current period. This may be compared with the observed value of 45%. In general, these theoretical estimates fit the observed values for the proportion of sales accounted for by repeat-buyers in Table 2.1 to within an average of 3 percentage points**.

(ii) Next, the proportion of sales accounted for by buyers who made at least two purchases in the given time-period. The theoretical estimate for this is also

$$q.$$

The fit for the observed figures in Table 2.1 – the $r = 2$ column – is to within an average of 2 percentage points.

(iii) These two formulae both take the value q and therefore say that the percentage of sales accounted for by repeat-buyers should in theory

* The quantity q is the single parameter of the LSD model. The relationship between w and q is $w = -q/(1-q)\ln(1-q)$, where "ln" stands for the "Napierian" or "natural" logarithm to base e (a table of natural logarithms is given in Appendix B). The equation gives w in terms of q, but it cannot be written the other way round. In other words, it is algebraically impossible to express the unknown quantity q directly in terms of the known quantity w. Hence we give a table such as Table 2.2 from which q can be read off. For routine use, a more extensive table of this kind is set out in Appendix B, or q can be calculated by iteration. For w greater than 2, it is also possible to use the approximate formula $q \doteq (w-1.4)/(w-1.15)$: this holds to within $\pm$ 0.01, which is accurate enough for many practical purposes.

** Some of the differences may be statistically significant (even though the sample sizes in the table are often fairly small). But as in most of the work in this book, the important point here is not so much whether any of the discrepancies are real (or merely sampling errors), but that the same theoretical formula accounts for much of the greater part of the observed variation, and that the remaining deviations are relatively small and more or less unbiased. Some sampling error formulae are discussed in §6.4.

equal that accounted for by "more-than-once" buyers. This was found to be so for the observed data in Table 2.1 to within an average of 3 percentage points (as already noted as item (iii) in §2.2).

(iv) The theoretical formula for the proportion of sales accounted for by buyers who make at least r purchases in the period is

$$q^{r-1}.$$

This is a more general result corresponding to the expression q in (ii), i.e. q^{r-1} with $r = 2$. This formula holds for all the observed values given in Table 2.1, to within an average of 3 percentage points (other than for the values in brackets).

(v) The bracketed values in Table 2.1 are exceptions to the pattern in item (iv). Comparison with the theoretical formula q^{r-1} shows that these bracketed values are all lower than q^{r-1}, as already noted empirically in §2.2, so that the theoretical values serve to show up this particular sub-pattern as being itself systematic*.

(vi) Finally we have the theoretical expression for the average frequency of purchase per "new" buyer. The theoretical formula for this is $q/\ln(1+q)$. The numerical value of this expression varies very little for different values of q when w is greater than about 2, so that the "new" buyers' theoretical average frequency of purchase may in fact be taken as

$$1.4,$$

a virtual constant. This agrees with the observed values in Table 2.1 to within an average of about .2.

The significance of these theoretical formulae is three-fold. First and foremost, they succesfully describe the observed results to a close degree of approximation (as can readily be checked in detail by using the value of q given in Table 2.2 for the observed value of w of each item in Table 2.1). Second, the formulae are very simple, in that they all turn on one single statistic q, which is itself uniquely determined by w, the observed average frequency of purchase among all buyers. Third, they are not merely "best-fitting" isolated formulae for various sets of data, but are all inter-related. This is clear from their common formulation in terms of the quantity q, and is made explicit in the LSD theory from which the formulae stem.

* This is the so-called "shelving" or "variance discrepancy" phenomenon which is discussed in §7.8; it is of relatively little practical importance at this stage, although it has basic theoretical implications.

Many additional aspects of repeat-buying under stationary conditions can also be succesfully deduced from the LSD theory and from the more general NBD theory. Examples are formulae for the number of repeat-buyers, for their average frequency of purchase, for the number of "new" buyers (whose average purchase frequency $q/\ln(1+q)$ has already been given in (vi) above), and so on. These additional results and their theoretical background will be developed in Chapters 4, 7 and 8, together with certain simplifying approximations. Here we now turn to illustrate a practical application.

2.4. A Practical Application: A Seasonal Trend for Brand M

We now illustrate the use of the repeat-buying theory in a simple case-history which turns on evaluating a *non-stationary* situation, namely a change in the sales level of a certain Brand M. The sales increase in question was due to a recurrent seasonal trend, i.e. it was due to natural causes.

The problem was not to establish whether or not there had been a sales increase – that was clear from the aggregate data. Thus Table 2.3 shows how the purchasing level of Brand M had increased during the peak-season quarter to 48 purchases per 100 households from a level of 36 purchases in the preceding off-season quarter. The latter is the "stationary norm" for the aggregate sales in the peak quarter, i.e. what the sales would then have been if there had in fact been no increase.

Table 2.3. The Seasonal Trend in the Aggregate Level of Purchasing Brand M

Peak-Season Quarter	Observed	Stationary Norm*	Difference
Purchases (per 100 households)	48	36	12

* As in off-peak quarter.

The problem was therefore to understand the *nature* of the increase. Were the extra peak-season sales due to there being more repeat-buyers, to their buying more, to there being more "new" buyers, to *their* buying more, or to some combination of these various possibilities?

Table 2.4 gives two breakdowns of the observed data (based on a sample of about 2000 households). Firstly, it shows that the peak-season sales of 48 purchases per 100 households were made up of 16%

of the sample buying on average 3 times each in the quarter, whereas the 36 purchases per 100 households in the off-season quarter came from 12% of the sample buying on average 3 times each. The sales increase was therefore due to an increase in the number of households buying the brand, and not due to any change in the average rate of buying. That is one finding.

Secondly, the peak-season buyers are classified in the table as either "repeat-buyers" or "new" buyers, according to whether or not they had also bought Brand M in the preceding (off-season) quarter. The table shows that there was an equal division into 8% of the sample being repeat-buyers (who on average made 4 purchases each) and 8% "new" buyers (who on average made 2 purchases each). The total sales of 48 purchases per 100 households were therefore made up of 32 purchases by repeat-buyers and 16 purchases by "new" buyers.

This latter tabulation however does not show whether the sales increase of 12 purchases per 100 households from the previous period came more from abnormally high repeat-buying or more from an abnormally high influx of new buyers. To answer this, we need to know how many repeat-buyers and how many "new" buyers there would have been (and how often each group would have bought) if there had been no seasonal trend. In other words, we have to compare the observed repeat-buying behaviour with the pattern to be expected if there had been *no* trend, i.e. with the theoretical norm of *stationary* repeat-buying behaviour.

Now the findings which have been illustrated in §3 (and which are elaborated in Parts II and IV) are that stationary repeat-buying should depend only on the observed buying behaviour in the *preceding* period. Thus given the average rate of buying of $w = 3$ purchases per buyer observed in the off-peak quarter (see Table 2.4), the no-trend NBD/LSD estimate is that about two-thirds of all the buyers of Brand M in that quarter would have been repeat-buyers (i.e. 8% of the total sample), and that they would on average have bought 4 times each in the quarter*. This is shown in Table 2.5 and agrees exactly with the observed results for repeat-buyers. The seasonal sales trend therefore did not affect the repeat-buying of previous off-peak buyers, either in the *number* who were repeat-buyers or in their average *rate* of buying. If there had been no sales increase or no "season", repeat-buyers would have accounted for about 32 purchases per 100 households in the peak-season (8% times 4 purchases), just as was observed when the trend did actually occur.

* The actual calculations here were based on the NBD version of the theory.

Table 2.4. Observed Purchasing in the Peak-Season Quarter broken down into Repeat-Buyers and "New" Buyers

Peak-Season Quarter	Observed		Stationary Norm*		Difference
All Buyers	16% of the sample buying at av. rate of 3 purchases	= 48	12% of the sample buying at av. rate of 3 purchases	= 36	12
Repeat-Buyers HAD bought in previous quarter	8% of the sample buying at av. rate of 4 purchases	= 32		?	?
"New" Buyers had NOT bought in previous quarter	8% of the sample buying at av. rate of 2 purchases	= 16		?	?

* As in off-peak quarter.

Table 2.5. Observed Repeat-Buying Compared with Stationary NBD Norms

(All observed and theoretical figures rounded to the nearest whole number for expository simplicity)

Peak-Season Quarter	Observed		Stationary Norm*		Difference
All Buyers	16% buying at av. rate of 3 purchases	= 48	12% buying at av. rate of 3 purchases	= 36	12
Repeat-Buyers HAD bought in previous quarter	8% buying at av. rate of 4 purchases	= 32	8% buying at av. rate of 4 purchases	= 32	0
"New" Buyers had NOT bought in previous quarter	8% buying at av. rate of 2 purchases	= 16	4% buying at av. rate of 1 purchase	= 4	12

* Predicted from off-peak quarter.

The seasonal increase of 12 purchases per 100 households was therefore all due to "new" buyers. The observed incidence was that 8% of the sample were "new" buyers, making an average of 2 purchases each and accounting for an aggregate sales level of 16 purchases per 100 households. This compares with the theoretical no-trend incidence of

4% new buyers making about 1 purchase each (the theoretical norm of 1.4 rounded to the nearest whole figure), and thus accounting for about 4 purchases per 100 households. The difference between the observed and theoretical sales figures of 16 and 4 therefore pinpoints the observed sales increase of 12 purchases.

It follows that the market for Brand M was segmented into two kinds of buyers. There were all-the-year-round buyers whose repeat-buying from the off-peak to the peak-season was in no way affected by the seasonal trend, and peak-season-only buyers – the *extra* 4% "new" buyers – who did not buy in the off-season, and who bought at an above-normal rate of almost 3 purchases each*.

2.5. Summary

The nature of repeat-buying could vary with a lot of different factors. It might be different for market-leaders than for smaller-selling brands, it might depend on the number and the popularity of *other* brands or pack-sizes or varieties available, it might vary with the nature of the product (its usage patterns, retail distribution, etc.) or with the weight and nature of promotion (advertising, consumer deals, etc.). However, it is found empirically that under stationary no-trend conditions, none of these factors matter explicity. Instead, various indices of repeat-buying follow regular patterns which generalise across a wide range of brands, products, time-periods and other conditions. The numerical values in question all vary primarily with the average frequency of purchase per buyer, w.

It follows that theoretical formulae to "model" these regularities must all be interrelated. This is illustrated in §2.3 by some simple formulae from the repeat-buying theory which is described in more detail in Parts II and IV.

This is a theory of *stationary* repeat-buying but it can also be used to interpret *non-stationary* situations. An example is given in §2.4, where a seasonal sales increase is found to be due to extra buyers coming into the market during the peak-season, the all-the-year-round buyers being quite unaffected by the seasonal uplift.

* Half the observed new buyers in Tables 2.4 and 2.5 are *normal* new buyers, i.e. all-the-year-round buyers who would buy relatively infrequently, on average about 1.4 times each in any quarter in which they buy at all. The additional peak-season-only new buyers had therefore to be more heavy buyers – buying on average three times each – so as to bring the *overall* average frequency of purchase to the value of 2 actually observed.

PART II

REPEAT-BUYING

CHAPTER 3

THE REPEAT-BUYING STRUCTURE OF A MARKET

An Empirical Example

The purpose of Part II of this book is to try and describe the nature of repeat-buying theory for the general reader. The aim is not to propound an abstract theory but to help in describing and understanding empirical buyer behaviour itself – the forms which it takes and the factors on which it depends. That is where the theoretical work started and where its applications lie.

In the present chapter we describe the theory's major practical application, which is to help structure the observed facts of buyer behaviour in any given market. More specifically "problem-solving" orientated studies are illustrated in Part III. The theoretical results of the NBD/LSD model are outlined in Chapter 4, and the numerical calculations are described in Appendix A*.

The analysis in the present Chapter is closely based on a standardised type of report which has been used in the comprehensive analysis of many different product-fields in the last few years [e.g. Aske Research 1970]. The variety of product-fields covered so far in those intensive studies range from Instant Potatoes, Frozen Foods, Breakfast Cereals and Soup, through Confectionery, Cigarettes (which is rather different), "Take-home" Beer, Dentifrice and Detergents, to Petrol and Motor Oil. The results for one particular product-field are set out here in "coded" form. This product-field is typical not only in that the observed buying patterns mostly follow the theoretically predicted lines rather closely, but in that there are also a number of specific exceptions.

The analysis centres on repeat-buying and penetration growth for five leading brands, labelled A to E, during a 48-week period (i.e. effectively a year). The results are set out in ten tables or groups of tables, and commentary. The purchasing behaviour of each brand is studied here in isolation from that for all other brands, and corresponding results for "multi-brand" buying, relating purchasing of the different brands to each other, are given in Chapter 9.

* The theoretical calculations in this chapter are based on the NBD version of the theory.

The distinctive approach is that the analyses presented are already well understood so that the results usually follow simple and predictable patterns (as shown by the NBD/LSD theory which is set out in Chapter 4). This allows us to see which of the findings are normal and which are exceptional. A wide range of tabulations can therefore be reduced to manageable proportions and interpretation can become more positive. The main structure of repeat-buying behaviour in this product-field is seen to be normal, and the exceptional features which may deserve special attention are readily isolated.

Table 3.1 Penetration and Penetration Growth

For branded consumer goods, the penetration of a brand is usually the feature which most clearly differentiates one brand from another in terms of sales level or market-share (as given in the first column of figures in Table 3.1). The "penetration" of an item is the proportion of the population who buy the item at least once in a given time-period. This is examined in the two facing tables for the total product-class ("Any Brand") and the five leading brands A to E (which account for 75% of the market).

In Table 3.1 we examine how many households buy in each of the four 12-week quarters of the 48-week "year" analysed here*. We see that the penetrations of the total product and of each of the five individual brands are relatively stable quarter by quarter. The variation is statistically significant but slight, and mainly reflects some seasonality in this market (more buyers in Quarters II and III)**. Such movements could form the basis of more detailed investigations (as was illustrated in the seasonal analysis in §2.6 of Chapter 2), but in the first instance the relative stability of the quarterly penetrations means that the "Average" column is a fair approximation to the penetration in any one quarter. Thus *about* 62% bought the product in each quarter, *about* 42% bought Brand A, *about* 17% bought Brand B, and so on***.

* In this and subsequent tables, results are usually set out to at most two significant figures. (An additional figure may occasionally be needed in detailed calculations.)

** On a 4-figure sample base, the standard error of an individual quarterly penetration is of the order of 1 percentage point, and with a continuous panel, the sampling error of a trend is reduced (see also §6.4 and Appendix A).

*** The fact that the values of the quarterly penetrations closely resemble the market-share values of the five brands is a numerical coincidence for *quarterly* periods; it does not occur in other length time-periods (see Table 3.1a).

Table 3.1. Penetration

% of Households Buying in each Quarter

	% Market Share* (Annual)	QUARTER I	II	III	IV	Av.
		%	%	%	%	%
ANY BRAND	100%	58	63	65	61	62
Brand A	46	39	45	44	41	42
Brand B	12	15	17	18	16	17
Brand C	6	7	8	8	6	7
Brand D	5	5	6	8	6	6
Brand E	6	7	6	7	7	7
Average Brand	15	15	16	17	15	16

* In terms of purchase occasions, with Brands A to E accounting for 75% of the product-class ("Any Brand").

Table 3.1a. Penetration

% of Households Buying in Periods of Various Lengths

(Observed Values "O" and Theoretical Norms "T")

	Period of length (in weeks) 1 O	1 T	4 O	4 T	12 O	12 T*	24 O	24 T	48 O	48 T
ANY BRAND	22	**19**	45	**42**	62	**(62)**	74	**72**	79	**80**
Brand A	12	**10**	28	**26**	42	**(42)**	55	**52**	62	**61**
Brand B	3	**3**	9	**9**	17	**(17)**	25	**22**	32	**28**
Brand C	2	**1**	4	**4**	7	**(7)**	12	**10**	17	**13**
Brand D	1	**1**	3	**3**	6	**(6)**	11	**8**	14	**11**
Brand E	2	**1**	4	**4**	7	**(7)**	9	**9**	12	**11**
Average Brand	4	**3**	10	**9**	16	**(16)**	22	**20**	27	**24**

* Used in fitting.

These quarterly averages are shown again in Table 3.1a in the context of penetration growth, i.e. penetration figures for time-periods of lengths ranging from 1 week through 12 weeks to the full 48 weeks analysed here. Thus 22% of the population bought the product in a typical week, 62% in the average quarter as already noted, and 79% in the full 48-week year. Similarly, for Brand A the penetration growth is from 12% in a week to 62% in the year. And so on.

These observed figures ("O") are compared with theoretical norms ("T") which are predictions from the 12-week penetration and the corresponding 12-week buying frequency (given in Table 3.2a), the calculations being based on the NBD/LSD theory (see § 4.8 and Appendix A for numerical details).

These norms allow us to assess the penetration growth of each item against what *normally* happens. Table 3.1b for example gives the results for the average brand in the last six product-fields analysed in this way at the time of writing, and the fit clearly is good. In the case of Table 3.1a, given that 62% of the population buy the product-field in the typical quarter and that they do so on average 5 times during that quarter (see Table 3.2), the prediction is that 19% of them will buy the product in the average week and that 80% will buy during the "year". For Brand A the predictions are that about 10% should buy Brand A in a week and 61% in the year. And so on.

Table 3.1b. Typical Penetration Growth: The Average Brand in 6 Product-Fields

% of Households Buying in Periods of Various Lengths

(Observed Value "O" and Theoretical Norms "T")

Six Product-Fields	Period of length (in weeks) 1		4		12		24		48	
	O	T	O	T	O	T*	O	T	O	T
The Average Brand	3	**2**	6	**6**	11	**(11)***	17	**16**	22	**21**

* Used in fitting.

The agreement between the predicted penetration figures and the observed ones is mostly close. (For 24 predictions ranging from 1% to 80%, the average discrepancy is about 1½ percentage points, and only 2 values are out by more than 3 percentage points.) This agreement shows the extent to which penetration growth from a week to a year largely follows a predictable or "normal" pattern. There is therefore little in the longer-term (e.g. annual) penetration figures which is not already implicit in (or predictable from) the shorter-term buying patterns. The amount and complexity of the information that has to be considered is therefore greatly reduced.

The exceptions are not only fairly small but also rather systematic. They occur mainly for Brands B, C and D. These have penetrations in 48 weeks which are several percentage points higher than is predicted from the shorter-term buying patterns. The reason is thought to be the seasonality of the market (noted in Table 3.1) which has brought in some extra buyers for Brands B, C and D. This does not occur for the product-field as a whole, so that the implication is one of greater brand-switching for these three brands during the peak-season (a specific kind of finding which can be followed up in more detail by further analysis). The brand-leader A however gained virtually no extra buyers in this way. Brand E also shows virtually no excess annual penetration but Table 3.1 also showed no seasonal trend for this brand; Brand E is in fact somewhat "different" in a number of other respects as well, as will be seen in subsequent tables.

Table 3.2 The Frequency of Buying per Buyer

In Table 3.2 we turn to the average rates of buying per quarter, i.e. the number of occasions on which the average buyer of an item in a given quarter bought the item during that quarter. There is little variation quarter by quarter, and the "Average" column provides a good approximation to the buying frequencies in any one quarter.

In examining the relationship between the average purchase frequency of different brands, a general finding is that the average buying frequency does not differ greatly from one brand to another (as is discussed more fully in § 10.2 of Chapter 10). Typically this also occurs here: on average about 3 purchases of a brand are made per quarterly buyer of the brand, irrespective of the fact that the penetration levels or market-shares of Brands A to D or E differ by a factor of almost 10. This uniformity in buying rates imposes a major constraint on marketing action – it is unlikely that the existing buyers of a brand will turn (or can be turned) into much heavier buyers of it: any *major* change in average purchase frequency would be out of line with the general pattern.

Specific exceptions to this regularity occasionally arise. These are sometimes due to the "growing pains" of new brands, or relate to some physical segmentation of the product-field. More generally however, this apparent "constancy" of the average purchase frequencies is in fact not quite the whole story. There tends to be a trend, usually quite a small one, with penetration level: the more people there are who buy a brand the more often (or *slightly* more often) they tend to buy it. The trend is not very clear in Table 3.2 but the brand-leader with an average purchase-frequency of 3.7 certainly stands out.

Table 3.2. The Frequency of Buying the Brand

The Average Number of Purchases of a Brand in each Quarter per Buyer of the Brand in that Quarter

	% Market Share* (Annual)	QUARTER I	II	III	IV	Av.
ANY BRAND	100	4.9	5.1	5.0	5.1	5.0
Brand A	46	3.7	3.8	3.6	3.7	3.7
Brand B	12	2.5	2.4	2.5	2.5	2.5
Brand C	6	3.1	3.0	2.6	2.7	2.9
Brand D	5	2.4	2.6	2.2	2.8	2.5
Brand E	6	2.7	3.2	3.2	3.1	3.0
Average Brand	15	2.9	3.0	2.8	3.0	2.9

* In terms of purchase occasions.

Table 3.2a. The Frequency of Buying the Brand

The Average Number of Purchases per Buyer, in Periods of Various Lengths

(Observed Values "O" and Theoretical Norms "T")

	Periods of length (in weeks) 1		4		12		24		48	
	O	T	O	T	O	T*	O	T	O	T
ANY BRAND	1.16	**1.39**	2.3	**2.4**	5.0	**(5.0)**	8.8	**8.6**	15.7	**15.7**
Brand A	1.02	**1.26**	1.8	**2.0**	3.7	**(3.7)**	6.0	**6.0**	10.1	**10.3**
Brand B	1.01	**1.14**	1.5	**1.5**	2.5	**(2.5)**	3.3	**3.6**	5.0	**5.8**
Brand C	1.01	**1.19**	1.7	**1.7**	2.8	**(2.8)**	3.9	**4.2**	5.1	**6.7**
Brand D	1.01	**1.17**	1.5	**1.6**	2.5	**(2.5)**	3.1	**3.7**	4.3	**5.7**
Brand E	1.01	**1.21**	1.6	**1.7**	3.0	**(3.0)**	4.9	**4.6**	6.8	**7.4**
Average Brand	1.01	**1.19**	1.6	**1.8**	2.9	**(2.9)**	4.2	**4.4**	6.3	**7.2**

* Used in fitting.

This kind of trend normally takes the form that it is the average purchase frequency per buyer multiplied by the proportion of *non-buyers* which does not vary greatly from brand to brand (see §10.2 of Chapter 10 and §§11.4 and 11.5 in Chapter 11). The effect is quite marked in longer time-periods, as is discussed overleaf, but when penetrations are low (as in the quarterly periods here other than for Brand A

– see Table 3.1), the "correction-factor", i.e. the proportion of non-buyers, is itself close to unity and is therefore almost negligible. Thus for the average quarterly results in Table 3.2 we have, using the complement of the percentage buying figures in Table 3.1, that

Average Quarterly Purchase Frequency

Brand	Observed	× Non-buyers
A	3.7	2.1
B	2.5	2.1
C	2.8	2.6
D	2.5	2.3
E	3.0	2.8
Average	2.9	2.4

The variability of the figures is generally reduced (from a range of 1.2 – i.e. 3.7 to 2.5 – to one only half as big at .7).

In particular, the high purchase frequency of Brand A is seen to be accounted for by its high penetration – it is nothing intrinsic to Brand A as such. Brands C and E however now stand out a little as having attracted a somewhat high frequency of purchase amongst their buyers, and these exceptions merit further attention. For Brand C the explanation lies with the quarter-by-quarter trend in the figures of Table 3.2, it being the results in the first two quarters which are too high. The picture for Brand E is that this brand in fact differs in product-formulation from the other brands (rather as in the U.K., Ribena or Lucozade differ from other soft drinks, or All Bran or Stergene differ from other cereals or detergents). We begin to see to what extent such a difference in potential product-appeal is reflected in actual consumer response: Brand E has no seasonal trend (see Table 3.1) and attracts somewhat more frequent buyers, but the difference is not vast.

In Table 3.2a we turn to the average buying frequencies per buyer in periods of varying length. The differences between brands just discussed become more marked in the longer periods such as the 48-week one, but in the main (i.e. other than the Brand E effect), they remain predictable simply in terms of the relation with the proportion of non-buyers, as can readily be checked by using the penetration figures in Table 3.1a.

Another general feature of the average purchase frequencies shown in Table 3.2a is that they increase less than "pro rata" with the increasing length of the analysis-period. This is because extra buyers in a longer period are generally lighter (or less frequent) buyers – they buy in the *longer* period but not in the initial shorter period. This process tends to follow a general pattern. Theoretical NBD norms can therefore be calculated from the quarterly results and compared with the observed figures. Thus given that the quarterly buyer of Brand A bought it on average 3.7 times during the quarter, the prediction is that the larger number of buyers in the full year will buy the brand on average 10.3 times during the year.

As in the comparison with the penetration norms in Table 3.1a, the agreement between the observed and theoretical buying rates in Table 3.2a tends to be close. Looking at the 48-week figures for example, the observed and theoretical figures are identical for the product-field as a whole and almost so for the leading brand. For the other brands, the observed figures are somewhat lower than the theoretical norms, which ties in with their above-normal annual penetrations in Table 3.1a. The implication is mostly that there were extra "seasonal" buyers for each of these brands, and that these bought less often than normal (i.e. in the peak-season only).

In the shorter time-periods, especially the single week ones, the theoretical norms consistently overstate the observed figures. This is a general finding, namely that in short time-periods buying patterns are different from those in longer periods, and hence the NBD/LSD extrapolations do not apply, as discussed more fully in Chapter 4 and in Part IV.

Table 3.3 Packs per Purchase

The simplest results in examining buyer behaviour are generally obtained by working in terms of *purchase occasions*, as already mentioned in §1.2 of Chapter 1. Table 3.3 provides the link with *sales*, showing the number of packs bought per purchase occasion*.

With most products that have been studied, only a single pack is purchased on most occasions. This is the case with all the brands in

* The decomposition of the sales of an item in a given time-period is into the number of buyers in the period *times* the average number of purchases per buyer *times* the average number of packs per purchase (*times* the average size or value per pack). This decomposition is valuable in as far as none of the components other than penetration vary very much from brand to brand.

Table 3.3, and does not depend on whether people are frequent or infrequent buyers of the product-group. (In some product-fields, several units tend to be bought per purchase, e.g. gallons of petrol, and analyses as in this table tend then to carry more information.)

In some markets (such as the present one), each brand is sold in several pack-sizes. Detailed analyses of penetration growth and frequency of purchase for the different sizes are then also needed. However, it is a basic finding that repeat-buying loyalty can be successfully examined for each brand as a whole (i.e. irrespective of pack-size), if the *purchase occasion* is treated as the analysis unit, as in this chapter. The theoretical reason for this is discussed in § 11.5 of Chapter 11 and largely stems from the empirical finding (§ 10.3, Chapter 10) that average rates of buying a pack-size per buyer of that size vary little from size to size.

Table 3.3. Packs per Purchase

The Average Number of Packs bought per Purchase Occasion

		Buyers of Product *		
48 WEEKS	ALL	Light	Medium	Heavy
ANY BRAND	1.05	1.03	1.03	1.08
Brand A	1.05			
Brand B	1.03			
Brand C	1.13	Not calculated		
Brand D	1.04			
Brand E	1.04			

* Light = 1 – 12 purchases of the product per year (55% of all buyers).
Medium = 13 – 25 purchases of the product per year (25% of all buyers).
Heavy = 26 + purchases of the product per year (20% of all buyers).

Table 3.4 Light and Heavy Buyers

The purchasing rates referred to in Table 3.2 were *averages*. In Table 3.4 we therefore consider the way in which individual purchase frequencies are distributed about these averages, over the 48-week period analysed here. (Similar results are obtained in shorter time-periods.)

Of all the households buying Brand A during the year, 18% bought only once, 9% twice, and so on, 52% buying six or more times during the year. And so on for other brands.

These observed frequencies are compared in the table with theoretical NBD norms which are calculated from the percentage of buyers of the brand in the period and their average frequency of purchase. Thus, given that 62% of the population bought Brand A at an average frequency of about 10 purchases in the year (as shown in Tables 3.1a and 3.2a), we predict that about 19% of these buyers will prove to be once-only buyers, 12% to have bought twice, and so on.

The agreement for the various brands is generally close. Something of a discrepancy however occurs for the product-field as a whole. This is often found for products which most of the population buy (here 79%

Table 3.4. Light and Heavy Buyers

The % of Buyers Making 1,2,3,.... Purchases in the Year

(Observed Values "O" and Theoretical Norms "T")

48 WEEKS	% of Population buying at all (T3.1a)		Number of Purchases in the Year 1	2	3	4	5	6+	Average*
ANY BRAND	79 = 100%	O:	8	8	6	4	4	70	16
		T:	**12**	**9**	**7**	**6**	**5**	**62**	
Brand A	62 = 100%	O:	18	9	8	7	6	52	10
		T:	**19**	**12**	**9**	**7**	**6**	**49**	
Brand B	32 = 100%	O:	32	18	11	6	5	29	5
		T:	**31**	**17**	**11**	**8**	**6**	**28**	
Brand C	17 = 100%	O:	38	16	13	7	3	23	5
		T:	**33**	**16**	**10**	**7**	**6**	**28**	
Brand D	14 = 100%	O:	45	20	9	3	0	23	4
		T:	**36**	**17**	**11**	**7**	**5**	**24**	
Brand E	12 = 100%	O:	36	12	6	5	6	36	7
		T:	**30**	**15**	**10**	**7**	**5**	**34**	
Average Brand	27 = 100%	O:	34	15	9	6	4	33	6
		T:	**30**	**15**	**10**	**7**	**6**	**33**	

* Average Number of purchases per buyer (Table 3.2a).

buying in the year): buying is somewhat more regular for the product than for individual brands (or for the NBD model). There is here an excess of frequent buyers (e.g. 6+ times in the year) compared with the norm, but the difference is never very large.

Another small discrepancy occurs for Brands C, D and E. These show an abnormally high proportion of once-only buyers. This suggests that there was a higher than normal level of switching between brands on a very *occasional* basis (i.e. during the peak-season, as noted earlier).

Brand E was earlier noted to be a brand with an unusually high average frequency of purchase, for its penetration level. Table 3.4 now shows that this is not due to any particular excess of very heavy buyers as such. Other than the slight excess of *light* buyers just noted, the distribution of purchasing frequencies about the average of 7 is in fact closely predictable along the standard lines.

Table 3.5 The Sales Importance of Light and Heavy Buyers

Having examined the numbers of lighter and heavier buyers of each brand, we turn in Table 3.5 to the proportion of the total sales of the brand which they account for.

Table 3.5. The Sales Importance of Light and Heavy Buyers

The Percentage of the total Purchases of an Item going to People Who Bought the Item Once, Twice, Three times etc. in the Year

(Observed Values "O" and Theoretical Norms "T")

48 WEEKS	% Market share (Purchases)		Number of Purchases in the Year 1	2	3	4	5	6+
ANY BRAND	100 = 100%	O:	0.5	1.0	1.2	1.0	1.4	94.9
		T:	1.5	1.5	1.5	1.5	1.4	92.6
Brand A	46 = 100%	O:	2	2	2	3	3	88
		T:	2	2	3	3	3	88
Brand B	12 = 100%	O:	7	7	6	5	5	70
		T:	6	7	7	6	6	68
Brand C	6 = 100%	O:	7	6	8	6	3	70
		T:	6	6	6	6	5	71
Brand D	5 = 100%	O:	10	9	6	3	0	72
		T:	8	8	7	7	6	64
Brand E	6 = 100%	O:	5	4	2	3	4	82
		T:	4	5	4	4	4	79
Average Brand	15 = 100%	O:	6	6	5	4	3	76
		T:	5	6	5	5	5	74

For the product-field as a whole, the 30% of buyers who bought less than six times in the year (Table 3.4) account for only about 5% of sales.

For the average brand, something like 80% of its annual sales are accounted for by the heavier buyers, i.e. those who made six or more purchases of the brand in the year (about 25-50% of its buyers – Table 3.4). These figures agree closely with the corresponding NBD norms.

The ability to predict such figures from theory illustrates the extent to which the sales structure of a market is understood. Such use of theory can be quicker and cheaper than fresh tabulation of raw data. For example, similar results for periods shorter than a year can also readily be estimated. However, the more important outcome is the ability to *interpret* the observed data by judging which figures are normal and which require special attention.

Table 3.6 The Incidence of Repeat-Buyers

People who buy an item more than once in a given time-period are repeat-buyers, and the average frequency of purchase considered in Table 3.2 was therefore a measure of repeat-buying for that time-period. We now turn to another form of repeat-buying, namely the incidence of repeat-buying from one time-period to another. This is a particularly powerful form of analysis, as has already been illustrated in the practical application to a seasonal trend given in §2.5 of Chapter 2.

Table 3.6 sets out the incidence of repeat-buyers quarter by quarter, both for the product-field as a whole and for the five leading brands within it. Thus, of the 58% of the population who bought the product in the first quarter (see Table 3.1), 87% bought it again in the second quarter. Similarly, 87% of the second-quarter buyers of the product bought it also in the third quarter, and so on. In the absence of marked trends, the incidence of repeat-buyers should be the same for different pairs of quarters. The slight seasonality seen in Table 3.1 produces a fractionally lower repeat-buying level between Quarters III and IV, but despite this the average column in Table 3.6 reflects the general quarterly repeat-buying levels quite well.

Of themselves, such repeat-buying figures do not have a clear meaning. Thus for Brand A, something like 78% of those who bought it in one quarter bought it again in the subsequent quarter, but as already noted in Chapter 1, the question is whether such a repeat-buying level is high or low, good or bad.

Table 3.6. The Incidence of Repeat-Buyers

The Percentage of Buyers of an Item in One Quarter who bought it again in the Next Quarter

	QUARTERS			
	I/II	II/III	III/IV	Av.
ANY BRAND	87	87	81	85
Brand A	84	77	73	78
Brand B	61	58	58	59
Brand C	49	57	45	50
Brand D	55	58	46	53
Brand E	65	73	69	69
Average Brand	63	65	58	62

Table 3.6a. The Incidence of Repeat-Buyers

Percentage of Buyers in One Period Who bought again in the Next Period for Periods of Various Lengths

Observed Values "O" and Theoretical Norms "T")

	Period of Length (in weeks):							
	1		4		12		24	
	O	T	O	T	O	T	O	T
ANY BRAND	45	na	76	71	85	84	91	89
Brand A	40	na	69	58	78	77	83	84
Brand B	32	na	46	47	59	65	62	73
Brand C	37	na	52	52	50	67	52	73
Brand D	20	na	58	45	53	64	54	70
Brand E	34	na	65	48	69	69	63	75
Average Brand	33	na	58	50	62	68	63	75

na = Theory not applicable in very short time-periods.

To reach an answer, we compare the observed figures with what normally occurs, as reflected by the theoretical NBD norms. These are calculated from the average penetration and rate of buying in the *first* period of each pair. We have that 42% of the population bought Brand A in the typical quarter (Table 3.1) and made an average of 3.7 purchases each (Table 3.2). From these two figures, the theoretical model leads us to expect that about 77% of these buyers will buy again in the following quarter – almost exactly as was observed in this case. Table 3.6a gives such comparisons for various lengths of time-periods*.

Discrepancies between the observed and theoretical values help us to describe and to understand any peculiarities of the market. Considering first repeat-buying 12-weeks by 12-weeks, the agreement for the total market and Brands A and E is very close, but the observed incidence of repeat-buyers of Brands B, C and D is substantially below the normal level**. This shortfall seems to link up with the seasonality of these brands (Table 3.1) and will be further examined in Tables 3.9 and 3.10. The 24-week by 24-week pattern is similar, except for an additional 12-point shortfall of repeat-buyers for Brand E which goes with an abnormally heavy rate of buying, as is shown in Table 3.7.

Table 3.7. The Buying-Frequency per Repeat-Buyer

The Average Number of Purchases per Repeat-Buyer in Period of Various Lengths

(Observed Values "O" and Theoretical Norms "T")

	Period of Length (in weeks)							
	1		4		12		24	
	O	T	O	T	O	T	O	T
ANY BRAND	1.2	na	2.5	2.8	5.7	5.7	9.5	9.5
Brand A	1.0	na	2.0	2.1	4.2	4.3	6.8	6.9
Brand B	1.0	na	1.9	1.8	2.9	3.0	4.4	4.2
Brand C	1.1	na	2.1	2.1	3.8	3.5	4.9	4.8
Brand D	1.0	na	1.8	1.8	3.6	3.1	4.9	4.1
Brand E	1.0	na	1.8	1.9	3.9	3.8	6.6	5.5
Average Brand	1.0	na	1.9	1.9	3.7	3.5	5.5	5.1

* The form taken by the theoretical norms here differs from that in Tables 3.1a and 3.2a, although both stem from the same theory. Given the number of buyers and their average purchase frequency in one period, we predict either how many will buy *again* in the next equal period (Table 3.6a) or how many buy *at all* in a period of different length (Table 3.1a).

**Tests of statistical significance for sampling errors are not easy to carry out rigorously when picking on the larger exceptions in large bodies of data (see also Appendix A). But the discrepancies for Brands B, C and D appear fairly large compared with the kinds of sampling fluctuations that can occur in these data.

In very short time-periods, repeat-buying patterns are different and the theoretical model does not apply in any product-field. Thus within a week here, almost no-one makes more than one purchase of the product (an average of 1.01 purchases as shown in Table 3.2a); it would then follow from the NBD/LSD model that there should be virtually no repeat-buyers in the next week. But in practice, Table 3.6a shows that about a third of the weekly buyers of any one brand buy it again in the next week. Weekly shopping habits and "dead-period" effects (the need to more or less use up one purchase before another is made) tend to dominate over the longer-term questions of brand-choice which the NBD theory models. (This is discussed more fully in Chapters 4 and 7.) Even for the 4-weekly periods in Table 3.6a, the repeat-buying estimates tend still for this reason to be somewhat below the observed values.

Table 3.7 The Buying-Frequency per Repeat-Buyer

Tables 3.6 and 3.6a gave the percentage of the buyers of a brand in one period who bought it again in the next, equal-length, period. Table 3.7 now shows how often these repeat-buyers bought the brand in the second period.

The agreement with the theoretical norms is close and shows that there is generally nothing unusual about the repeat-buyers' average frequency of buying, even in those cases where the *number* of repeat-buyers was unusual. This is a typical example of the power of the present approach – to pin-point a deviation in a neat and simple manner.

The largest discrepancy occurs for Brand E in the 24-week periods, where the observed buying frequency of the repeat-buyers is high; this mirrors the shortfall in the 24-week incidence of repeat-buyers in Table 3.6a and shows that total sales accounted for by repeat-buyers (i.e. their number times their purchase frequency) was just about on the mark.

Table 3.8 The Buying-Frequency per "New" Buyer

Table 3.8 gives the corresponding buying frequencies for people who bought in the second period but not in the first. Under no-trend conditions, such "new" buyers are usually infrequent buyers, whose rate of buying is almost invariably low*. In most cases, the theoretical norm is about 1.4 purchases per "new" buyer in the second period.

In the present instance the agreement with the theoretical rates is close enough in the 4-weekly periods, but in the longer periods the observed buying frequencies are higher than the norm. This is an unusual finding. It is however very systematic here for all brands, and is even more marked for the product-field as a whole.

These discrepancies might have been explained by the seasonality of the market, with some "new" buyers being peak-season-only buyers and relatively heavy ones at that, but in fact, the phenomenon occurs all the year round. Instead, we seem to have a situation where some people have an intense but short-lived enthusiasm for a particular brand, but then stop buying it (the "jag" phenomenon, as denoted some time ago by Dudley Ruch). This might seem a normal enough form of buying behaviour, but experience shows that it does *not* occur in most product-fields (see for example Tables 2.1 and 5.1 in Chapters 2 and 5). At this stage, we have therefore established that "new" buyers in this particular product-class tend to be unusually heavy buyers, thus pointing up a phenomenon where further examination can be undertaken.

Table 3.8. The Buying-Frequency per "New" Buyer

The Average Number of Purchases per "New" Buyer in Period of Various Lengths

(Observed Values "O" and Theoretical Norms "T")

	Period of Length (in weeks):							
	1		4		12		24	
	O	T	O	T	O	T	O	T
ANY BRAND	1.1	na	1.5	1.5	2.1	1.6	3.2	1.6
Brand A	1.0	na	1.3	1.4	2.0	1.5	2.3	1.6
Brand B	1.0	na	1.3	1.3	1.8	1.4	2.3	1.4
Brand C	1.0	na	1.3	1.3	1.6	1.4	1.7	1.4
Brand D	1.0	na	1.3	1.3	1.5	1.4	2.3	1.4
Brand E	1.0	na	1.1	1.3	1.5	1.4	2.0	1.4
Average Brand	1.0	na	1.3	1.3	1.7	1.4	2.1	1.4

* As already mentioned when introducing the concept of "new" buyers in §2.2 of Chapter 2, a "new" buyer is defined here as someone who has not bought the item in question in the preceding equal period but may have bought it before that.

Table 3.9 Repeat-Buying in Non-Consecutive Periods

A symptom of one form of "jag-buying" would be a decay of repeat-buying loyalty over longer periods of time. More generally, a repeat-buying rate which falls away in non-consecutive periods might indicate increasing brand-switching and a possible erosion of the brand's franchise. If however buying behaviour is in an equilibrium state, the proportion of repeat-buyers in non-consecutive quarters such as from Quarter I to Quarter III should be the same as that found for consecutive periods: failure to repeat-buy would simply be attributable to infrequent buying and not to any fundamental change in buying habits.

Table 3.9 compares the incidence of repeat-buying in non-consecutive quarters with the figures for consecutive quarters and with the theoretical norms*. Some special turnover of buyers might be expected here because of the seasonality of the market, but in fact there is little difference between the consecutive and non-consecutive figures, irrespective of whether repeat-buying is at the normal level (as for the product-field as a whole and for Brands A and E) or somewhat below (as for Brands B and D). Only Brand C shows some erosion of repeat-buying and this is linked with the drop in the quarterly purchasing rate

Table 3.9. Repeat-Buyers in Non-Consecutive Periods

The Percentage of Buyers in One Quarter who Also Buy Two Quarters Later

(Average of Quarters I/III and II/IV)

	Non-Consec. Quarters	Consecutive (See T3.6)	Theoretical Norm
	%	%	%
ANY BRAND	**84**	**85**	**84**
Brand A	75	78	**77**
Brand B	54	59	**65**
Brand C	40	50	**67**
Brand D	52	53	**64**
Brand E	67	69	**69**
Average Brand	58	62	**68**

* According to the underlying NBD/LSD model as described in the next chapter, the same theoretical norms should apply in consecutive and non-consecutive periods.

of Brand C noted in Table 3.2. The evidence therefore points mainly to a steady level of repeat-buying even over extended time-periods, as is found in most product-fields.

Table 3.10 Repeat-Buying By Light and Heavier Buyers

The tables below present a more detailed analysis of repeat-buying than attempted so far. They consider separately non-buyers, lighter, and heavier buyers in one period and their buying in the subsequent period. This analysis provides a sensitive method of investigating trends in repeat-buying. It is frequently used in the detailed study of specific trend situations (as illustrated in § §5.5 and 6.2 in Part III).

Table 3.10. Repeat-Buying by Light and Heavier Buyers

% Buying in QIII by purchasing level in QII

	Buying of stated item in Quarter II					
	Non-buyers		Once only		More than once	
	O %	T %	O %	T %	O %	T %
NORMAL						
Any Brand	29	26	62	61	93	92
Brand A	18	18	51	55	87	89
Brand E	2	2	40	47	92	86
Average	16	15	51	54	91	89
BELOW NORMAL						
Brand B	10	7	39	47	78	82
Brand C	4	3	25	47	80	85
Brand D	5	2	35	46	87	83
Average	6	4	33	47	82	83

Table 3.10 deals with Quarters II and III, treating separately the total product (i.e. "Any Brand") and Brands A and E, for each of which the incidence of repeat-buying in Table 3.6a had been normal, and Brands B, C and D. For each item, the households in the sample are firstly grouped into those who did not buy the item (i.e. the product or

specific brand) in Quarter II, those who bought it only once, and those who bought it more than once. Thus the table shows what proportion of each group bought it at least once in the *next* quarter (i.e. Quarter III).

The observed buying levels in Quarter III are then compared with the appropriate theoretical NBD norms (see §7.6). We see close agreement at all points for the product-field as a whole ("Any Brand") and for Brands A and E. These are the three items for which the more aggregated repeat-buying levels 12-weeks by 12-weeks in Table 3.6a had already matched up to the theoretical norms.

For Brands B to D the total incidence of quarterly repeat-buyers in Table 3.6a was however below expectations. This could have arisen either from an excess of "new" and "lapsed" buyers moving in and out of the market periodically, or from an actual shortage of repeat-buyers as such – two alternatives with very different implications.

Table 3.10a. Repeat-Buying by Light and Heavy Buyers

Average number of purchases per buyer in Quarter III

	Buying of stated item in Quarter II					
	Non-buyers		Once only		More than once	
	O	T	O	T	O	T
NORMAL						
Any Brand	2.1	**1.6**	2.6	**2.1**	6.1	**6.5**
Brand A	1.9	**1.5**	1.9	**2.0**	4.6	**5.1**
Brand E	1.6	**1.4**	2.3	**1.9**	4.5	**4.8**
Average	1.9	**1.5**	2.3	**2.0**	5.1	**5.5**
BELOW NORMAL						
Brand B	1.8	**1.4**	2.2	**1.8**	3.4	**3.5**
Brand C	1.4	**1.4**	2.6	**1.9**	3.8	**4.4**
Brand D	1.6	**1.4**	2.1	**1.8**	3.8	**3.9**
Average	1.6	**1.4**	2.3	**1.8**	3.7	**3.9**

The analysis in Table 3.10 now shows that the repeat-buying franchise amongst *heavier* buyers of Brands B, C and D is almost precisely at the predicted (normal) level. The shortfall of repeat-buyers in fact occurs amongst *light* buyers of each brand – i.e. those who bought only once in the first period. (There is a compensating excess of "new" buyers in the second period – i.e. people who did not buy in the first period.)

This evidence therefore points once more to seasonal variation with a somewhat higher than normal level of brand-switching amongst *light* buyers. It also shows that there is no fundamental weakness in the repeat-buying loyalty of the heavier buyers of these brands.

Table 3.10a gives the corresponding buying frequencies in Quarter III for the groups shown in Table 3.10. The results for heavier buyers are again close to the norms, but those for the "new" buyers and for the light buyers tend to be somewhat higher than expected. This reflects the unusually high buying rates for "new" buyers which were already noted in the 12-weekly results in Table 3.8.

Summary and Conclusions

The preceding analyses show that penetration growth and repeat-buying for the different brands of the product-field covered here are largely predictable.

One brand may have more buyers than another, but repeat-buying and the growth of penetration over different time-periods generally have the same form. In many respects this is as occurs in other product-fields (as is summarised by the theoretical NBD model), but there are also certain systematic differences from these more general norms. These discrepancies tend to be fairly small and can therefore only be spotted by analyses such as those illustrated here, rather than by direct examination of the observed data as such.

Thus for Brands B, C and D there is some excess of penetration in the longer periods (Table 3.1a), and a substantial shortage of repeat-buyers of each brand from one quarterly (or longer) period to the next (Table 3.6a). For Brands C, D and E there is a small excess of once-only buyers, and more generally there is quite a marked tendency here for "new" or "lapsed" buyers to buy in abnormally heavy bursts or "jags" (Table 3.8).

The analysis of repeat-buying in non-consecutive periods (Table 3.9) shows that there is no marked tendency for the incidence of repeat-buyers (whether low or normal) to be eroded as time progresses. The indications are not so much that "jag-buyers" stop buying a given brand altogether, but that they stop temporarily and that they mostly return to the brand subsequently.

These deviations are partly linked to the seasonal trend in the market but there seems here also to be evidence that a few people temporarily "get tired" of a brand and this is something which could be subjected to further analysis. Thus the analysis in Table 3.10 already shows that from one quarter to the next it is the *light* buyers of each brand (not the *heavier* ones) who tended not to buy again to the normal extent.

For the product-field in total the number of repeat-buyers is normal. The relatively low incidence of repeat-buying for some of the individual brands is therefore a matter of extra switching between brands, rather than a question of moving out of or into the market altogether. A more direct attack on brand-switching behaviour in this product-field is described in Chapter 9.

The analyses have shown that the apparent short-fall of repeat-buyers for Brands B, C, and D is not due to any weakness in repeat-buying at all, but reflects an *excess* of occasional buyers – something with radically different marketing implications.

Despite these various exceptions, the outstanding characteristic of the data analysed here remains their basic regularity. Even the various departures from the norms are too systematic to hide this. The findings are therefore simple enough to lead to a better understanding of buyer behaviour and to provide a basis for evaluating specific marketing problems.

CHAPTER 4

BASIC THEORY

4.1. Three Forms of Repeat-Buying

In this chapter the basic concepts and working formulae of the NBD and LSD theory of repeat-buying will be set out *. Repeat-buying may be regarded as any situation where a consumer buys more than one unit of a particular item, such as a particular brand or pack-size of a consumer product. Three cases of repeat-buying can be distinguished.

Firstly, if a consumer buys the item at all in the given time-period, he may buy it on more than one purchase occasion. Different buyers differ in how often they repeat-buy in this sense. The resultant frequency distribution of purchases (i.e. the number of consumers making 0 or 1 purchases, or, if repeat-buyers, making 2, 3, 4 etc. purchases) can then generally be modelled by the Negative Binomial or Logarithmic Series Distributions. This is discussed in §§4.2 to 4.4 of this chapter.

Secondly, a consumer may buy the item in more than one time-period. In §4.5 a certain underlying model of buyer behaviour is introduced which could represent this form of repeat-buying under stationary no-trend conditions, and which in practice yields formulae which successfully do describe the incidence of period-to-period repeat-buying (§§4.6 and 4.7) as well as related phenomena like penetration growth over different length periods (§4.8). This model also explains the basic NBD or LSD frequency distributions which occur in a *single* period.

The third form of repeat-buying is that more than one unit may be bought on the same purchase occasion. For many frequently-bought branded consumer goods this hardly ever happens (most buyers buy one unit on most occasions – see for example Table 3.3 in Chapter 3). But even for product-fields where multi-unit purchases do occur (e.g. dog foods or, in a sense, petrol), the tendency is for the average number

* The mathematical formulae can be passed over quickly by the reader, especially on a first or second reading: all one needs to note is that certain formulae exist for the stated purpose. The more mathematical details and proofs are in any case left over to Chapters 7 and 8, for the more technical reader. A worked numerical example for the calculations involved is given in Appendix A.

of units bought per purchase not to vary much from brand to brand. This form of repeat-buying therefore seems relatively easy to deal with by straightforward analysis. We therefore by-pass it here through using the *purchase occasion* as analysis unit, as has already been indicated in Chapter 1.

4.2. The Number of Purchases in a Time-Period

For most kinds of consumer products there tends to be something like a minimum time-interval between one purchase and another. This is a week for many types of household products (which are normally bought *at most* once a week), or a day or so for cigarettes. We now consider the frequency of purchase in time-periods which are relatively long compared with this minimum (e.g. months or quarters or years, rather than a week). Special problems arise for periods near the minimum and these will be discussed in § 4.9 and in Chapter 7.

Even in a relatively long time-period such as a month or a year, many buyers of any particular brand buy it only once, some buy it twice, fewer buy it three times, and so on, with smaller frequencies of still heavier buyers. Furthermore, the total number of consumers who buy the brand at all in the analysis-period is often a relatively small proportion of the total population of potential buyers, so that the largest single observed frequency is usually that of *non-buyers* of the item.

Typically, we therefore have a highly skew distribution (as illustrated by the column of "observed" figures in Table 4.1). The theoretical finding then is that such distributions can generally be fitted by one particular mathematical function, namely the Negative Binomial Distribution (or NBD for short)*. A typical example — the first one published — is given in Table 4.1 from a certain 26-week period, with 1612 households out of a sample of 2,000 not buying the item at all,

* The need to find some such theoretical model for the observed frequency distributions of purchases was the start of the present work on repeat-buying. It arose specifically from a suggestion by Mr. D.A. Brown in dealing with a practical data-handling problem in the late 1950s which concerned a possible excess of "heavy buyers" in some observed data. The most obvious theoretical distribution to try and fit to the distribution of relatively rare events like consumer purchase is the Poisson, but this was found to be not skew enough for the observed data. The NBD is more skew than the Poisson. It was one of several theoretical distributions which was tried next and it gave a good fit, both in the initial studies and subsequently (see also § 11.4 in Chapter 11).

Table 4.1. The Earliest Published Example of the Fit of the Negative Binomial Distribution to Purchasing Data

(26-week data for a 2,000 household sample)

Number of Purchases	No. of Households	
	Observed	Theoretical
0	1,612	(1,612)*
1	164	157
2	71	74
3	47	44
4	28	29
5	17	20
6	12	15
7	12	11
8	5	8
9	7	6
10	6	5
11–15	11	12
16+	8**	7

Proportion of non-buyers: p_0 = .806*
Proportion of buyers: $b = 1-p_0$ = .194

Average number of purchases *per household*: m = 0.64*
Average number of purchases *per buyer*: $w = m/b$ = 3.3

Negative Binomial exponent: k = 0.115, and $a = m/k$ = 5.53

Standard deviations: Observed (root-mean square) = 2.12
Theoretical $\sqrt{\{m(1+a)\}}$ = 2.04

* Used in fitting the theoretical distribution.
** Actual values: 17, 17, 20, 22, 22, 25, 26, 26.

164 households buying it once, 71 twice, and so on [Ehrenberg 1959]. The theoretical NBD purchase frequencies are 157 households buying once, 74 buying twice, and so on, and the agreement is clearly good.

Two coefficients or "parameters" have to be calculated in order to fit the theoretical distribution. This is usually done by using the observed proportion of zeros (the number of non-buyers) and the average number of purchases per household (the mean of the distribution) as inputs. These two quantities are usually denoted by p_0 and m, and the technical details are set out in the next few paragraphs.

The Mathematics of the NBD. The Negative Binomial Distribution is based on the non-negative integers 0, 1, 2, 3, 4 and in general r. The distribution is specified by two parameters, the mean, m, and the exponent k. The probability p_r of observing r purchases is then given by

$$p_r = \frac{(1+m/k)^{-k}\,\Gamma(k+r)}{\Gamma(r+1)\,\Gamma(k)} \left(\frac{m}{m+k}\right)^r ,$$

which follows from expanding the binomial expression $(1-m/m+k)^{-k}$, in which the exponent k has a negative sign. The negative binomial distribution is always positively skewed, and has one mode, which is at zero for the fairly small values of m and k which generally occur with consumer purchasing data. The variance of the distribution is $m(1 + m/k)$ or $m(1 + a)$, where the quantity $a = m/k$ is another useful way of expressing the parameters of the NBD.

In fitting a negative binomial distribution to empirical data, the two parameters m and k have therefore to be estimated. The best estimate of the mean m is simply the observed mean (for sample data it is the maximum-likelihood estimator and is also unbiased).

Estimating the second parameter k (or a) is less straightforward. (The maximum-likelihood equations for the second parameter are for example very cumbersome to solve.) Various ways of estimating k or a have however been developed [cf. Anscombe 1950]. An estimating procedure common in statistics is the method of moments, which here means estimating a by equating the observed sample variance to its expected value $m(1 + a)$. This is however not particularly efficient statistically for the NBD (sometimes less than 50%), although it would be adequate with large samples; but in practice it would often be laborious to have to compute the observed variance, since with market research data the basic frequency distribution is not usually tabulated.

An alternative method utilizes the number of non-buyers. We equate the observed proportion of zero readings p_0 to its expected value, i.e. we write

$$p_0 = (1+m/k)^{-k}, \qquad \text{or} \qquad (1+a)^{-m/a} .$$

This equation cannot be solved directly for k or a, from given values of p_0 and m, but various indirect procedures exist *. One is by iteration and this can readily be computerised if it is to be done routinely. Another is to work out the quantity $c = -m/\ln p_0$ from the observed values of m and p_0 and read off the corresponding value of a from Table B.3 in Appendix B, taken from Chatfield [1969]. (A transformation such as $c = -m/\ln p_0$ makes for ease of accurate interpolation; this is a simpler version, developed by G.J. Goodhardt, of an earlier suggestion by Evans [1953]†.)

In general, it is very convenient with consumer purchasing data to estimate the second parameter k or a of the NBD from the proportion of non-buyers p_0 in this way. Thus the mean m and the proportion of non-buyers p_0 – or the proportion of *buyers* $(1-p_0)$, usually denoted by b – are often all the observed figures that are routinely tabulated. Statistically, this method is in any case at least 90% efficient for most such data, and often a good deal more so [cf. Anscombe 1950, for very low values of k].

Having estimated the two parameters m and k (or a), we need to calculate the theoretical proportion p_r of the sample who make r purchases. A relatively simple procedure is to use an

* Both the NBD and LSD theories appear mathematically more cumbersome than they really are because each involves, at its very earliest stage, a relationship between the theoretical parameters and the observed data which cannot be solved directly.

† Data for which $m < -\ln p_0$ cannot be fitted by an NBD.

iterative formula (starting with the observed value of p_0), namely

$$p_r = \left(\frac{a}{1+a}\right)\left(1 - \frac{a-m}{ar}\right)p_{r-1}\,.$$

The values tend to be tedious to compute by hand if r is at all large, but the procedure can be readily computerised for routine use.

The goodness of fit of the estimated distribution can be tested by calculating the chi-squared value for the observed and theoretical frequencies along standard statistical lines. A quicker test in practice is to compare the observed variance with the theoretical value $m(1 + a)$, the distribution having been fitted by the mean and the proportion of zeros. (A test of significance for this difference has been given by Evans [1953].)

4.3. The Fit of the Negative Binomial Distribution

In fitting the negative binomial distribution to consumer purchasing data, a good fit has been obtained in most stationary cases, i.e. when there is no marked trend in the aggregate sales level. A typical example, taken from the earliest data analysed, was shown in Table 4.1, the fitting as usual being by equating the number of zeros and the mean of the *theoretical* distribution to the corresponding *observed* values. The fit in this case was clearly close, as may be seen by eye. (A summary measure of the fit is obtained by comparing the standard deviations of the observed and theoretical distributions, at 2.12 and 2.04.) Many thousands of further cases with widely differing characteristics have now been successfully analysed, as is summarised in Table 4.2. Some recent examples were shown in Table 3.4 in Chapter 3, where certain specific discrepancies between the observed and theoretical frequencies are also discussed.

In general then, the fit of the theoretical NBD to observed purchasing data is good. Specific exceptions can however also occur. Some were already noted in the earliest work and include a clustering tendency at or near the number of weeks in the analysis unit [Ehrenberg 1959]. This is a manifestation of the minimum inter-purchase time-period effect which was referred to briefly at the beginning of § 4.2 and will be discussed more fully in § 4.9 and in Chapter 7. Thus for products which tend to be bought at most once a week, the number buying *more* than 26 times in a 26-week period, say, is less than the theoretical value. There is then something of a compensating excess of people who regulate their purchases to roughly the once-a-week cycle. Examples were given in Table 2.1 in Chapter 2, but this effect only shows up for products where the number of buyers who *do* buy once a week is

Table 4.2. Conditions under which the Negative Binomial Distribution has Generally been Found to Hold ***(updated as Table 1 in Preface)***

– For a variety of product-fields, *viz*:

Breakfast Cereals, Butter, Canned Vegetables, Cat and Dog Foods, Cocoa, Coffee, Confectionery, Convenience Foods, Cooking Fats, Detergents, Disinfectants, Flour, Food Drinks, Household Soaps, Household Cleaners, Instant Potatoes, Jams and Marmalade, Margarine, Motor Oil, Petrol, Polishes, Processed Cheese, Refrigerated Dough, Sausages, Shampoos, Soft Drinks, Soup, Take-home Beer, Toilet Paper, Toilet Soap

– The leading brands in each product-field

– Large, medium and small pack-sizes and the brand as a whole

– Great Britain, Continental Europe, U.S.A.

– 1950–1970

– Various Demographic Subgroups

– Analysis Periods ranging from 1 week to 12 months

appreciable *. The shortfall of heavier buyers then leads to a lower value for the standard deviation (or variance) of the observed distribution as against that of the theoretical NBD. This is the so-called "variance discrepancy" which will be discussed more fully in Chapter 7 †.

There are also some products where the fit is not good in ways that are not yet fully understood. Essentially this seems to apply particularly to very regularly bought items such as perhaps bread, cigarettes and milk, or in certain cases also the total product-class (even when individual brands or pack-sizes give a good fit). Further discussion of discrepancies occurs at later stages in this book.

4.4. The LSD Approximation

A simplifying approximation to the Negative Binomial Distribution which holds under certain circumstances is the Logarithmic Series Distribution or LSD. (This is the model used in § 2.3 of Chapter 2.) It

* It hardly shows up in the example of Table 4.1, although the three heaviest buyers (out of the 8 buyers who made 16 or more purchases) bought 25 or 26 times in the 26-week period, and no one bought *more* often than 26 times.

† The title "variance discrepancy" arose historically but does not refer to what is now regarded as the most pertinent feature of this phenomenon.

applies when the proportion of the population buying the item in the analysis-period is relatively low – roughly that b, i.e. $(1-p_0)$ is less than about 0.2 or 20%. It is then possible to fit the distribution of *buyers* (i.e. excluding the zeros or non-buyers) by a certain theoretical distribution, the LSD. The number of non-buyers is therefore treated separately from the buyers, and the LSD has only one parameter and hence is simpler than the NBD which has two (cf. Chatfield et al. 1966).

The LSD is not an *alternative* to the NBD, but merely a special case of it (see Chapter 8) which under certain conditions (i.e. roughly for $b < .2$) gives virtually the same results, but more simply. Since the LSD has only one parameter, many formulae for repeat-buying etc. are particularly straightforward in their LSD form, as was illustrated in Chapter 2 and is discussed more fully in § § 4.7 and 4.8 of this chapter. An extreme example is the formula for the proportion of the total sales of an item which is accounted for by those relatively "heavy" buyers of the item who buy it more than r times in the analysis-period. In the LSD theory there is the very simple formula q^r which was used in § 2.3 of Chapter 2 and which is discussed below and in § 8.4 of Chapter 8. In the NBD theory there is *no* equivalent (simple) formula.

Conceptually the importance of the LSD is that it shows how buyer behaviour is independent, or at least largely independent, of the precise definition of the population of potential buyers, as long as this is large enough compared with the number of *actual* buyers (i.e. $b < .2$ or so).

The Mathematics of the LSD. We let p'_r stand for the proportion of all buyers in the given time-period who made r purchases of the item in that period. The values of p'_r for all $r \geq 1$ can then in practice be represented by the one-parameter Logarithmic Series Distribution, irrespective of the precise number of non-buyers (as long as this is large). It is important to note that p'_r is used here to stand for the proportion of *buyers* in the given period who made r purchase, whereas p_r without the dash stands for the proportion of the total sample (i.e. including the non-buyers) who made r purchases. We therefore have

$$p'_r = p_r/(1-p_0)\,.$$

The mean of the distribution is denoted by w, where $w = m/(1-p_0)$ in terms of the mean m of the total distribution (including zeros).

The LSD probabilities for there being p'_r buyers making r purchases (for $r \geq 1$) are best expressed in terms of a certain parameter q:

$$p'_r = \frac{-1}{\ln(1-q)}\,\frac{q^r}{r}\,.$$

the logarithm (ln) being to base e. (Tables of such "Natural" or "Naperian" logarithms to base e are reproduced in Appendix B.) The parameter q can be related to w, the mean of the observed distribution of purchases, by the implicit equation already noted in Chapter 2,

$$w = \frac{-q}{(1-q)\ln(1-q)}\,.$$

This equation cannot be solved directly for q in terms of w (see also the second footnote in §4.2). The numerical value of q for a given w can be obtained from Table B.2 given in Appendix B, of which Table 2.5 in Chapter 2 was a sample extract. Alternatively, q can be calculated by iteration (e.g. on a computer), or by the simple approximation

$$q \doteq (w-1.4)/(w-1.15)\,, \qquad \text{which holds to within } \pm.01 \text{ for } w > 2.$$

The value of q lies between 0 and 1. It quickly reaches high values close to 1 – e.g. it is as high as .9 for a relatively low value of w such as $w = 4$ – so that considerable accuracy is often needed in its arithmetical calculation. A better way of calculating q from w is through forming a new parameter $a = q(1 + q)$ and using Table B.3 in Appendix B, as is discussed in §8.3.

One aspect of the distribution which is of some special interest is its "tail", i.e. the purchases made by heavier buyers. This amounts to Σip_i for values of i greater than some number r. Expressed as a proportion of total purchases, this can easily be shown to reduce (see §8.4) to the simple expression q^r, as already noted earlier.

4.5. An Underlying Model of Stationary Buyer Behaviour

So far we have examined the incidence of purchases in a single time-period of some given length. Under stationary conditions, the distribution of purchase frequencies then tends to follow the NBD or LSD and this applies for a period of any length (subject to the "minimum time-period" limitation touched on in §4.2 and discussed further in §4.9). We have however not yet considered how the results in periods of different lengths are related, nor have we yet dealt with repeat-buying from one period to another. For all this, a more elaborate model is needed.

The NBD approach is fruitful in this respect because there is a theoretical model of a stochastic (or "quasi-random") kind which not only yields an NBD in any single time-period, but also provides formulae relating the results in different length periods and for period-to-period repeat-buying. Furthermore, the model in question not only seems "reasonable" on *a priori* grounds, but has also been found to work well in practice.

The model is a two-dimensional one, one dimension being time, and the other (an unordered one) being the individual consumers, as is

* In the modern literature, this type of model tends now to be referred to as a "mixed Poisson".

Table 4.3. A Stochastic Model over Time yielding the NBD in any given Period

	Successive Time-Periods						Long-run Averages	Distributions (horizontally)
	I	II	III	IV	.	.		
Consumer								
a	x	x	x	x	x	x	μ_a	Poisson
b	x	x	x	x	x	x	μ_b	Poisson
c	x	x	x	x	x	x	μ_c	Poisson
d	x	x	x	x	x	x	μ_d	Poisson
.	x	x	x	x	x	x	$\mu_.$	Poisson
.	x	x	x	x	x	x	$\mu_.$	Poisson
Mean	m	m	m	m	m	m	m	
Distributions (vertically)	NBD	NBD	NBD	NBD	NBD	NBD	Gamma	

Note: The x-values in the body of the table represent varying observed numbers of purchases and are not intended to imply equality.

shown schematically in Table 4.3. The model is of a so-called "compound Poisson" type * [e.g. Anscombe 1950, Feller 1957, Haight 1965, Boswell and Patil 1970], the details being broadly as follows:

The Incidence of Purchases over Time. Any particular consumer makes some sequence of purchase of an item, e.g. 2, 0, 3, 1, 1, etc. purchases, in successive equal periods of time (see for example Table 1.2 in Chapter 1). The model then requires that these purchase frequencies should behave like independent random samplings from a system where the event (i.e. a purchase) has the same probability at any given point in time, and where these probabilities are independent of each other. This is then a so-called Poisson distribution.

This Poisson formulation for individual purchase sequences is a plausible *a priori* assumption under two conditions which should normally be more or less fulfilled, namely that

(a) Successive analysis-periods must not only be of equal length but also be "similar" to each other, e.g. weeks or longer periods measured in weeks for household-products (rather than days, because shopping patterns on Mondays tend to differ from those on Tuesdays or Saturdays and hence tend to be non-stationary, whereas in periods longer than a week, these short-term effects are in effect balanced) *. This "similarity" of the successive periods has to extend to actual purchasing behaviour itself. In other words, there must be no trend in the aggregate sales or penetration figures (or in the parameters of the underlying model). This is the "stationarity" assumption.

(b) The analysis-periods must not be too short, so that the purchases made in one period do not directly affect those made in the next. Periods of one week for instance may be too short for some products. For example, if a tin of cocoa is bought in one week, no such purchase is likely to be made until the initial purchase is more or less used up. This is part of the minimum time-interval problem which is discussed further in §4.9.

In practice, these conditions are often likely to hold for periods which are long enough in terms of the minimum inter-purchase or usage habits for the product in question †. The basic test is of course how well a model based on such assumptions works in practice.

The Poisson distribution has one parameter, which may be denoted by μ, the consumer's average rate of purchasing "in the long run". We now consider the distribution of these mean values for different consumers.

Differences between Consumers. The second part of the model concerns the differences in the average purchasing rates of different consumers.

The model states that the frequency distribution of the long-run average purchasing rates μ_a, μ_b, μ_c, etc. of different consumers *a, b, c,* etc. should be proportional to a so-called "Gamma" distribution (with exponent k). This is a statistical probability distribution for non-negative values which is either reversed J-shaped or hump-backed, and always positively skewed. It is therefore of the right general shape (this being what the observed data generally look like), and is rather flexible, having two adjustable parameters. In fact, many different kinds of data

* At a more technical level, we note that purchasing behaviour is patterned in discrete time-periods, although the Poisson process is really continuous in time. When working in relatively long analysis-periods, this discrepancy does not greatly affect the fit of the model in most respects, although it underlies the problems already touched on in §4.3.

† Something is known about the way in which the Poisson assumption does not quite hold in practice (cf. §7.8), but this has little effect on most of the results here.

can be fitted by a Gamma-distribution, and this is therefore not a particularly stringent assumption *.

To summarise, the general model is one of Poisson distributions "compounded" by a Gamma-distribution, in that the *i*th consumer's purchases over time are to follow his own Poisson distribution with a mean μ_i, the means for different consumers following a Gamma-distribution, as was set out in Table 4.3. It is not necessary to assume that purchases actually follow this Poisson-Gamma model in successive periods, let alone in the long run. (The average "long-run" rates of purchasing μ_a, μ_b, μ_c, etc. which have been postulated for individual consumers are in fact not directly observable.) Instead, it is only necessary to suppose that for any one or more periods of time, consumers' purchases behave *as if* they were random samples from such a stochastic model. (In the next period, the parameters of the model – or the model itself – could have changed, due to some marketing disturbance or seasonal trend say, although usually this does not happen.)

A number of mathematical deductions can then be made from this formulation and tested empirically. The first deduction of this kind is that the distribution of purchases in any single time-period should follow the negative binomial distribution, (see § 7.2). This tends to be so in practice, as we have seen, and was of course the starting-point of the whole study.

Other deductions concern repeat-buying from one period to the next (to which we now turn), and the way in which the values of the penetration b and of the average purchase frequency w vary in periods of different lengths (as is discussed in § 4.8). It is the extent to which these deductions generally fit the observed facts which determines the practical validity and usefulness of this Gamma-Poisson model.

4.6. Period-to-Period Repeat-Buying

The repeat-buying pattern from one period to the next is of fundamental interest both for the theoretical understanding of buyer behaviour and for practical applications. There are four main questions to be

* By the same token, it is not possible (or necessary) to adduce any strong reasons why a Gamma-distribution should hold. No special consequences follow from the assumption other than the various repeat-buying formulae etc. whose validity is tested directly against the observed data anyway. (This has been superseded – See Goodhardt and Chatfield 1973 and Chapter 13, §13.2).

Table 4.4. The Definition of Repeat, Lapsed and "New" Buyers in Two Time-Periods

Definition	Period I	Period II
Repeat-buyers:	Buying	Buying
Lapsed buyers:	Buying	Not buying
"New" buyers:	Not buying	Buying
Non-buyers:	Not buying	Not buying
All buyers in period:	Repeat + Lapsed	Repeat + "New"

posed to any theoretical formulation, namely whether the theory can successfully predict:

(a) What proportion of the buyers of the item in one period also buy it in the next period and by implication,

(b) What proportion of buyers in each period do *not* buy in the other, i.e. are "lapsed" or "new" buyers, as is set out schematically in Table 4.4. Next,

(c) How often the repeat-buyers buy in each period, and finally,

(d) How often the "lapsed" buyers and the "new" buyers – i.e. the one-period-only buyers* – buy in the period in which they buy at all.

We now set out the answers to these questions which are given by the NBD/LSD theory under stationary (i.e. no-trend) conditions from one period to the other. The point is that if we are given data about purchasing behaviour in the *first* period, the theory predicts repeat-buying behaviour concerning the *next* period†.

4.7. Three Levels of Theory

The NBD/LSD repeat-buying theory gives results at three levels. These levels are of increasing mathematical simplicity but of decreasing generality. They are the NBD, the LSD, and certain numerical approximations to the latter.

* One-period-only out of the *pair* of periods being analysed.

† The NBD theory also copes with various elaborations of the above questions, such as the incidence of repeat-buying in the second period amongst light, medium, or heavy buyers in the first period, or the nature of repeat-buying in unequal time-periods, or in more than two periods, or in non-successive periods. The answers to these more elaborate questions are set out in Chapter 7 (with some practical applications being already given in Chapters 3, 5 and 6).

In using the NBD and the LSD theories to estimate how many repeat-buyers there should be in the second of two periods and how often they should buy, it is first of all necessary to calculate certain characteristic numbers or "parameters" for the *first* time-period. In the NBD there are *two* such parametes. One is m, which is simply the observed mean number of purchases per informant in the first period. The second is the special NBD parameter k (the negative binomial "exponent"). Its value can be calculated from the observed values of m and b (the observed proportion of buyers) in the first period. The NBD equation linking these observed values to k is set out in Table 4.5, and the calculations involved in solving this equation have already been outlined in §4.2 earlier in this chapter. (A numerical example of this and the remaining calculations in this chapter is set out in Appendix A.)

In the LSD theory, there is only *one* special parameter value that needs to be worked out. This is the LSD parameter q which can be calculated, as already mentioned in §4.4 and in Chapter 2, from the observed value of w, the average number of purchases per buyer in the first time-period. The relevant equation is also given in Table 4.5. Since the LSD theory requires only this single parameter it gives simpler results than the NBD. But the theory holds to a reasonably close degree of approximation – i.e. gives virtually the same results as the NBD – only for data where the proportion of buyers, b, is less than 20%. For items with a higher penetration, the LSD and NBD repeat-buying estimates diverge, and it is the NBD values which give a good fit.

The third type of repeat-buying formulae consists of numerical approximations to the LSD ones. Here there is no need to calculate any special parameter, as the formulae are expressed directly in terms of the

Table 4.5. The Basic Equation for the NBD and LSD parameters k and q, in terms of the Observed values of b, m and w

NBD:

$$b = 1-(1+m/k)^{-k}, \text{ or}$$
$$= 1-(1+a)^{-k}, \text{ where } a = m/k.$$

LSD:

$$w = -q/\{(1-q)\ln(1-q)\}, \text{ where } w = m/b.$$

b = the proportion of informants buying the item in the period
m = the average number of purchases per *informant*
w = the average number of purchases per *buyer* ($w = m/b$)

observed value of the average purchase frequency, w, in the first time-period. Arithmetically, these are therefore the simplest formulae, but they give close approximations to the NBD/LSD results only for a certain range of values of w, generally that w is greater than about 2 but less than about 20 (and of course that the penetration b is less than 20%, as for the LSD theory generally).

Given the appropriate values in the first time-period, each of the three levels of theory can now be used to give estimates of the various aspects of repeat-buying in the next equal period, under stationary conditions. (Whilst the LSD or approximation formulae are simpler – in those situations where they give the right answers at all – the more general NBD formulae are normally used when relatively large-scale or repetitive applications are involved, the calculations being simple enough when routinised on a computer.)

We start with the incidence of repeat-buyers, i.e. the proportion b_R of the population who buy in both periods. The three formulae for b_R are set out in Table 4.6 (the LSD and the "Approximation" formulae being expressed as b_R/b, i.e. as the proportion of the buyers in the first period who buy again in the next, as they are simpler in this form). The numerical values given by these formulae are illustrated in Table 4.15 at the end of this chapter.

The earliest published examples of the fit of the NBD formula to empirical data is reproduced in Table 4.7: there are some discrepancies (partly due to some minor non-stationarity and to sampling errors), but

Table 4.6. Three Formulae for b_R, the Proportion of the Population who are Repeat-Buyers in Two Equal (Stationary) Time-Periods, in terms of the Observed Parameters in Table 4.5.

NBD:

$$b_R = 1 - 2(1+a)^{-k} + (1+2a)^{-k}$$

LSD (for $b < .2$):

$$\frac{b_R}{b} = 1 + \frac{\ln(1+q)}{\ln(1-q)}$$

Approximation (for $b < .2$ and $2 < w < 20$):

$$\frac{b_R}{b} \doteq \frac{2(w-1)}{2.3w-1}$$

Table 4.7. Example of the Fit of the NBD Formula for the Percentage of Repeat-Buyers

(The first published examples of NBD estimates of quarter-by-quarter repeat-buying for 12 different items [Ehrenberg 1964])

	Brand or Pack-Size											
	L	M	N	O	P	Q	R	S	T	U	V	W
The Given Data in Quarter I:												
% buying = 100 b	1.3	3.8	4.5	4.9	6.5	8.9	10	12	15	15	24	34
Mean buying rate = m	.03	.09	.14	.15	.20	.26	.28	.53	.67	.66	.80	1.5
% of Sample Buying in both Q.I and Q.II:												
Theoretical	.07	2.4	3.1	3.4	4.5	6.1	7	9	11	11	18	26
Observed	.06	1.4	2.2	2.5	4.1	4.5	7	9	10	8	18	26

there clearly was major agreement overall. This has been confirmed in many thousands of cases which have been examined since, some examples being given in Chapter 3 (Table 3.6) and in Chapters 5 and 6, where discrepancies are also further discussed (see also § 4.9 and Chapter 7).

Next, we consider the average frequency with which these repeat-buyers buy in the second period, expressed as m_R on a "per informant" basis in the NBD theory, or as w_R on a "per repeat-buyer" basis in the LSD, i.e. dividing through by the number of repeat-buyers. Table 4.8 gives the three formulae, the numerical approximation to the LSD giving the particularly simple result

$$w_R \doteqdot 1.23w\ ,$$

for values of w such that $1.5 < w < 20$.

Table 4.8. Three Formulae for w_R, the Average Purchase Frequency per Repeat-Buyer under Stationary Conditions

NBD :

$$m_R = m\ \{1 - (1+a)^{-k-1}\}\ ,\ \text{where } m_R = w_R b_R$$

LSD (for $b < 0.2$) :

$$w_R = -q^2 / (1 - q) \ln (1 - q^2)$$

Approximation (for $1.5 < w < 20$) :

$$w_R \doteqdot 1.23w$$

Turning to the "new" buyers in the second period, i.e. those buyers in the period who had not bought the item in the preceding period, it follows from Table 4.4 that b_N, the proportion of the population who are "new" buyers, is the difference between b, the total incidence of buyers in the period, and b_R, the incidence of repeat-buyers, i.e.

$$b_N = b - b_R\ .$$

The formulae for b_R or b_R/b which were given in Table 4.6 therefore indirectly provide estimates of b_N *.

* Under stationary conditions, b_L, the incidence of "lapsed" buyers – buying in the first but not the second period – is numerically equal to b_N. The value of b, the total number of buyers, in the first period is (by the definition of stationarity) equal to the value of b in the second period, and the values of b_R are necessarily identical. Under stationary conditions the average purchase frequencies per repeat-buyer in the first and second periods are also equal, and so are the average purchasing rates w_N and w_L per "new" and per lapsed buyer.

Table 4.9. Three Formulae for w_N, the Average Purchase Frequency per "New" Buyer under Stationary Conditions

NBD :

$$m_N = m\,(1+a)^{-k-1}, \text{ where } m_N = w_N b_N$$

LSD (for $b < 0.2$) :

$$w_N = q / \ln(1+q)$$

Approximation (for $w > 2$) :

$$w_N \doteq 1.4$$

Finally, w_N, the average purchase frequency per "new" buyer (or w_L, the average per "lapsed" buyer) is given by the formulae in Table 4.9. For items bought at an overall average frequency of at least 2 purchases per buyer (i.e. $w \geqslant 2$), the theoretical LSD value of w_N varies only between 1.35 and 1.44 and in the numerical approximation it can therefore be treated as a quasi-constant, i.e.

$$w_N \doteq 1.4\,,$$

the simplest result of all. (For $w < 2$, the theoretical value of w_N or w_L decreases from 1.3 down to 1.0; for $b > .2$, the NBD values of w_N and w_L are of course appropriate and these can increase beyond 1.4.)

The tendency for w_N (and w_L) to be virtually "constant" at 1.4 under a wide range of conditions is less surprising than it might seem at first sight. Thus in the underlying model of stationary buyer behaviour of § 4.5, "new" buyers in a given period are essentially *infrequent* buyers, and mostly buy once or perhaps twice in the period. Hence it is not surprising that on average the "new" buyers buy *about* 1.4 times. In longer period, there are fewer "new" (or "lapsed") buyers than in shorter ones, but they are still mostly once-only or twice-only buyers, and hence it remains "intuitively" acceptable that the values of w_N and w_L need not vary with the length of analysis-periods *.

For the various other aspects of repeat-buying, this virtual independence from the length of the analysis-period does however not occur.

* Under no-trend conditions, people who buy *often* in the first period would in general not stop buying the item altogether in the second period, especially given that the choice of time-periods (i.e. the particular dividing-line between them) is essentially arbitrary anyway.

Thus the number of "new" (or "lapsed") buyers, the number of repeat-buyers, and their average purchase frequency, all vary with the length of the analysis-period, as does the *total* number of buyers (the penetration b) and their average purchase frequency w. For example, in a longer period there are not only more buyers than in a shorter period, but a higher proportion of them will buy again in the next equal period.

Since the various repeat-buying formulae depend at most on the two basic parameters b and w for *all* buyers in the analysis-period, we only need to know how these two basic parameters vary with the length of analysis-period. This then provides the input for calculating the repeat-buying estimates according to the length of the analysis-period.

4.8. Time-Periods of Varying Length

The general form of the relationship between the penetration and the length of the time-period is implicit in the repeat-buying formula for two successive periods of a given length which was discussed in the previous section (Table 4.6). Thus under stationary no-trend conditions, the number of buyers in a time-period made up of two equal shorter sub-periods is the sum of the buyers in each sub-period *minus* the number of repeat-buyers (the latter being counted amongst the buyers in each sub-period – see Table 4.4). In symbols, if b_2 is the proportion of the population buying in a period of length 2, then

$$b_2 = 2b - b_R,$$

where b is the proportion buying in the typical "unit" period and b_R is again the proportion of repeat-buyers from one of the two unit periods to the other. The formulae in Table 4.7 then allows us to calculate the penetration in the double period, b_2, from the parameters of the unit period.

More generally, for any time-period of length T, the proportion of the population, b_T, buying in period T can be expressed in terms of the observed data (or derived parameters such as k or q) for a "unit" length time-period. The alternative NBD and LSD formulae are set out in Table 4.10, and Table 4.11 gives the earliest empirical results for some 14 different brands or pack-sizes examined in this way, where the penetration in 6 months is successfully predicted from 4-weekly data (to within about ± 1 percentage point). More recent examples are illustrated in Table 3.1a in the previous chapter.

Table 4.10. Three Formulae for b_T, the Proportion of the Population Buying in a Period of Length T

(The parameters b, m, a, k or q are for the "unit" period under stationary conditions – see Table 4.5)

NBD :

$$b_T = 1 - (1 + m_T/k_T)^{-k_T}, \text{ by definition of the NBD,}$$
$$= 1 - (1 + Tm/k)^{-k}, \text{ since } m_T = Tm \text{ and } k_T = k.$$

LSD (for $b_T < 0.2$) :

$$\frac{b_T}{b} = 1 - \frac{\ln\,\{1 + (T-1)\,q\}}{\ln\,(1-q)}$$

Approximation (for $w > 1.5$ and $w_T < 20$) :

$$\frac{b_T}{b} = \frac{Tw}{\{1 + (w-1)\,T^{0.82}\}}$$

Table 4.12 sets out the corresponding relationships between w, the average number of purchases per buyer in "unit" period and w_T, the average number of purchases in period T made by the larger number (b_T) of buyers then. The approximation to the LSD formula here is rather simple. Thus

$$(w_T - 1) = T^{.82}(w - 1)\,,$$

as long as $w > 1.5$ and $w_T < 20$ (and $b < .2$). The quantity $T^{.82}$ (which also occurred for b_T in Table 4.10) can be readily calculated using log tables, but for simplicity, its values for some commonly occuring values of T are set out in Table 4.13. A case of specific interest is $T = 2$, when

$$(w_2 - 1) = 1.76(w - 1)\,.$$

We note that the expression $(w-1)$ here in fact represents the average number of *repeat-purchases* made per buyer (i.e. after discounting the first purchase); it is a way of expressing the data which numerically leads to various other simplifications in the LSD theory (see Chapter 8).

These various "time-period" formulae have already been illustrated in Chapter 3 (Tables 3.1 and 3.2a) for monitoring the growth of penetration and of buying frequency, and are also used in §6.3 of Chapter 6 for interpretative extrapolations to longer periods.

Table 4.11. Examples of the Fit of the NBD Formula for the Percentage of Buyers in a Longer Time-Period

(The first published examples of predicting the 24-week penetration from 4-weekly data for 14 different items [Ehrenberg 1962])

	Brand or Pack size													
	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	*m*	*n*
The Given Data in 4 Weeks:														
% buying = $100b_4$	1.4	1.7	2.7	2.9	3.8	4.4	5.1	6	7	7	10	11	12	22
Mean buying rate = m_4	.03	.03	.06	.06	.12	.12	.11	.11	.15	.14	.22	.28	.36	.49
% of Sample Buying in 24 Weeks = $100b_{24}$:														
Theoretical	4	4	6	7	7	8	11	15	16	17	20	22	20	41
Observed	4	4	5	8	8	8	13	15	16	18	19	22	22	40

Table 4.12. Three Formulae for w_T, the Average Purchase Frequency per Buyer in a Period of Length T under Stationary Conditions

(The parameters m, k, w and q refer to the time-period of "unit" length)

NBD :

$$w_T = Tm/b_T$$
$$= Tm/\{1 - (1+Tm/k)^{-k}\}$$

LSD (for $b < 0.2$):

$$\frac{w_T}{w} = \frac{T \ln(1-q)}{\ln(1-q) - \ln\{1+(T-1)q\}}$$

Approximation (for $w > 1.5$ and $w_T < 20$) :

$$(w_T - 1) = T^{0.82}(w-1)$$

Table 4.13. Values of $T^{0.82}$ for some Commonly Occurring Values of T, to Use in the LSD Approximations for b_T and w_T

T	2	3	4	5	6	12
$T^{0.82}$	1.76	2.46	3.12	3.74	4.34	7.67

4.9. Conditions of Fit

The technical details of the NBD/LSD repeat-buying theory outlined in this chapter are discussed more fully in Chapters 7 and 8 (including various elaborations such as the formulae for the "conditional" kind of repeat-buying analyses used in Table 3.10 in the previous chapter). The theory tends to fit the kinds of empirical regularities that were illustrated in Chapters 2 and 3, further examples being given in Chapters 5 and 6. Six common types of discrepancy can however also be distinguished.

Firstly, a major requirement for a good fit is of course the effective absence of a trend, i.e. the "stationarity" of the situation. In practice, it is unusual to observe *complete* stationarity, and most of the cases studied are ones of *near-stationarity* (e.g. changes in the sales-level m, in the penetration b, or in the rate of buying per buyer w, of not more than ± 10%, say). Such departures from strict stationarity will introduce some discrepancies between the observed and theoretical repeat-buying patterns, and this is one of the commonest causes for at least *minor* discrepancies between the theoretical and observed values.

Secondly, there is sampling error. With a panel of 1,000 households for example, and a "penetration" for the brand in question of perhaps 5% in the analysis-period, the sample base is only 50 buyers. However, because one is usually examining internal patterns within a given sample, the *effective* sampling errors tend to be reduced (for example, the sampling errors of the average purchase frequency w in one period and the incidence of repeat-buyers in the next period tend to be positively correlated and the sampling variations tend therefore to some extent to cancel out when estimating the one quantity from the other). Repeat-buying analyses are therefore more sensitive, and somewhat less subject to extraneous "noise", than might at first be thought. (Some sampling error formulae are discussed in §6.4).

Thirdly, discrepancies will arise if there are measurement errors in the data. Some case-history examples (especially for non-panel data) are given in Chapter 6.

Fourthly, there is the minimum time-period limitation to which reference has already been made earlier. Repeat-buying in time-periods of a length close to the shortest period between successive purchases in the product-field tends to be different. The minimum inter-purchase period is not necessarily something which can as yet be altogether rigidly defined, although for many grocery-products for example it tends to be a week. It is made up of more than one component, but basically, there is a tendency for purchasing acts to manifest themselves in discrete units of time like days or weeks, rather than in continuous time (as is assumed by the Poisson formulation of §4.5). This can be partly a matter of shopping habits – many people shop for certain types of goods only at specific times, e.g. first thing each morning for milk from the milkman, or on a certain day of the week for meat and groceries (on Tuesdays for some people, or Fridays for others, and so on), and only at still longer intervals for other products (see for example §5.3 in Chapter 5). This imposes a certain regularity on individual purchasing habits, as well as *non-stationary* aggregate behaviour in the short-term (e.g. by day of week).

A related short-term shopping pattern is that for any sub-sample of people who buy the given brand twice in a fortnight, almost all will buy it once in the first week and once in the next, and almost no one will buy it twice in one week (and not at all in the other). This is what generally occurs, as a matter of direct observation. In other words, the independent-random-events (or Poisson) hypothesis of §4.5 does not apply in such short time periods (one would expect only half the buyers buying once each week, with the other half buying twice either in week 1 or twice in week 2). The traditional pattern of buying cigarettes day-by-day is an extreme example of such *spreading out* of purchasing, with small, rather regular acts, instead of some bulk-buying. This leads to an abnormally *low* average purchase frequency per buyer (i.e. a low w) in each week, and to an abnormally *high* incidence of repeat-buying from one week to the next. The combined effect is to produce a very striking discrepancy between a high observed repeat-buying and a low theoretical estimate (from the low w) in short time-periods. (No theoretical estimates for 1-week periods were therefore given in Table 3.6.)

Linked to, but not necessarily identical with, such short-time shopping habits is the fact that there is often a "dead-period" between one purchase and the next, where the initial purchase has first of all to be more or less used up before another purchase is made.

These various short-term shopping and usage habits have little if anything to do with people's *longer-term* repeat-buying and brand-switching behaviour, i.e. those aspects of buyer behaviour with which brand-loyalty concepts are primarily concerned. The short-term patterns tend to "wash-out" in longer periods *. The basic finding is of course that the same NBD/LSD patterns tend to apply for periods of *any* length, as long as they are relatively long compared with the minimum inter-purchase interval, with only the numerical values in the pattern differing (predictably, as in §4.8). In longer time-periods, only the so-called "variance discrepancy" phenomenon tends to remain as a general discrepancy problem (i.e. a short-fall of heavy buyers who buy more than once a week – see §4.3). This is in fact also an indirect manifestation of short-term buying patterns and is further discussed in Chapters 7 and 8. Here we note that it does not usually affect the fit of repeat-buying formulae such as those in §4.7.

The fifth general type of discrepancy to be noted is that as already mentioned in §4.3 for the negative binomial frequency distribution as such, there is also some evidence – not yet altogether clear-cut – of excess regularity for certain very frequently bought items and for some

* For example, in a longer period, people who buy in just two successive weeks make up only a small proportion of all buyers, and the fact that their purchases are spread out over both weeks instead of some being bunched in one or the other week, has little effect on the overall purchasing patterns.

total product-classes (i.e. purchasing of *any* brand of detergent, or of margarine, or of petrol). Such purchasing may be somewhat more regular than the NBD/LSD model would imply, especially perhaps for apparently "saturated" markets with little or no growth-potential left. (The example treated in Chapter 3 is not such a case, the fit for the product-field – i.e. "Any Brand" – there being generally as good as for the individual brands, except possibly for the frequency distribution in Table 3.4.) In the earliest work, the study of repeat-buying was confined to each separate pack-size of an individual brand, and the results of aggregating different brands (or of aggregating the different pack-sizes of a brand) have only been examined more recently (on a "purchase occasion" basis – see § 1.3 of Chapter 1). The theoretical aspects of aggregating different NBD's are discussed in Chapter 7, but more empirical study is needed.

Finally, the wrong definition of the population of consumers is potentially another general factor which can lead to discrepancies. Thus for some products it is not very clear whether the "buyer" is a household or some specific individual. (This is quite apart from questions of measurement errors and biases, including vicarious purchasing, as when person X actually buys on behalf of person Y.) Furthermore some segments of the population may not be potential buyers of the product at all (or hardly at all), such as non-motorists for petrol and moter oil, non-owners of dogs or cats for pet foods, non-smokers for cigarettes, and so on. In other product-classes, the "never-buyer" may be less obvious to identify [see also Morrison 1969, Ehrenberg 1970a]. However, one outcome of the NBD/LSD theory has been to show that the precise definition of the population at risk does not usually affect the form of the observed repeat-buying patterns or the fit of the theoretical models, as long as the proportion of non-buyers is high enough. In the extreme case of the LSD version (i.e. where the proportion of buyers b has to be less than 20% or so for it to apply), the repeat-buying results in fact depend only on w, the average purchase frequency per *buyer*, and not on the number of non-buyers.

Apart from these more general types of discrepancies, more specific or isolated discrepancies tend also to occur. They are relatively rare, the analyses in Chapter 3 being typical in this respect (e.g. systematic but unusual discrepancies in Table 3.8, and otherwise a rather good fit apart from some seasonal non-stationarity). Such less general discrepancies are as yet of necessity largely unexplained.

Table 4.14 gives another example of a new, and hence as yet isolated and unexplained, discrepancy. Here the observed incidence of repeat-

Table 4.14. An Unusual Case of a REDUCING Incidence of Repeat-Buyers as the Length of Analysis Period Increases

(The percentage of buyers in one period who buy again in the next period, for periods of various lengths – Observed values "O" and Theoretical NBD norms "T" for the three leading brands in a certain product field X on pilot-scale data)

Product X	Period of length (in weeks)							
	1		2		5		10	
	O	T	O	T	O	T	O	T
Brand A	83	**73**	81	**80**	77	**85**	74	**88**
Brand B	85	**70**	83	**78**	80	**83**	77	**86**
Brand C	80	**73**	74	**79**	73	**84**	60	**86**
Average	83	**72**	79	**79**	77	**84**	70	**87**

buying *decreased* from an average of 83% in one-week periods to 70% in 10-week periods, whereas the theoretical values increased (as usual) from about 72% to 87% *. This is the only occasion on which such a tendency – systematic for all the brands in question – has so far been observed. The data come from a pilot-study and there is some reason to suspect measurement biases. On the other hand, the product-field is unusual, i.e. it is on the fringe of the kind of empirical conditions covered so far (see Table 4.2) in terms of frequency of buying and related characteristics. Further data would need to be collected (on a larger scale) to establish first of all whether this occurrence is repeatable, and if so, whether it is specific to the product-field or to the measurement procedure used.

This example serves to illustrate two major principles about discrepancies in the context of the present theory. Firstly, it is of course always necessary to check any new kinds of data against the theoretical model, to see if perhaps some additional factor is at play. And this is particularly important when empirical conditions are on the border-line of what has so far been covered – examples being perhaps very *frequently* bought products like bread, cigarettes or milk (as mentioned in § 4.3) or rather *infrequently* bought ones like clothing (as discussed in § 5.3 in the next chapter).

* The first thought with any discrepancy must always be of a clerical or computing error, e.g. that the observed data in this case were recorded the wrong way round with the 1- and 10-week results and the 2- and 5-week ones inter-changed (the fit would be good!). But checks showed that this was *not* the case here.

Secondly, the important thing with a discrepancy is not that the observed data differ from the theoretical model as such (let alone that the theory itself could be "disproved" by one such happening), but that the new data differs from the patterns observed in all the other cases studied so far where the model does give a good fit. Thus the conclusion about the buyer behaviour for the product in Table 4.14 is that it is different from the kind of behaviour generally observed in other product-fields.

4.10. Summary: The Nature of the NBD/LSD Theory

To summarise the results outlined in this chapter, there is a *single* theory – the NBD model, together with the simplifying LSD approximations – which applies to a very wide range of circumstances, i.e. different kinds of product-fields and brands, different lengths of time-period, and so on.

Table 4.15. Numerical Values of the LSD and NBD Formulae for the Percentage of Buyers of an Item in One Period Who Buy it Again in the Next Equal Period

(The values for $100b_R/b$ as given by Table 4.6, for various values of b and w)*

$100b_R/b$	Average Purchase Frequency per Buyer, w							
	1.1	1.3	1.5	2	3	5	10	20
NBD, for Proportions of Buyers, b =								
.8					85	89	92	93
.6				70	78	84	88	90
.4			52	64	74	80	85	88
.2**		37	47	60	70	77	83	86
.1		35	45	58	69	76	82	85
.01	16	34	44	57	68	75	81	85
LSD: Exact	16	34	44	57	68	75	81	85
Approximate	13	30	41	56	68	76	82	84

* Data for which $w < -\ln(1-b)/b$ cannot be fitted by an NBD.

** The percentage of repeat-buyers varies little for values of b less than .2.

The theory therefore is essentially a simple one, and the required input is also very simple – two observed quantities, primarily w, the average frequency of purchase of the item per buyer in some given period, and secondly b, the proportion of the population buying the item in that period. The way in which w is the dominant factor is illustrated in Table 4.15 in terms of the percentage of repeat-buyers from one period to the next, as given by the NBD and LSD formulae in Table 4.6 for various values of b and w.

One conclusion is that the observed patterns of repeat-buying loyalty are generally not intrinsic to the individual brand or product-field – it is not a case of "My brand (or my product) is different". Instead, repeat-buying patterns turn largely on the average purchase frequency of the item. Any two items with the same average purchase frequency will have the same repeat-purchasing patterns. Furthermore, as far as any two brands in a product-field tend to have similar w's, they will have repeat-buying characteristics which are similar not only in their general structure but also numerically *.

The underlying NBD/LSD model, and the empirically verified deductions from it, imply that there is generally no erosion of repeat-buying over time. Thus if there is no general trend in sales, the incidence of repeat-buyers in non-successive time-periods is virtually no lower than it is in successive periods. People who buy in one period but not in the other – so-called "lapsed" and "new" buyers – are not leaving the market altogether or coming in for the first time, but are simply *infrequent* (not necessarily *irregular*) buyers. The theory turns on the basic notion that different consumers have different long-run frequencies of purchase, and that this differing propensity-to-purchase manifests itself over time in a more or less random (i.e. stochastically regular) manner for each consumer.

Exceptions to this formulation can occur when there is a real turnover of buyers (the "leaky-bucket" theory) or when people have a special short- or medium-term enthusiasm for an item (the "jag buying" theory), but these *are* exceptions. (The "jag" type of exception seems for example to occur for the product-field analysed in Table 3.8 of Chapter 3, but – and this is the point – *not* in most other product-fields analysed, as is illustrated by Table 2.1 in Chapter 2 and Table 5.1 in Chapter 5.)

The simplicity of the repeat-buying model arises largely because the question of whether or not buyers of the item in question also buy any *other* item does not enter into any of the equations. This is essentially a matter of empirical fact – a good fit can be obtained by the NBD/LSD

* In contrast, the penetration b of different brands varies widely (and largely determines their different sales levels), but does not greatly affect their repeat-buying characteristics. The relation between b and w is discussed further in §§10.2, 11.4 and 11.5.

model as such, without making allowance for other brands. The reason behind all this is not that there is no multi-brand buying, as the reverse is generally true, but lies in the nature of multi-brand buying, as is discussed in Chapter 11.

All this is not to say that the NBD model is fundamentally "true". As has been pointed out already, the model breaks down at the boundaries, e.g. for very short time-periods and for very heavy buyers, and probably for very frequently-bought items and possibly for very infrequently bought ones. In particular, neither the Poisson- nor the Gamma-distribution assumption of the underlying stochastic model set out in § 4.5 can be altogether right. Some quite fundamental reformulation is required, as is discussed a little further in Chapter 11 and has in fact occurred in terms of the Dirichlet Model in Chapter 13, since the first edition of this book. But this must essentially give the same results and largely the same insights as the NBD theory in the vast majority of situations covered so far which the theory "models" rather successfully. The justification of the theory is therefore not the absolute truth of the theory in itself but that it works in practice and helps us to know and understand a good deal about empirical buyer behaviour and its essential simplicities.

PART III

PRACTICAL APPLICATIONS

CHAPTER 5

SOME PRACTICAL APPLICATIONS

5.1. Theory and Applications

In this and the next chapter, we illustrate several practical uses of the NBD/LSD repeat-buying theory. They are drawn from the increasing variety of applications that have been developed in the last few years.

The first application deals with the simple question of whether repeat-buying patterns in one country are like those in another. The second application probes possible extensions of the results for frequently-bought goods to semi-durable goods such as clothing; included here are also major problems with the available data.

Various cases of below-normal levels of repeat-buying are examined in § § 5.4 to 5.6. They relate to a new brand, to market segmentation, and to price-cutting.

These rather specific applications follow on from the more general application of the theory already illustrated in Chapter 3 of first of all understanding the general repeat-buying structure of any given product-field.

5.2. A Comparison of American and British Repeat-Buying Habits

The NBD/LSD theory was initially developed from British data. One task was therefore to see what repeat-buying was like elsewhere, and in particular perhaps in the U.S.

The existence of an empirically validated model was crucial to this purpose. Thus only after repeat-buying patterns in the U.K. had been found to be of the same general form for different brands, products, time-periods, and so on, was it plausible to suppose that repeat-buying elsewhere might take a sufficiently simple and general form to allow one to ask what "it" was like.

Furthermore, it was unnecessary to try and match specific American conditions to comparable British ones (e.g. types of product, frequency of purchase, length of analysis period, etc.). Instead, the easiest way to learn what American repeat-buying habits were like was to check whether they obeyed the British laws, i.e. the NBD/LSD theory as such. This simple use of a general model is one of the most basic forms of practical pay-off to be gained from successful theoretical work. It is much more efficient than reference back to raw data or the use of controlled experimentation.

The American data analysed here come from George Brown's [1952/3] pioneering studies of brand-loyalty. They consist of certain small sub-samples from the *Chicago Tribune* Consumer Panel in 1951 which Brown published in full, thus facilitating re-analysis. Four of the product-fields covered by Brown are reported on here, namely margarine, detergents, flour and regular coffee (selected as being least fragmented – bearing in mind Brown's small sample sizes – and non-seasonal). In these, 19 cases were identified (as listed in Table 5.1) which satisfied two criteria: first, the item in question was bought by at least ten households in each of two successive quarters of 1951 (10 was an arbitrary lower limit of sample size); second, the number of buyers and their average rate of buying were more or less steady or "stationary" over the two quarters.

To introduce the detailed findings, a typical example is given in Table 5.2. It is for Hills Bros. 1 lb. pack of regular coffee in Quarters I and II of 1951 (taken from halfway down the list of 19 cases in Table 5.1). In Brown's sample (exactly 100 buyers of coffee), there were just over 20 buyers of Hills Bros. 1 lb. pack in each of the two quarters. They bought nearly 60 packs, at an average frequency of 2.4 or 2.7 packs per buyer, as shown in Table 5.2*. The penetration and purchase frequency therefore showed little change from quarter to quarter.

The average quarterly buying rate of 2.55 packs per buyer for *both* quarters was used as input (averaged across the two quarters because of the small sample sizes). The NBD/LSD theory then leads to the following predictions about repeat-buying from quarter to quarter.

The number of repeat-buyers: There should have been 14 households who bought Hills Bros. 1 lb. size both in Quarter I and in Quarter II.

The average rate of buying per repeat-buyer: The households who bought the 1 lb. size in both Quarters I and II should on average have bought about 3.1 packs per quarter.

The proportion of total sales accounted for by repeat-buyers: In each quarter, these repeat-buyers should have accounted for 81% of the total purchases of the Hills Bros. 1 lb. size.

The average rate of buying by "new" or "lapsed" buyers: The households who bought the Hills Bros. 1 lb. size *either* in Quarter II only *or*

* The analysis reported here was carried out in terms of "number of packs bought" rather than "purchase occasions" (see discussion in §1.4 of Chapter 1 and §4.1 of Chapter 4).

Table 5.1. Estimated and Observed Repeat-Buying Statistics for 19 Near-Stationary U.S. Cases in 1951

(Theoretical Norms "T" and Observed Values "O")

Brand	Size	Product	Quarters	The given quarterly data: Number of buyers		The given quarterly data: Average packs bought per buyer		Repeat-buying statistics: Number of repeat-buyers		Average packs per repeat-buyer		Percent sales by repeat-buyers		Average packs per 1-Q-only buyer	
				1st Q	2nd Q	1st Q	2nd Q	T	O	T	O	T	O	T	O
Am. Fam. Flakes	20 oz.	Soaps/Suds.	I & II	21	21	4.3	5.0	16	14	5.6	5.6	92	81**	1.4	2.6**
All Sweet	½ lb.	Margarine	III & IV	47	53	4.6	3.9	37	36	5.1	5.2	91	90	1.4	1.6
Am. Fam. Flakes	20 oz.	Soaps/Suds.	III & IV	21	18	3.8	4.6	15	14	5.1	5.2	91	91	1.4	1.4
Hills Bros.	2 lb.	Coffee	II & III	12	12	3.3	3.3	8	8	4.0	4.3	87	89	1.4	1.3
Hills Bros.	2 lb.	Coffee	III & IV	12	15	3.3	3.2	9	9	3.9	4.3	87	90	1.4	1.5
Nutley	½ lb.	Margarine	I & II	18	17	2.9	3.0	12	12	3.6	3.7	85	85	1.4	1.4
Tide	52 oz.	Soaps/Suds.	III & IV	13	11	3.0	2.9	8	7	3.6	3.7	85	77	1.4	1.6
Hills Bros.	1 lb.	Coffee	I & II	24	21	2.4	2.7	14	15	3.1	3.2	81	85	1.4	1.2
Chase & Sanborn	1 lb.	Coffee	III & IV	14	16	2.4	2.6	10	9	3.0	3.4	81	82	1.4	1.2
Parkay	½ lb.	Margarine	III & IV	26	30	2.5	2.3	17	17	2.9	3.0	78	76	1.4	1.5
Eight O'Clock	1 lb.	Coffee	III & IV	17	18	2.1	2.6	11	10	2.8	3.2	77	78	1.4	1.2
Eight O'Clock	1 lb.	Coffee	II & III	18	17	2.4	2.1	10	9	3.1	2.8	76	70	1.4	1.4
Tide	52 oz.	Soaps	I & II	11	10	1.7	2.2	6	7	2.3	2.4	69	80	1.3	1.2
Pillsbury	5 lb.	Flour	III & IV	25	30	1.8	2.2	16	14	2.4	2.8	71	69	1.3	1.3
Pillsbury	5 lb.	Flour	II & III	26	25	1.7	1.8	13	15	2.1	2.1	66	68	1.3	1.4
Pillsbury	2 lb.	Flour	III & IV	14	12	1.9	1.7	7	7	2.1	1.8	64	59	1.3	1.4
Gold Medal	5 lb.	Flour	III & IV	18	18	1.4	1.4	7	10	1.7	1.6	49	62	1.2	1.2
Gold Medal	5 lb.	Flour	II & III	17	18	1.2	1.4	6	9	1.3	1.5	41	56	1.2	1.2
Average				20	21	2.7	2.8	13	13	3.3	3.4	77	78	1.4	1.4
Average discrepancy (1st Q–2nd Q)* or (T–O)*				2		0.3		1		0.2		5		0.2	

* Ignoring sign.

** See text.

Table 5.2. The Observed Quarterly Data for Hills Bros. Coffee 1 lb. Pack in Chicago. Quarters I and II, 1951

	Q.I	Q.II
Number of Buyers	24	21
Number of Packs bought	58	57
Packs per Buyer	2.4	2.7

Table 5.2a. The Estimated and Observed Repeat-Buying from Quarter I to II

(Theoretical Norms "T" and Observed Values "O")

	T	O
Number of Repeat-Buyers	14	15
Packs per Repeat-Buyer	3.1	3.2
Total Sales accounted for by Repeat-Buyers	81%	85%
Packs per "New" or "Lapsed" Buyer	1.4	1.2

in Quarter I only should on average have bought about 1.4 packs in the quarter in question.

These predicted figures are compared in Table 5.2a with the repeat-buying results tabulated directly from Brown's data for the two quarters. Clearly, the agreement is good. Thus the observed rate of buying per household who bought Hills Bros. 1 lb. size in both quarters was 3.2 packs, compared with the NBD/LSD estimate of about 3.1 packs. Again, there were 15 actual repeat-buyers in Brown's sample, compared with the estimated 14, and they accounted for 85% of the total purchases compared with the estimate of 81%. And so on.

The corresponding results for all of the 19 near-stationary cases are set out in Table 5.1. There is virtually no difference between the *average* observed and predicted values, and the discrepancies for the individual cases tend also to be small. For example, the number of repeat-buyers ranges from 36 to 7 sample households and is given by the theory to within about one household. The number of packs bought per repeat-buyer ranges from 5.6 down to 1.5 for the different cases and these figures are estimated to within an average of about 0.2 of a pack. And so on.

Some of the largest numerical departures from the theoretical results are the differences of 13 and 15 percentage points for the proportion of total sales accounted for by repeat-buyers in the last two entries in Table 5.1 (Gold Medal Flour, the 5 lb. pack). These discrepancies are essentially caused by there having been three more repeat-buyers than expected (a 15% difference for a total sample of about 20 buyers) and are in fact not statistically significant.

A more interesting discrepancy occurs in Quarters I and II for the 20 ounce pack-sizes of American Family Flakes in the "Soaps and Sudsers" field (the first item in Table 5.1). The observed percentage of sales accounted for by repeat-buyers is 11 points lower than would be normal, and the average rates of buying by "lapsed" and "new" buyers (2.9 and 2.4 – represented by their average of 2.6 in the table) are almost double the normal level of 1.4, a discrepancy which is well outside the limits of the other cases in Table 5.1 (or for stationary cases more generally).

Since the number of repeat-buyers and their average rate of buying are near the theoretical norms, it is the observed rate of buying by one-quarter-only buyers that is abnormal. Further analysis showed additional peculiarities in the data. *All* the "lapsed" and "new" buyers bought more than one pack, and only 5 of the 14 repeat-buyers bought just a single pack in one quarter or the other – a very unusual distribution. There was in fact a marked tendency in Quarters I and II for buyers of this brand to have purchased an *even* number of packs.

In Quarters III and IV (the third entry in Table 5.1), the average buying frequency of the "new" and "lapsed" buyers of American Family Flakes was however back to normal (i.e. 1.4), and there was also no longer any clustering of "even" purchases. This suggests that in Quarters I and II there may have been something like a manufacturer's offer of two 20-ounce packs physically "banded" together or sold at a reduced price. It would follow that this Quarters I and II discrepancy – about the only sizeable deviation from the theoretical norms that there was – was of a kind which would also have shown up in British data under similar marketing conditions*.

The conclusion from this study was therefore that American repeat-buying habits and British ones were the same. More recent U.S. data from the MRCA panel have widely confirmed this conclusion.

* If *purchase occasions* had been used as the analysis unit, the discrepancy for American Family Flakes in Quarters I and II would not have shown up in repeat-buying terms, but only as a special quirk in the average number of packs bought per purchase occasion. (We also note that American Family Flakes was generally marketed in a somewhat unusual way and was withdrawn from the market a few years later, but the Quarters III and IV results indicate that there was nothing *instrinsically* different about its repeat-buying patterns at the time.)

5.3. Repeat-Buying of Clothing

In the second case-history we move from "non-durable" branded food and other household products to clothing, i.e. "semi-durable" items most of which tend to be bought less frequently. We also move from full-scale panel data to an extremely limited set of data which had been obtained from single-interview sample surveys. This brings us up against major data problems. The study summarised here was an exploratory one*.

Prior to the study, Kemsley [1965] had already briefly reported that the frequency-distribution of purchases of clothing items in a given time-period tended to follow the Negative Binomial Distribution or NBD form. He had also mentioned certain measurement biases and other discrepancy problems in his data. The general tendency for the theoretical NBD-distribution to fit is confirmed here with data different from Kemsley's, but repeat-buying patterns in different time-periods and various discrepancy problems are pursued in more detail.

Most garments are bought relatively infrequently by any one person, at intervals of several weeks, months or even years. However, the main data available for analysis (a sample of 600 housewife interviews from the Continuing Clothing Survey carried out for ICI by Research Services Ltd) cover only very limited time-periods for any one informant's purchases: 2 weeks for stockings and 4 weeks for other garments. This restriction to a short time-span has limited the analysis that was possible and results will be discussed here only for the more frequently-bought items such as stockings, knitting-yarn, and socks, the patterns for the latter being similar to those for items such as panties and briefs and for other still less frequently-bought items.

A second difficulty with the data is that they suffer from measurement biases which occur in certain kinds of survey work: asking informants to recall their purchases over some previous period such as 2 or 4 weeks can lead to decreasing numbers of purchasing claims as the length of recall-period increases. The repeat-buying claims analysed here therefore represent a mixture of purchasing behaviour and measurement error which require separating out.

In discussing the results we consider firstly the extent to which people buy two or more items of the same kind in a single period such as a week. Secondly, we consider repeat-buying between two time-periods such as successive weeks – i.e. the proportion of buyers in one week

* Based on a report for ICI Fibres Ltd.

who buy also in the other week, and how much they buy. This work was undertaken before much was known about repeat-buying patterns for grocery goods in weekly time-periods (see §4.9 in the previous chapter), but in any case, the question here was simply to establish what the empirical repeat-buying patterns were like. Major questions are of course whether there are any generalisable patterns in these results, and how they are to be interrelated or explained.

Stockings. Starting with the most frequently-bought garment, stockings, Table 5.3 shows the extent to which buyers buy more than 1 pair of stockings per week. Two observed frequency distributions are shown, for the week immediately prior to the interview, and for the week before that*.

In the week immediately prior to the interview, 54% of the 600 women in question had acquired no stockings, 30% had bought 1 pair, 12% 2 pairs, and so on. This pattern is almost exactly in line with the theoretical NBD which can be fitted using the percentage of non-buyers, 54%, and the average rate of buying per buyer, 1.44 pairs, as input. The good degree of fit is clear to the eye and is also summarised by the equality of the standard deviations of the observed and theoretical distributions. The fit for the "last-but-one-week's" data in Table 5.3 is equally good. The numerical differences between the two weeks' results – 46% claiming an average of 1.4 purchases in one week and 35% claiming 1.3 purchases for the earlier week – reflect the length-of-recall period measurement bias in the data**. The fact that the NBD fits despite this bias suggests that the bias is not all concentrated at some particular level of purchasing frequency.

Turning to the repeat-buying between the two weeks, the NBD formulae say that given the observed average rate of buying 1.4 pairs per buyer in the last week before the interview, 50% of these buyers should also have bought stockings in the preceding week and that such "repeat-buyers" should each on average have bought about 1.5 pairs of stockings per week. These theoretical predictions agree quite well with the repeat-buying pattern actually observed for buyers in the week prior to the interview, as is shown in Table 5.4 (57% buying 1.4 pairs on average).

* This was dictated by the form of questioning used in the surveys (essentially of the form "When did you last buy . . .?" and "When before that?"), with interviewing carried out on Mondays and Tuesdays and purchase claims for the two preceding calendar weeks being analysed.

** The data are averaged over a number of different interviewing weeks and there was no marked trend in sales during this period.

Table 5.3. Weekly Purchasing Distributions of Stockings

(Observed Values "O" and Theoretical Norms "T")

Number of pairs bought	Week before interview: Last O	Last T	Last-but-one O	Last-but-one T
	%	%	%	%
0	54	(54)*	65	(65)*
1	30	31	26	26
2	12	11	7	7
3	2	3	1	2
4 +	2	1	1	1
Pairs per buyer	1.4	(1.4)*	1.3	(1.3)*
Standard deviation	0.9	0.9	0.8	0.8

* Fitted directly from the observed data.

However, when the number of repeat-buyers is expressed as a percentage of the (smaller) number of claimed buyers for the week before that, there is a large discrepancy with the theoretical figures. Thus 73% of the buyers in the last-week-but-one actually claimed also to have bought stockings in the following week, compared with the *theoretical* NBD estimate of only 43%*. The special interest of this observed excess of repeat-buyers is that it occurs only in the earlier week and therefore relates to the basic measurement bias in the data, such as the week-by-week drop of 11 percentage points (from 46% to 35%) in the number of buyers in Table 5.3. The repeat-buying results in Table 5.4 now indicate that this trend was not so much due to people who bought in both weeks but failed to claim the purchases they made in the earlier week, but to people who bought only in the *earlier* week but failed to report this. This possible explanation that *both* discrepancies – the excess of repeat-buyers and the week-by-week trend – are due to the same specific measurement error is something which could be tested experimentally in future work.

A second discrepancy from the general NBD pattern arises when we examine the frequency distribution of claimed purchases in the two weeks combined. This is shown in the "2-weeks" portion of Table 5.5.

* This estimate is based on the last-week-but-one's penetration of 35% and average purchase frequency of 1.3 (see Table 5.3); estimates based on the "last" week's figures would still be 50% repeat-buyers buying 1.5 times.

Table 5.4. Week-by-Week Repeat-Buying of Stockings

	Week before interview			
	Last		Last-but-one	
	O	T	O	T
No. of repeat-buyers / No. of buyers	57%*	50%	73%**	43%
Pairs per repeat-buyer	1.4	1.5	1.2	1.4

* % of buyers who had also bought in week *before.*
** % of buyers who had also bought in week *after.*

Here the theoretical NBD gives a marked excess of purchases of 1 pair of stockings in the 2 weeks (26% versus the observed 17%) and a deficit of purchases of 2 pairs (14% versus 27%). This is not a question of the 2 pairs of stockings being bought on the same purchase occasion because there is no such "excess" of 2 pairs in the 1-week distribution of Table 5.3. Nor can the known week-by-week measurement bias in the data be the whole explanation, since an "excess" of purchases of 2 pairs also occurred in some experimental data (again obtained by Research Services Ltd) covering a 6-week period but using a somewhat different measurement technique (1-week recall with a succession of six weekly interviews on the same informants), as is shown in the "6-weeks" portion of Table 5.5.

Having isolated this abnormally large group of people who regularly make one purchase in successive weeks, more experimental work would be required to explain it further*. Some questions are whether they tend to buy stockings of the same type each week (and whether the delay between the first and second purchase is due to uncertainty about the fit, or lack of spending money, or what), or whether they buy stockings of a different type each week (colour, make, price, etc.) and if so, why. Answers to such questions could affect practical merchandising decisions (such as 2-pair offers, and whether they should be of the same colour etc.).

Knitting-yarn. Next to stockings, hand-knitting-yarn is the most frequently bought textile item. The data cover the 4 weeks prior to each interview in the ICI Survey and are again subject to a length-of-recall-period bias, with fewer purchasing claims being made for earlier weeks than for ones closer to the interview (a trend from 6% buying to 13%).

* Subsequent work on full-scale panel data for grocery products has shown similar spreading out of purchases week by week, as already discussed in §4.9.

Table 5.5. Purchasing Distributions of Stockings in Periods of 2 and 6 Weeks

(Observed Values "O" and Theoretical Norms "T")

Number of pairs bought	Length of purchasing period			
	2 weeks**		6 weeks***	
	O	T	O	T
	%	%	%	%
0	45	(45)*	35	(35)*
1	17	**26**	22	**26**
2	27	**14**	22	**17**
3	6	**7**	11	**10**
4	3	**4**	5	**6**
5	1	**2**	2	**3**
6	1	**1**	2	**2**
7+	–	**1**	1	**1**
Average pairs per buyer	2.0	(2.0)*	2.3	(2.3)*
Standard deviation	1.4	**1.5**	1.8	**1.8**

* Used in fitting.
** Two-week recall.
*** Weekly recall in six successive interviews.

The most striking feature of the knitting-yarn data is that the incidence of repeat-buyers from one week to another is quite exceptionally high compared with the theoretical NBD result. This is illustrated in Table 5.6. Of the people claiming a purchase in the last week before the interview, 37% had also bought knitting-yarn in the preceding week, compared with a theoretical 25%. The results for the two earlier weeks are even more extreme still, in that for the last-but-third week before the interview, the percentage of the (low) number of buyers who also claimed to have bought in the following week is more than 40 percentage points higher than the theoretically expected NBD value (55% versus 12%).

A high incidence of week-by-week repeat-buying relative to the NBD norms might seem in line with what occurs for non-durable household products such as foods in short time-periods (see the "minimum time-period" discussion in Chapters 3 and 4), but it differs radically from what occurs for most other garments (see for example Table 5.9 for socks). The finding can hardly be explained by the measurement biases in the present data (as was the case for stockings in Table 5.4), since it occurs both forward and backward in time. Nor was it due to any sales trends. Instead, it seems to reflect something real for knitting-yarn.

The crucial point seems to be that we are dealing here with purchases of *yarn*, and not of completed garments. Thus there is known to be some tendency among consumers to buy the knitting-yarn needed for any garment in two or more instalments. This would explain exceptional high repeat-buying over the relatively short periods of time needed to complete a garment.

Table 5.6. Week-by-Week Repeat-Buying of Knitting Yarn

(Observed values "O" and Theoretical Norms "T")

	Pairs of Weeks before interview				Pairs of Weeks before interview			
	Last		Last-but-one		Last-but-two		Last-but-three	
	O	T	O	T	O	T	O	T
No. of repeat-buyers / No. of buyers	37%*	25%	48%**	38%	44%*	16%	55%**	12%
Oz. per repeat-buyer	1.2	1.3	1.2	1.2	1.1	1.2	1.1	1.1

* % buyers also buying in the week *before*.
** % buyers also buying in the week *after*.

Table 5.7. Quasi-Repeat-Buying of Knitting-Yarn in Periods of 2-weeks Lengths

	Pairs of Weeks before interview			
	(Last + first)		(Middle two)	
	O	T	O	T
No. of repeat-buyers / No. of buyers	44%	43%	40%	44%
Oz. per repeat-buyer	1.6	1.6	1.8	1.6

A test of this hypothesis is whether repeat-buying covering time-periods longer than weekly ones comes nearer to the "normal" NBD level under stationary conditions. The scope for such a test is very limited with only 4 weeks' data available per informant, and with the data moreover being subject to a trend due to measurement bias. However, grouping the purchasing claims for the second and third weeks on the one hand, and the purchases claimed for the first and fourth weeks on the other hand, leads to two periods each of 2-week length which are of a fairly stationary form (i.e. with about 15% of the informants buying in each such "period", at an average rate of roughly 1.5 oz. each) and for which the effect of any week-by-week instalment-buying is largely eliminated. This therefore provides a test, even if a rather artificial one. Table 5.7 shows that the "repeat-buying" patterns for these two 2-week periods in fact agree well with the appropriate theoretical NBD norms. This return to normalcy therefore suggests that for analysis-periods which do not relate to the instalment buying of knitting-yarn for any given garment, the "normal" NBD repeat-buying patterns might occur. But more direct study with better data is clearly needed.

Socks. The buying patterns for men's, boys' and girls' socks are similar, and we need only describe that for men. The pattern differs from the ones described above for stocking and knitting-yarn in two respects, but its major features are common to those of all *other* garments that have been examined (panties, briefs, shirts, blouses, etc.).

The observed weekly and fortnightly frequency distribution of claimed sock purchases are illustrated in Table 5.8. At first sight they seem in fair agreement with the theoretical patterns. However, there is one systematic discrepancy. Thus the observed distributions are in fact bimodal, there being a definite tendency for the number of purchases of 2 pairs of socks to exceed the number of single pairs being bought (e.g. 4% against 3% in the last week before interview). An "excess" of purchases of 2 pairs was, of course, also found for the two-week and

Table 5.8. Purchasing Distribution of Men's Socks

(Observed Values "O" and Theoretical Norms "T")

Number of pairs bought	Weeks before interview: Last		Last two	
	O	T	O	T
	%	%	%	%
0	92	(92)*	83	(83)*
1	3	4	6	9
2	4	2	8	4
3	1	1	2	2
4+	1	1	1	2
Pairs per buyer	1.9	(1.9)*	2.0	(2.0)*
Standard deviation	0.6	0.7	0.8	0.8

* Fitted directly from the observed data.

six-week purchasing of stockings (Table 5.5), although not in a single week (Table 5.3). But for socks, the excess *does* occur within a single week, i.e. generally on the same purchasing occasion. It might seem like commonsense that people will tend to buy 2 pairs of socks at a time, but in terms of quantitative detail, the question is whether it is commonsense (or even true) that *more* people should buy 2 pairs than buy 1 pair in a week. (This discrepancy appears to relate to the distinction between the *amounts* bought and the number of *purchasing occasions*, which has already been discussed in earlier chapters.)

Quantitatively more important than this discrepancy is the fact that the *week-by-week* repeat-buying pattern for socks which is shown in Table 5.9 is completely different from that for stockings, for knitting-yarn, and for non-durable household goods in general. Thus for other products which are bought at an average rate of about 2 units per buyer in one period, it has been found (as summarised by the theoretical NBD results) that a little over 55% of buyers in one time-period normally buy again in the adjacent period. For socks there is however an almost complete absence of repeat-buyers week-by-week (or fortnight-by-fortnight). Furthermore, the rare week-by-week repeat-buyers of socks that *do* occur buy at a rate of only 1 pair each, compared with the theoretically expected average rate of about 2.4. These findings are not unique to socks but occur also for garments such as panties and briefs, and for even less frequently-bought items such as knitwear, blouses, dresses, skirts, shirts and ties.

Table 5.9. Week-by-Week Repeat-Buying of Men's Socks

	Weeks before interview			
	Last		Last-but-one	
	O	T	O	T
No. of repeat-buyers / No. of buyers	4%	56%	4%	57%
Pairs per repeat-buyer	1.0	2.4	1.0	2.6

This radical departure from the NBD repeat-buying patterns does not seem to have anything to do with measurement biases. Instead, it fits in with commonsense experience that an item like socks is seldom bought in two successive periods as short as a week or so. The problem of what repeat-buying of socks is like in *longer* time-periods, and whether or not it then fits in with the general NBD type of patterns, cannot be answered directly because appropiate data are not available.

Relating the low repeat-buying for socks to the general repeat-buying theory discussed in Part II, we note that one of the main empirical findings summarised in this theory is that under stationary equilibrium conditions, repeat-buying follows a single pattern for time-periods of various lengths such as months, quarters or half-years; only the numerical values of the parameters vary with the length of the period. These parameter values are themselves interrelated, which is what gives the NBD theory its power for descriptive, evaluative and forecasting purposes. All this has been found to be true for a wide range of household goods, but generally of course only for time-periods of *at least* a week or so, as discussed in § 4.9 of Chapter 4. There was a minimum time-period effect, but this took the form that the observed repeat-buying from one week to the next was *higher* than the NBD prediction. However, *day-by-day* repeat-buying patterns of most grocery goods, toiletries, etc. are generally quite different from week-by-week or month-by-month ones. Few people who buy detergents, corned beef, or toothpaste on one day do so again the next day. A certain "dead-period" generally intervenes before any repeat-purchasing occurs. Such day-by-day repeat-buying (or rather the lack of it) is a reflection of *short-term* shopping and usage habits and does not relate to the longer-term factors of brand-loyalty, brand-switching etc. which really underlie the structure of a market and the effectiveness of marketing policies.

The indications from the results here are that the same considerations apply to the garment market, except that the "dead-periods" are longer. The very low *week-by-week* repeat-buying found for most garments (socks, etc.) may be very much the same kind of short-term phenomenon as the low *day-by-day* repeat-buying for most non-durable household goods. In contrast, the weekly results for stockings in Table 5.4 – a product like non-durable household goods in average length of usage life – and the (rather artificial) fortnightly results for knitting-yarn in Table 5.7 indicate that it may be possible for repeat-buying levels of textile garments in suitably long time-periods to be of the general NBD kind. It is therefore possible that repeat-buying would once again fit in with the NBD pattern once some *minimum* length of period is exceeded, depending on the product. The further study of purchasing patterns in the textile market and the practical problem-solving application of the results largely depend on the availability of more suitable data.

5.4. Low Repeat-Buying: A New Brand R

The preceding case-history contained several examples where the incidence of repeat-buying was low compared with the theoretical NBD estimates, either because of faults in the data or because the time-periods analysed were short compared with the minimum inter-purchase time for the item in question. We now consider three other cases of low repeat-buying (taken from unpublished reports), and start with the case of a certain new Brand R: repeat-buying of R was low, particularly in non-consecutive time-periods.

For a new brand, the ultimate question is not so much whether people will try it at all or even whether they will buy it a second time. Thus with good product research, promotional support and retail distribution, a new brand should find no great difficulty in getting quite a high proportion of its *potential* users to try it, and even to buy it a second or even a third time. Instead, the problem is whether a sufficient proportion of initial users will go on using the product more or less indefinitely.

In the present case-history, sales of a certain new Brand R had achieved a steady if unexciting level a year or so after its launch: about 4% of the population bought it each quarter, at an average frequency of about 1.4 purchases per buyer in the quarter.

The proportion of buyers of R in one quarter who bought R again in the next quarter was found to be 21%. This was relatively low, not only in absolute terms but also compared with a theoretical NBD/LSD norm in the high 30's (given 4% buying 1.4 times each in the first quarter). However, it was known that repeat-buying for new brands may not settle down to the "normal" level for an established brand for as long as a year or two [e.g. Ehrenberg and Goodhardt 1968d, Ehrenberg 1970b], there still being a high in-and-out flow of first-time triers. Brand R might well have been suffering from this.

The question therefore was whether the 20% quarter-by-quarter repeat-buying did in fact represent a solid core of long-term buyers (with a high turnover of additional new triers), or whether the product was merely sufficiently attractive to be bought just two or three times. This was explored by carrying out quarterly repeat-buying analyses for the new brand in *non-consecutive* quarters (as has already been illustrated for *established* brands in Table 3.9 of Chapter 3).

In the present instance, the analysis showed that there was in fact a serious erosion of the repeat-buying of Brand R in the longer run: of the buyers in one quarter, only some 11% bought again two quarters later, compared with 21% in the *next* quarter (and compared with a theoretical norm in the high 30's for established brands, whether in consecutive or non-consecutive quarters).

This form of analysis therefore showed that the new brand was failing to build up any sizeable following of buyers who would be loyal to it in the longer-term. The brand's steady sales level was due to its still continuing to attract new first-time buyers, but it was clearly working its way through the whole potential population in this respect, and sales would collapse before long. Management in fact withdrew the brand before this happened.

5.5. A Shortage of Repeat-Buyers or an Excess of Occasional Buyers?

Another case of low repeat-buying occurred in dealing with a certain food-product S. It was found that repeat-buying for virtually every brand was below the expected norm, by something like 7 percentage points on a quarter-by-quarter basis (and more half-year by half-year). There was no obvious explanation such as some large sales trend (i.e. non-stationarity) or the like. Repeat-buying for the product-field as a whole was almost exactly on the norm, so that the low repeat-buying for each individual brand was not a case of buyers moving in or out of the market as a whole.

Most brands were sold in three or four different pack-sizes, ranging from "Small" through "Medium" and "Large" to "Giant". The Large and Giant sizes had mostly been launched only two or three years previously, and the marketing management of the client company felt that the relatively low repeat-buying of each brand might be due to an as yet irregular buying pattern for these larger sizes.

Repeat-buying was therefore studied for each individual pack-size of each brand. The results are illustrated in Table 5.10 and showed differences in the repeat-buying level of the pack-sizes, but in precisely the opposite direction to that hypothesised by marketing management: repeat-buying of the Large and Giant sizes was virtually normal (42% versus 38%) and it was that for the two *smaller* sizes that was markedly below the theoretical level (37% versus 52%). Table 5.10 also shows that the discrepancy lay only in the sheer *proportion* of repeat-buyers, and not in their frequency of buying or in that of the "new" buyers in the second quarter (where the observed and theoretical rates still agree closely).

Table 5.10. Quarter-by-Quarter Repeat-Buying for Individual Pack-Sizes of Product S

(Observed Values "O" and Theoretical Values "T")

	% Buyers in one Quarter who buy in next Quarter		Av. frequency of purchase per repeat-buyer		Av. frequency of purchase per "new" buyer	
	O	T	O	T	O	T
Small Size	34	51	2.2	2.1	1.3	1.3
Medium Size	40	52	2.2	2.1	1.2	1.3
AVERAGE	37	52	2.2	2.1	1.2	1.3
Large Size	44	42	1.7	1.6	1.2	1.2
Giant Size	39	33	1.5	1.5	1.2	1.2
AVERAGE	42	38	1.6	1.6	1.2	1.2

The question next arose as to whether the discrepancy for the two smaller pack-sizes signified a real short-fall in repeat-buying or instead perhaps an excess of occasional buyers. The point is that the preceding analysis is *relative*, in that "too many" once-only buyers in one period would show up as an apparent *short-fall* of repeat-buyers in the next.

Table 5.11 illustrates the more detailed analysis required here, namely examining repeat-buying amongst previous light, medium and heavy buyers (a form of analysis already illustrated in Table 3.10 in Chapter 3). For the Small size (where in Table 5.10 there was a discrepancy of 17 percentage points in the observed and theoretical incidence of repeat-buyers), there was a marked short-fall of repeat-buyers amongst *light* buyers (i.e. ones who had made 1 or 2 purchases in the first quarter), whereas amongst *heavier* buyers (3 or more purchases in the first quarter) the incidence of repeat-buyers was on the mark. For the Medium size there was similarly a short-fall only amongst the initial once-only buyers.

Table 5.11. Repeat-Buying of the Small Size by Light and Heavier Buyers

		Purchases of the Small Size in One Quarter				
		0	1	2	3	4+
		%	%	%	%	%
% Buying Small size in the Next Quarter	O	3	27	48	73	89
	T	3	39	62	76	88

The apparent short-fall of repeat-buyers in this product-field was therefore tracked down firstly to the two smaller pack-sizes, and secondly even there not to any real failure to attract repeat-buyers, but rather to an excess of occasional buyers. This appeared to be related to very marked price-cutting competition for the smaller sizes (especially amongst Own Label brands) and possibly to patchy retail availability, two factors which led to increased brand- or size-switching on an *occasional* basis and hence an abnormally high number of light buyers.

5.6. Low Repeat-Buying and High Brand-Switching

In a certain non-food product-field M, repeat-buying for each brand was generally at the normal level, as illustrated in Table 5.12 for the four leading brands M_1 to M_4. An exception occurred for two particular brands, called M_5 and M_6 here: quarter-by-quarter repeat-buying was at least 10 percentage points below the expected NBD/LSD norms.

These two brands had a certain product-characteristic in common. This appeared to be exceptionally relevant to consumers, as shown in their brand-switching behaviour. Thus in the kind of brand-duplication and switching analyses that are introduced in Part V (as illustrated for example by Table 9.7 in Chapter 9), there was a very clear clustering of brands M_5 and M_6: switching between these two brands was at a substantially higher level than would be predicted from the general pattern of switching in this product-field.

Table 5.12. Quarterly Repeat-Buying in Product-Field M

(Observed Values "O" and Theoretical Norms "T")

	% Bought in next Quarter II	
Buyers in QI of:	O	T
Brand M_1 = 100%	70	63
Brand M_2 = 100%	60	58
Brand M_3 = 100%	51	52
Brand M_4 = 100%	58	58
AVERAGE	60	58
Brand M_5 = 100%	41	54
Brand M_6 = 100%	43	53
AVERAGE	42	53

Table 5.13. Quarter-by-Quarter Repeat-Buying and Brand-Switching between Brands M_5 and M_6

	In Next Quarter					
	% Buying the Same Brand		% Buying the Other Brand		Total: % Buying M_5 and/or M_6*	
Buyers in QI of:	O	T	O	T	O	T
Brand M_5 = 100%	41	54	16	5	57	59
Brand M_6 = 100%	43	53	11	4	54	57

* Double-counting 2 or 3% buying both.

This extra switching is shown in Table 5.13: the proportions of buyers of M_5 also buying M_6 or vice versa were almost 10 percentage points higher than the theoretical levels. This excess virtually made up for the short-fall of repeat-buying for each of the two brands, as is shown in the last column of Table 5.13 where the tendency for buyers of one of the two brands to buy *either* that brand *or* the other one was at virtually the predicted level (e.g. 57% versus 59% for M_5). There was therefore a definite tendency for consumers to treat brands M_5 and M_6 as substitutable for each other – it was more important to buy that particular *type* of brand (with the common product-characteristic) rather than a particular brand-name*.

The analysis therefore served to explain this particular short-fall in repeat-buying in "segmentation" terms, i.e. that there was a sub-group of brands which were similar in product-formulation and appeared to be treated as such by the consumer. Such special kinds of purchasing pattern however occur very rarely – repeat-buying of most brands follows the normal pattern, *without* having to take into account what other brands are or are not bought as well.

5.7. Summary

In this chapter, the NBD/LSD theory has been used to examine repeat-buying patterns under various previously unexplored conditions, such as for a different country, for a different type of product-class, and for data subject to major errors of measurement. This essentially involves comparisons of new kinds of data with previous results.

* This special degree of switching between the two brands may have been caused by uneven distribution and occasional lack of availability of one or the other brand, rather than by any positive breakdown of the "normal" intensity of repeat-buying for each brand. This possibility is so far unexplored in this particular market.

Such comparisons are facilitated by the use of a theory which effectively summarises previous empirical results and which therefore eliminates in these cases the need for controlled experimentation or for other matching procedures, or indeed the need for any direct recourse to the previous data in "raw" form.

The findings discussed show that repeat-buying in the U.S. takes the same form as in the U.K., and that the same repeat-buying patterns *may* also hold for semi-durable products like clothing in time-periods that are long enough to exclude the "dead-period" between one purchase and another (but further work on better data is needed here).

The analyses also show that usable information can sometimes be extracted from data which are subject to major measurement biases. Cases of abnormally low repeat-buying can also be tracked down to "real" factors such as the failure of a newly-launched brand to build up a repeat-buying franchise, an excess of occasional once-only buyers (due to price-cutting or patchy retail availability), or abnormally high brand-switching between brands which share a particular product-characteristic.