# LIST OF CONTRIBUTORS


| | |
|---|---|
| *Frank M. Bass* | The University of Texas at Dallas, USA |
| *Albert C. Bemmaor* | Ecole Superieure des Sciences Economiques et Commerciales (ESSEC) BP 105, 95021 Cergy Pontoise, Cedex, France |
| *Moshe Ben-Akiva* | Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA |
| *R. Carter Hill* | Department of Economics, Louisiana State University, Louisiana, LA, USA |
| *Siddhartha Chib* | Professor of Econometrics and Statistics, John M. Olin School of Business, Washington University, St Louis, USA |
| *Pradeep Chintagunta* | Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL, USA |
| *Jean-Pierre Dubé* | Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL, USA |
| *Dennis Fok* | Econometric Institute, Erasmus University Rotterdam, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands |
| *T. Fomby* | Department of Economics, Southern Methodist University, Dallas, TX, USA |
| *Philip Hans Franses* | Econometric Institute, Erasmus University Rotterdam, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands |

| | |
|---|---|
| *Heiko Großmann* | Westfalische Wilhelms-Universitat, Psychologisches Institut IV, Fliednerstr. 21, D-48149 Munster, Germany |
| *Heinz Holling* | Westfalische Wilhelms-Universitat, Psychologisches Institut IV, Fliednerstr. 21, D-48149 Munster, Germany |
| *Cheng Hsiao* | Department of Economics, University of Southern California, USA |
| *Daniel McFadden* | Department of Economics, University of California, Berkeley, CA, USA |
| *Alan L. Montgomery* | Carnegie Mellon University, Graduate School of Industrial Administration, 255A Posner Hall, 5000 Forbes Avenue, Pittsburgh, PA, USA |
| *Taka Morikawa* | Graduate School of Environmental Studies, Nagoya University, Nagoya, 464-8603, Japan |
| *Vicki G. Morwitz* | Leonard N. Stern School of Business, New York University, USA |
| *Lars Muus* | Department of Economics, Universitetsparken, 8000 Aarhus C, University of Aarhus, Denmark |
| *Richard Paap* | Rotterdam Institute for Business Economic Studies, Erasmus University Rotterdam, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands |
| *Leonard J. Parsons* | Georgia Institute of Technology, USA |
| *Jeffrey S. Racine* | Department of Economics BSN3403, College of Business Administration, University of South Florida, 4202 East Fowler Avenue, Tampa, FL, USA |

| | |
|---|---|
| *Hiek van der Scheer* | Department of Economics, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands |
| *Rainer Schwabe* | Eberhard-Karls-Universitat, Institut fur Medizinische Biometrie, Westbahnhofstr. 55, D-72070, Tubingen, Germany |
| *P. B. Seetharaman* | Assistant Professor of Marketing, John M. Olin School of Business, Washington University, St Louis, USA |
| *Vishal Singh* | Kellogg Graduate School of Management, Northwestern University, USA |
| *Shuba Srinivasan* | University of California, Riverside, USA |
| *Andrei Strijnev* | Doctoral Candidate in Marketing, John M. Olin School of Business, Washington University, St Louis, USA |
| *Baohong Sun* | Graduate School of Industrial Administration, Carnegie Mellon University, 255A Posner Hall, 5000 Forbes Avenue, Pittsburgh, PA, USA |
| *Udo Wagner* | Institute of Management, University of Vienna, Bruenner Strasse 72, A-1210 Vienna, Austria |
| *Tom Wansbeek* | Department of Economics, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands |

# ECONOMETRIC MODELS IN MARKETING: EDITORS' INTRODUCTION

Philip Hans Franses and Alan L. Montgomery

## INTRODUCTION

This volume of the research annual, *Advances in Econometrics*, considers the application of econometric methods in marketing. The papers were selected from submissions provided by authors in response to a call for papers after undergoing a peer-reviewed process. Although these papers represent only a small fraction of the work that is currently in progress in the field of marketing, they are representative of the types of problems and methods that are used within marketing. It is our hope that this volume will help to educate econometricians and marketers about the application of econometric methods that can both further the discipline of econometrics and the study of marketing. Furthermore, we hope that this volume helps foster communication between these two areas, and through this interaction advance the study of each discipline.

Marketing focuses on the interaction between the firm and the consumer. Economics encompasses this interaction as well as many others. Economics, along with psychology and sociology, provides a theoretical foundation for marketing. Given the applied nature of marketing research, measurement and quantitative issues arise frequently. Quantitative marketing tends to rely heavily upon statistics and econometrics. There is a rich history of marketing bringing in ideas from econometrics as exemplified by the recent special issue of the

**Econometric Models in Marketing, Volume 16, pages 1–9.**

**ISBN: 0-7623-0857-5**

*Journal of Econometrics* (Wansbeek & Wedel, 1999). For good introductions to marketing models see Leeflang et al. (2000), Lilien et al. (1992), and Hanssens et al. (2001). However, quantitative marketing can place a different emphasis upon the problem than econometrics even when using the same techniques. Consider the recent flurry of work in Bayesian modeling (for a survey see Rossi & Allenby, 2000). The focus of much of this work has been measuring heterogeneity, which in econometrics tends to be treated as a nuisance parameter; while in marketing can form the basis for personalized marketing strategies.

A basic difference between quantitative marketing research and econometrics tends to be the pragmatism that is found in many marketing studies. While theory is important and a guiding influence in research due to the discipline it can bring to a problem, at the heart of most marketing problems is a managerial problem that is foremost in the researchers mind. Therefore theory often is balanced against empirical concerns of being able to translate the research into managerial decision making. This pragmatism can benefit theory, since it can highlight deficiencies of the current theory and serve as a guide to developing new ones.

Another important motivating factor in marketing research is the type of data that is available. Applied econometrics tends to rely heavily on data collected by governmental organizations. In contrast marketing often uses data collected by private firms or marketing research companies. Table 1 provides a listing of various types of data and examples of each. Observational and survey data are quite similar to those that are used in econometrics. However, the remaining

***Table 1.*** Types of Data that are Commonly Used in Marketing Research and Examples of Each Type.

| Description | Examples |
|---|---|
| Observational | Advertising exposure data, Nielsen People meter used to monitor television viewing, Store Audit, Pantry Audit |
| Interview and Survey | Personal interviews, Computer aided interviews, Telephone interviews, Mail surveys |
| Panel | Commercial panels that monitor television usage (ACNielsen's Homescan), retail purchases (IRI), purchase and attitude (NPD), web usage (Jupiter Media Metrix) |
| Transactional | Point-of-sale purchases collected using bar codes scanners, Salesperson call reports, Warranty registration cards, Clickstream or Web access from server logs or ISP requests |

types of data, panel and transactional, can look quite different from what may be familiar to econometricians. The automation and computerization of much of the sales transaction process leaves an audit trail that results in huge quantities of data. A popular area for study is the use of scanner data collected at the checkout stand using bar code readers. These datasets can easily run into hundreds of millions of transactions for moderately sized retailers. Often techniques that work well for small datasets do not scale well for these larger datasets. Therefore scalability is a practical concern that is frequently overlooked.

Nor is technology likely to abate any time soon, as the recent wave of e-commerce applications has resulted in new sources of data such as clickstream data, that may be magnitudes of size larger than scanner datasets. Clickstream data provides a record of the movement of a consumer through a web site, which can be associated with their choice and purchase information (Montgomery, 2001). This is analogous to recording not just what a consumer purchases, but everything they considered, along with a record of the information shown to the consumer. It requires that we must think more integratively about consumer behavior, incorporating elements of knowledge, search, learning, and choice. The ability of this new technology provides a rich, potential resource for developing new insights into consumer behavior, as well as representing a new challenge to quantitative marketers and econometricians.

## OVERVIEW OF THE VOLUME

The chapters in this volume reflect current research in marketing research. We provide a listing of the chapters in Table 2, along with a description of the type of data used, methodology employed, and application considered. To help group the papers we choose the first dimension, the type of data employed, to order the papers. Starting with the finest level of data at the individual level, and ending with the most aggregate data. Within these segments the papers are in alphabetical order. We briefly discuss each of the papers in this volume.

*Stated Preferences and Revealed Choices:* Two key questions that marketers face are: what consumers want (or say they want) and what they effectively do. The research problem is that the answers to these two questions can diverge. Additionally, there are measurement issues about which design to use to analyze stated preferences and which type of marketing performance measure should be used to understand revealed preferences (say, sales versus frequency of purchase, for example). The recent explosion of available data also started serious thinking about how all these data should be captured in ready-to-use

***Table 2.*** Summary of Data, Methods, and Applications Considered by the Papers in this Volume.

| Author(s) | Data Type | Methodology | Application |
|---|---|---|---|
| Hsiao, Sun, and Morwitz | Stated Preferences and Revealed Choices | Discrete Choice Model | New Product Sales |
| Morikawa, Ben-Akiva, and McFadden | Stated Preferences and Revealed Choices | Discrete Choice Model and Linear Structual Equation | Travel Mode |
| Chib, Seetharaman, and Strijnev | Individual Purchase Incidence from Store Scanner | Multivariate Probit Model | Cross category pricing and promotion |
| Großmann, Holling, and Schwabe | Individual Choice from Survey | Optimal Experimental Design | Conjoint Analysis |
| Muus, van der Scheer, and Wansbeek | Individual Choice from Transactions | Probit Model | Direct Marketing |
| Racine | Individual Choice from Transactions | Non-parametric Models | Direct Marketing |
| Bemmaor and Wagner | Aggregate Store Scanner | Multiplicative Modeling | Sales Promotion |
| Chintagunta, Dubé, and Singh | Aggregate Store Scanner | Aggregation of Logit Choice Model | Brand Mapping |
| Fok, Franses, and Paap | Aggregate Store Scanner | Market Share Attraction Model | Pricing and Sales Promotion |
| Montgomery | Aggregate Store Scanner | Hierarchical Bayesian Modeling | Pricing and Sales Promotion |
| Bass and Srinivasan | Aggregate Sales | Nonlinear Modeling | New Product Sales |
| Parsons | Aggregate Sales | Stochastic Frontier Analysis | Salesforce Management |

and, perhaps more importantly, read-to-understand models. Indeed, it turns out that many marketing questions, combined with available marketing data, require the development of new methods and techniques. The first two chapters deal with questions related to reconciling stated preferences and revealed choices.

The need to forecast customer attitudes are quite prevalent in new product sales, where established trends and relationships cannot be observed. A direct technique to assess the potential sales of a product is to survey customers and ask their intention to purchase. Cheng Hsiao, Baohong Sun, and Vicki G. Morwitz consider several models that relate purchased intention to actual purchase behavior in "The Role of Stated Intentions in New Product Purchase Forecasting". They show that stated intentions can be biased and need to be scaled and modeled appropriately to achieve unbiased estimates of product purchases.

Taka Morikawa, Moshe Ben-Akiva, and Daniel McFadden consider the combination of stated and revealed preferences in "Discrete Choice Models Incorporating Revealed Preferences and Psychometric Data". The framework consists of discrete choice models which models reveal and stated preferences and a linear structural model that identifies latent attributes from psychometric perceptual indicators. The model is illustrated using choices of travel modes.

*Individual Choice:* A common theme in the next four chapters is the use of individual choice or incidence. All of the data considered come from transactions that the company engages in with the consumer, whether it is a purchase at a register or a record of shipment from a mail catalog. At the same time the methodologies employed are diverse reflecting the managerial application.

Siddhartha Chib, P. B. Seetharaman, and Andrei Strijnev present an "Analysis of Multi-Category Purchase Incidence Decisions Using IRI Market Basket Data". Typically, product choice within a category is considered independently. However, a purchase in one category may reduce the chance of purchase in a substitute category (e.g. refrigerated juice will reduce the chance of buying frozen juice), while purchasing in a complementary category may increase the chance of purchase (e.g. purchasing cake mix may increase the chance of purchasing cake frosting). The authors present an analysis of a high-dimensional multi-category probit model. They find that existing models underestimate cross-category effects and overestimate the effectiveness of the marketing mix. Additionally, their measurement of household heterogeneity shows that ignoring unobserved heterogeneity can have the opposite effect.

The chapter by Heiko Großman, Heinz Holling and Rainer Schwabe is about "Advances in Optimum Experimental Design for Conjoint Analysis and Discrete Choice Models". Marketing studies often have the ability to collect primary data through experiments, which is less common in econometrics. The authors review new developments in the area of experimental design and provide methods to compare these designs. This chapter gives a good overview of the material and rightfully draws attention to the importance of formally comparing designs.

Lars Muus, Hiek van der Scheer, and Tom Wansbeek present "A Decision Theoretic Framework for Profit Maximization in Direct Marketing". The managerial problem is to decide which addresses to select for a future mailing from a mailing list. In this problem the analyst must estimate the probability of a consumer responding. Often analysts ignore the decision context of the estimation problem, which can result in sub-optimal decisions. In this chapter the authors derive an optimal Bayes rule that considers parameter uncertainty when formulating a mailing strategy. This research illustrates the importance of the decision context.

Jeffrey S. Racine proposes a non-parameteric technique for predicting who will purchase from a direct mail catalog in "'New and Improved' Direct Marketing: A Non-parametric Approach" choosing who to send a catalog. Racine discusses and compares parametric, semi-parametric, and non-parametric techniques in this chapter. He finds that conventional logit and probit models perform quite poorly, while nonparametric techniques perform better.

*Aggregate Store Scanner Data:* The most common type of transactional data available to a retailer or manufacturer is sales data that is aggregated through time and reported at a store or market level. The next four chapters deal with issues related to modeling data derived from these sources. The general theme is that managers wish to extract information to make better pricing and promotional decisions.

The applied nature of many marketing problems brings the data to the forefront. Often data is not in a form that is consistent with economic theory. In "Estimating Market-Level Multiplicative Models of Promotion Effects with Linearly Aggregated Data: A Parametric Approach", Albert C. Bemmaor and Udo Wagner consider the estimation of market level data when the models are postulated at a store-level. Market level data is frequently encountered in practice, yet many researchers focus on finer level analyses. They propose a technique for creating aggregate level data that is consistent with multiplicative sales response models. This chapter addresses the aggregation problem that plagues many econometric models by suggesting that more appropriate indices and data measures may help to alleviate aggregation issues, rather than focusing upon the models themselves.

The chapter entitled "Market Structure Across Stores: An Application of a Random Coefficients Logit Model with Store Level Data" by Pradeep Chintagunta, Jean-Pierre Dubé, and Vishal Singh presents an econometric model based upon the logit brand choice model. They consider the aggregation of this model to the store level while accounting for price endogeneity. Their estimation approach yields parameters similar to those from household data unlike other aggregate data studies. The reason for this methodology is the easy availability of aggregate level data to retailer managers. This paper illustrates the emphasis that marketers place on visualization of the model to communicate the results to managers, such as the creation of brand maps to illustrate market structure.

A popular approach in the analysis of sales is through the analysis of market shares using an attraction model. Dennis Fok, Philip Hans Franses, and Richard Paap present an "Econometric Analysis of the Market Share Attraction Model". The authors consider issues concerning the specification, diagnostics, estimation, and forecasting of market share attraction·models. They illustrate this model with an application to supermarket scanner data.

In "Reflecting Uncertainty about Economic Theory when Estimating Consumer Demand", Alan L. Montgomery explicitly considers the fact that most economic theory is uncertain. Frequently an analyst will pretest a theory. If the test is accepted, the analyst proceeds under the assumption that the restrictions from the theory hold exactly. However, this procedure overstates the confidence in the estimates. On the other hand if the theory is rejected, even if it is approximately correct, then all information from the theory is discarded. Montgomery proposes a Bayesian model that allows the analyst to shrink a consumer demand model towards a prior centered over an economic theory. Both the analyst who holds to theory dogmatically or agnostically can be represented as extreme cases. More importantly, when prior beliefs fall somewhere in between, the model can borrow information from the theory even if it is only approximately correct, in essence the estimates are "shrunk" towards the theory.

*Aggregate Sales:* The final two chapters conclude by considering aggregate sales data. This data may occur at a very broad level, for example all the sales of clothes dryers in a given year, or monthly sales for a given market. The common theme in both of them is the desire to predict and control the underlying process.

Time series econometricians have been intently focused on the issue of spurious regression and the effects of cointegration. Frequently the cumulative sales of a new product follow an S-shaped trend. The Bass Model describes this commonly observed curve using a diffusion argument. Along with sales, price and advertising generally have a trend also. In "A Study of 'Spurious Regression' and Model Discrimination in the Generalized Bass Model", Frank M. Bass and Shuba Srinivasan consider the problem that coincident trends can have in identifying a nonlinear model. They compare different nonlinear models and consider how nonlinearity can acerbate the problems in model selection.

Leonard J. Parsons' chapter on "Using Stochastic Frontier Analysis For Performance Measurement and Benchmarking" is different from the other papers in this volume, in the sense that it is trying to bring existing econometric methods to bear on an important problem in marketing, namely how to assess performance. This chapter also illustrates the slow speed with which some econometric ideas take to be adopted into common marketing practice. Although refined over the years, stochastic frontier analysis originated in the 1960s and 1970s. The benchmarking problem is how to focus on the frontier or best performance and not the average performance of the salesforce. A key point is that standard regression techniques do not work well since the error term is truncated.

## CONCLUSIONS

The last two decades have witnessed an increasing interest in marketing to use quantitative data to address substantive questions using quantitative models. This interest arouses from a firm's ability to easily collect and store data on the actual and stated behavior of their current and prospective customers. Hence, it has become possible to identify causes and effects of marketing instruments and environmental variables.

The essential gain of combining marketing problems with econometric methods is that marketing problems might get solved using serious and well-thought methods, while on the other hand the econometrics discipline benefits from new methodological developments due to the specific problems. Hence, this combination is a two-sided sword, and we expect to see many more such developments in the future.

## ACKNOWLEDGMENTS

We would like to thank the editors of this series Thomas Fomby and Carter Hill for the honor of guest editing this volume, Lisa Muscolino the Publishing Editor of this volume at Elsevier, and the following reviewers: Peter Boatwright, Eric Bradlow, Sachin Gupta, Jeffrey Inman, Frenkel ter Hofstede, Christophe Van de Bulte, Richard Paap, Kristine Vlagsma, Jedid-Jah Jonker, Nanda Piersma, and Zsolt Sandor.

## REFERENCES

Hanssens, D. M., Parsons, L. J., & Schultz, R. L. (2001). *Market Response Models: Econometric and Time Series Analysis*. Vol. 12 of the International Series in Quantitative Marketing, Kluwer Academic Publishers, Boston, Massachusetts.

Leeflang, P. S., Wittink, D. R., & Wedel, M. (2000). *Building Models for Marketing Decisions*, Vol. 9 of the International Series in Quantiative Marketing, Kluwer Academic Publishers, Boston, Massachusetts.

Lilien, G. L., Kotlier, P., & Moorthy, K. S. (1992). *Marketing Models*, Prentice Hall, Englewood Cliffs, New Jersey.

Montgomery, A. L. (2001). Applying Quantitative Marketing Techniques to the Internet. *Interfaces*, *30*(2), 90–108.

Rossi, P. E., & Allenby, G. M. (2000). Statistics and Marketing. *Journal of the American Statistical Association*, *95*, 635–638.

Wansbeek, T., & Wedel, M. (1999). Marketing and econometrics: Editors' Introduction. *Journal of Econometrics*, *89*, 1–14.

# THE ROLE OF STATED INTENTIONS IN NEW PRODUCT PURCHASE FORECASTING

Cheng Hsiao, Baohong Sun and Vicki G. Morwitz

## ABSTRACT

*In this paper, we develop four models to investigate the role of intentions (stated and true) and explanatory variables in forecasting purchase based on the social psychology view that true intentions determine purchase behavior. We found that a weighted average of stated intentions together with the complementary FED variables are powerful indicators of future purchase behavior. For intention survey designers, these results imply that a conversion scale is needed to convert stated intentions to true intentions and intentions questions would yield more useful information if it is formulated in terms of probabilities rather than in terms of yes/no answers.*

## INTRODUCTION

It is routine for market research to collect purchase intention information. However, the relationship between purchase intention and subsequent purchase behavior has been controversial. On the one hand, Manski (1990, p. 940) maintains that "researchers should not expect too much from intentions data". On the other hand, Fishbein and Ajzen (1975, p. 50) claim that "intentions should always predict behavior, provided that the measure of intention

**Econometric Models in Marketing, Volume 16, pages 11–28.**

**ISBN: 0-7623-0857-5**

corresponds to the behavioral criterion and that the intention has not changed prior to performance of the behavior". Indeed, studies by Adams and Juster (1974), Gormley (1974), Juster (1966), Kalwani and Silk (1982), McNeil (1974), Morwitz and Schmittlein (1992), Penny, Hunt and Twyman (1972), Tauber (1975), Warshaw (1980), Morrison (1979), Infosino (1986), Bemmaor (1995), Fitzsimons and Morwitz (1996), Morwitz (1997), Young, DeSarbo and Morwitz (1998), Hsiao and Sun (1999), Sun and Morwitz (2000), etc. have shown positive associations between purchase intention and actual purchase with varying strength. Tobin (1959) has also examined whether intentions supplement or merely repeat the explanatory information contained in financial, economic, and demographic variables. His regression results show that intentions do contain information about future purchases, but they are not an adequate substitute for the demographic and economic variables.

In order to better use stated intentions collected from survey research for forecasting purchase, it is important to understand the role of intentions (stated and true) and explanatory variables in forecasting purchase. In other words, there is a need for studying whether intentions supplement or merely repeat the explanatory information contained in financial, economic and demographic (FED) variables. In this paper we use a panel survey of intention to buy a home PC data to empirically investigate the link between the stated purchase intentions and actual purchase behavior at the micro level.

In Section 2 we construct various models linking stated purchase intentions with actual purchase behavior. Section 3 describes the data and estimation of various models using PC panel data. Section 4 provides an empirical estimation of purchase intention model. Conclusions are in Section 5.

## THE MODELS

In this section we present a basic framework that links various observed phenomenon between stated purchase intentions and actual purchase behavior. Obviously, there are many more possibilities than the ones considered here. Our main concerns are the consistency with known psychological models and the simplicity to estimate models which are capable of generating good predictions.

### *Model 1*

In the first model, we follow Fishbein and Ajzen (1975) and assume that behavior is determined by intentions alone and intentions are determined by attitudes and social norms. In other words, if $S_i$ denotes the information available for consumer $i$, $I_i^*$ denotes the latent true intentions, and $y_i^*$ denotes the latent response, then

$$S_i \rightarrow I_i^* \rightarrow y_i^*. \tag{1}$$

If the relationships are linear, then we have

$$y_i^* = \alpha + \beta I_i^* + u_i, \tag{2}$$

and

$$I_i^* = \gamma + \underset{\sim}{\delta}' \underset{\sim}{x}_i + v_i, \tag{3}$$

where $\underset{\sim}{x}_i$ denotes the observed social demographic variables and $u_i$ and $v_i$ denote the effects of all other omitted factors which are assumed to be uncorrelated with $I_i^*$ and $\underset{\sim}{x}_i$.

Let $y_i$ be the observed binary variable indicating whether actual purchase happens ($y_i = 1$) or not ($y_i = 0$). Suppose

$$y_i = \begin{cases} 1, & \text{if} \quad y_i^* > 0, \\ 0, & \text{if} \quad y_i^* \leq 0. \end{cases} \tag{4}$$

Then the probability that $y_i = 1$ given $\underset{\sim}{x}_i$ and $I_i^*$ equals

$$\text{Prob}(y_i = 1 \mid I_i^*, \underset{\sim}{x}_i) = P(y_i = 1 \mid I_i^*) = F(\alpha + \beta I_i^*), \tag{5}$$

where $F$ is determined by the probability distribution function of $u$. If the observed stated intentions, $I_i$, equal the latent true intentions $I_i^*$, then

$$\text{Prob}(y_i = 1 \mid I_i, \underset{\sim}{x}_i) = \text{Prob}(y_i = 1 \mid I_i^*) = F(\alpha + \beta I_i). \tag{6}$$

### *Model 2*

Sometimes, respondents may not report their true intentions. For instance, one may be asked to give a degree of intention such as "How likely are you to buy product X in the next six months" on a 5-point intentions scale (definitely will buy = 5; definitely will not buy = 1) or on an 11-point scale (certain or practically certain = 11; no chance or almost no chance = 1; e.g. Kalwani & Silk, 1982). Or one may be asked to give a timed intent measures such as intend to buy in the next six months, in the next seven to 12 months, etc. There are findings indicating that there could be tendencies to overstate the high stated intentions and understate the low stated intentions at the time of the survey (e.g. Duncan, 1974; Lord & Stocking, 1976) in the multi-level intention measures. In the second model, we incorporate the existence of measurement bias and construct a true intentions index from stated multiple intentions measures $\tilde{I}_J = (\tilde{I}_{i1}, \cdots, \tilde{I}_{iJ})$. Suppose that the true intentions are a weighted average of some stated intentions scale $\tilde{I}$,

$$I_i^* = \sum_{j=1}^{J} \delta_j \tilde{I}_{ij}, \tag{7}$$

then

$$P(y_i = 1 \mid I_i) = F\left(\alpha + \sum_{j=1}^{J} \beta_j^* \tilde{I}_{ij}\right), \tag{8}$$

where $\beta_j^* = \beta\delta_j$.

*Model 3*

In the third model, we will recognize the imperfection of the constructed true intentions index in a binary response framework and combine the constructed intentions index with FED variables to predict the outcome. Sometimes, the survey questions elicit binary response to the intentions as opposed to scaled measures. For instance, the question may be "Do you wish to buy a certain product in the next so many months?". Juster (1966, p. 664) notices that "consumers reporting that they 'intend to buy within $X$ months' can be thought of as saying that the probability of their purchasing $A$ within $X$ months is high enough so that some form of 'yes' answer is more accurate than a 'no' answer". In other words, a consumer facing an intentions question responds as would a statistician asked to make a best-point prediction of a future event. If the observed intentions, $I_i$, take the form

$$I_i = \begin{cases} 1, & \text{if} \quad I_i^* > 0, \\ 0, & \text{if} \quad I_i^* \le 0 \end{cases} \tag{9}$$

then even though $P(y_i = 1 \mid I_i^*, x_i) = P(y_i = 1 \mid I_i^*)$, $P(y_i = 1 \mid I_i, x_i) \ne P(y_i = 1 \mid I_i)$ and $P(y_i = 1 \mid I_i, x_i) \ne P(y_i = 1 \mid x_i)$ since even under the assumption that $u_i$ and $v_i$ are independent,

$$\begin{aligned} E(y_i^* \mid I_i) &= \alpha + \beta E(I_i^* \mid I_i) \\ &= \alpha + \beta \cdot \int I_i^* \cdot f(I_i^* \mid I_i) dI_i^* \\ &\ne E(y_i^* \mid I_i^*) \end{aligned} \tag{10}$$

and

$$E(y_i^* | \underset{\sim}{x}_i, I_i) = \alpha + \beta\gamma + \beta\underset{\sim}{\delta}'\underset{\sim}{x}_i + E(v | \underset{\sim}{x}_i, I_i)$$
$$\neq E(y_i^* | \underset{\sim}{x}_i). \tag{11}$$

The last two inequalities in (10) and (11) follow from $E(v_i | \underset{\sim}{x}_i, I_i) \neq 0$ even though $E(v_i | \underset{\sim}{x}_i) = 0$. Suppose instead of a single measure $I_i$, but a multiple intentions measure $\underset{\sim}{\tilde{I}}_i\}$ are available, we may approximate the nonlinear relation of $E(y_i^* | \underset{\sim}{x}_i, \underset{\sim}{\tilde{I}}_i)$ by a stepwise function, we have

$$\text{Prob}(y_i = 1 | \underset{\sim}{x}_i, \underset{\sim}{\tilde{I}}_i) = F\left(\alpha^* + \beta^* \sum_{j=1}^{J} \tilde{I}_{ij} + \underset{\sim}{\gamma}^{*\prime}\underset{\sim}{x}_i\right). \tag{12}$$

*Model 4*

In addition to the issue of the presence of possible differences between stated intentions and actual intentions, many things can also happen between the time of survey and the time of actual purchase. Therefore, it may not be just actual intentions at the time of survey determine behavior, the shocks during the time frame of interest also determine the actual purchase. In the fourth model, we allow true intentions to shift over time. Suppose that because of the shock during the time frame of interest, there is a probability $\pi_i$ that an individual will buy and with probability $(1 - \pi_i)$ that an individual has no change in his purchasing probability given the intentions, then

$$P(y_i = 1) = \pi_i + (1 - \pi_i) \sum_{j=1}^{J} F\left(\alpha^* + \sum_{j=1}^{J} \beta_j^* \tilde{I}_{ij}\right), \tag{13}$$

where $\pi_i$ may or may not be predictable from observed socio-demographic variables, $\underset{\sim}{x}_i$. If they are, then we may write, $\pi_i = \pi(\underset{\sim}{\eta}'\underset{\sim}{x}_i)$.

All these models ((6), (8), (12) and (13)) postulate that true intentions determine purchase behavior ((1)). The difference in associations between the stated intentions and actual purchase found in empirical studies are attributed to the different behavioral relations between the stated intentions and the true intentions, i.e. intention response bias and shift of true intentions over time. Table 1 summarizes the various assumptions underlying our purchase decisions and observed variables. There is nested structure in the first three models, but not for Model 4. Model 1 is nested within Model 2; Model 2 is nested within Model 3. However, the models are not nested within Model 4, although technically it is possible by also including $\underset{\sim}{x}_i$ as additional explanatory variables in $F(\cdot)$.

***Table 1.*** Purchasing Decision Models Based on the Assumption that the Latent Response $y^*$ is a Function of Latent Response $I^*$, $y^* = \alpha + \beta I^* + u$.

| | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| Explanatory Variables Available | $I$, $\underset{\sim}{x}$ | $\underset{\sim}{\tilde{I}} = (\tilde{I}_1, \ldots, \tilde{I}_J)$, $\underset{\sim}{x}$ | $\underset{\sim}{\tilde{I}} = (\tilde{I}_1, \ldots, \tilde{I}_J)$, $\underset{\sim}{x}$ | $\underset{\sim}{I}$ or $\tilde{I} = (\tilde{I}_1, \ldots, \tilde{I}_J)$, $\underset{\sim}{x}$ |
| Relation between latent intention $I^*$ and observed explanatory variables | $I^* = I$ | $I^* = \Sigma\delta_j\tilde{I}_j$ | $I$ or $\tilde{I}_j = 1$ if $I^* > 0$<br>0 otherwise | $I^* = I$ |
| Relation between latent response $y^*$ and observed purchasing decision y | $y = 1$ if $y^* > 0$<br>0 otherwise | $y = 1$ if $y^* > 0$<br>0 otherwise | $y = 1$ if $y^* > 0$<br>0 otherwise | with probability $\pi$ $y = 1$<br>with probability $(1 - \pi)$<br>$y = 1$ if $y^* > 0$<br>0 otherwise |
| Prob $(y_i = 1)$ | $F(\alpha + \beta I)$ | $F(\alpha + \Sigma\beta^*\tilde{I}_j)$ | $F(\alpha^* + \beta^*\Sigma_{j=1}^{J} I_j + \delta^* x')$ | $\pi + (1 - \pi)F(\alpha^* + \Sigma\beta_j^*\tilde{I}_j)$ |

## DATA AND ESTIMATION

In this section we use a panel survey of PC data to empirically investigate which of the above models ((6), (8), (12) and (13)) is more likely to describe the discrepancy between stated intentions and actual intentions in the new product survey.

The panel survey collected information about intentions to buy a home PC, a relatively new durable good. The survey took place approximately every six months from July 1986 (wave 1) to January 1989 (wave 7). The panel assembled was designed to be representative of U.S. households. During each wave the panel household were asked their timed intent to buy a PC in the future. Because the specific intent questions in the first 2 waves were different from the last five and because wave 3 (July 1986) and wave 4 (July 1987) were one year apart rather than six months apart we only analyze data of wave 4 to wave 7.

The intent question during waves 4–7 reads:

"Do you or does anyone in your household plan to acquire a (another) personal computer in the future for use at home?
Yes, in the next 6 months
Yes, in the next 7 to 12 months

***Table 2.*** Variable Description.

| Variable Name | Variable Descriptions | Mean or Frequency |
|---|---|---|
| intention 1 | intend to purchase PC in the next 6 months | 7.28 |
| intention 2 | intend to purchase PC in the next 7 to 12 months | 13.49 |
| intention 3 | intend to purchase PC after a year | 29.9 |
| cars | number of cars | 1.64 |
| young | age <= 30 | 32.53 |
| education | household education | 4.19 |
| new-household | new household | 14.51% |
| upscale | upscale families | 21.93% |
| mid age-no kids | mid age with no children | 23.14% |
| professional | professional | 23.66% |
| clerical | clerical | 27.59% |
| working-hours | number of working hours of householder | 2.77 |
| male-head | household head is male | 78.58% |
| white-collar | household head is white collar | 34.62 |

Yes, in the next 13 to 24 months
Yes, sometime, but not within 24 months
No, but have considered acquiring one
No, will not acquire one"

In addition to the intent question, extensive demographic information such as the size of household, annual household income, age of head of household, marital status, home ownership, household stage of life, occupation, education of head of household, race, number of cars owned, regional dummy, whether any household members had ever used a PC at work or at school, etc, were also collected.

The survey did not ask for actual purchase. However, it contained a question of whether households owned a PC previously. Using this information, we construct a new PC purchase data by comparing whether a household switched from being a non-owner to an owner from one survey wave to the next among those households that had not previously owned a PC at home.[1] Then the actual purchases are the purchases made within six months after each survey. Since the measurements are more noisy when a respondent states s/he will make a purchase of PC after a year, we focus on the first two intention measures, i.e. intend to purchase within six months and seven to twelve months.

Our criteria of the choice of the models are the stability of the relationship and good predictive power. Stability is important because a major function of any econometric model is to sustain inferences from observed regularities to conjectured causal dependencies. Theoretical models consist of the logically valid implications. The empirical relevance of a theory follows from the correspondence conditions (or measurement equations) mapping latent relations onto observable relations (e.g. Hendry & Richard, 1982). Good predictive power is also important in any modeling process. In fact, Klein (1988, p. 21) argues that "a severe test for an economic theory, the only test and the ultimate test is its ability to predict" (see also Friedman, 1957; Geisser, 1980; Zellner, 1988). "The real proof of the pudding is whether it produces a satisfactory explanation of data not used in baking it – data for subsequent or earlier years" (Friedman & Schwarz, 1991).

Our modeling strategy for converting time intent measures to the true intention measure and the use of socio-demographic variables to predict intentions or to predict the probability that an individual has a change on his intentions is based on the availability of relevant measures and a progressive general-to-specific approach of Hendry and Mizon (1990) and Hendry and Richard (1982). We start with specifications of most general models allowed in the light of the data, and subject them to a sequence of econometric estimation and tests of significance, and end with a model that achieves the maximum of explanation with the minimum of factors that are consistent with theory.

Assuming that $F(\cdot)$ has a logit form, the cross-sectional maximum likelihood estimates for waves 4 to 6 and the pooled estimates are presented in Table 3.[2] These estimates use all available social demographic variables and treating the true intention as a weighted average of the intention to buy within six months (intention 1), the next seven to twelve months (intention 2) and more than a year (intention 3) from now. As one can see, many of the socio-

***Table 3.*** Estimation Results with Longer-term Intentions and More FED Variables (Based on Pooled Data).

| Parameters | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| constant | –2.905(22.27) | 3.221(10.45) | –4.915(1.27) | 3.358(7.69) |
| intention 1 | 2.254(9.26) | 2.954(5.66) | 4.039(5.25) | 0.398(0.39) |
| intention 2 | | 0.037(0.034) | 0.1331(0.10) | –1.96(0.34) |
| intention 3 | | –0.699(0.66) | –1.069(0.92) | –0.987(0.84) |
| cars | | | –0.142(0.49) | 0.267(0.85) |
| baby | | | –0.068(0.04) | |
| young | | | 1.311(0.86) | |
| old | | | 0.729(0.66) | |
| large size | | | 0.745(2.47) | |
| household head | | | 0.147(0.49) | |
| income | | | –0.049(0.59) | |
| new household | | | –2.85(0.93) | 0.734(1.21) |
| new baby boomers | | | –56.341(0.000) | |
| low/mid income | | | –0.465(0.39) | |
| upscale family | | | 0.962(0.85) | |
| low/mid income | | | 0.611(0.39) | 0.976(0.64) |
| elderly | | | –0.909(0.31) | |
| professional | | | –0.82(0.27) | |
| managers | | | –0.876(0.28) | |
| clerical | | | 0.204(0.18) | |
| sales | | | 1.32(2.48) | |
| other professions | | | –0.829(0.53) | |
| work-hours | | | 0.070(0.05) | |
| white | | | 1.260(0.52) | 1.680(1.02) |
| male head | | | 0.296(0.32) | |
| own | | | 0.681(1.56) | 0.152(0.98) |
| Log-likelihood | –311.259 | –299.965 | –238.609 | –338.532 |

*t*-statistics are reported in the parenthesis.

demographic variables are not statistically significant. So are the weights of intention 3. Moreover, the weight coefficients of the intention 3 variables are negative, contradicting one's prior conjecture. Given the highly unreliable longer term intention measurement, in what follows we shall focus on models using intention 1 and intention 2 dummies.

Tables 4–7 presents the cross-sectional maximum likelihood estimates for wave 4 to 6 and the pooled estimates of the models 1 to 4 using only intention

***Table 4.*** Estimation Results of Model 1.

| parameters | Wave 4 (518) | Wave 5 (397) | Wave 6 (384) | Pooled (1299) |
|---|---|---|---|---|
| $\alpha_0$: constant | –2.651(14.26*) | –2.953(12.22) | –3.287(11.64) | –2.905(22.27) |
| $\beta_1$: intention 1 | 1.957(5.47) | 2.078(4.64) | 3.025(5.97) | 2.254(9.26) |
| log-likelihood | –144.788 | –92.216 | –71.720 | –311.259 |

***Table 5.*** Estimation Results of Model 2.

| parameters | Wave 4 (518) | Wave 5 (397) | Wave 6 (384) | Pooled (1299) |
|---|---|---|---|---|
| $\alpha_0$: constant | –3.010(12.80) | –3.123(22.44) | –3.296(11.21) | –3.128(20.54) |
| $\beta_1$: intention 1 | 2.316(6.00) | 2.248(4.83) | 3.033(5.91) | 2.477(9.68) |
| $\beta_2$: intention 2 | 1.505(3.80) | 1.213(2.02) | 0.118(0.11) | 1.307(4.32) |
| log-likelihood | –138.474 | –90.544 | –71.714 | –303.486 |

***Table 6.*** Estimation Results of Model 3.

| parameters | Wave 4 (518) | Wave 5 (397) | Wave 6 (384) | Pooled (1299) |
|---|---|---|---|---|
| $\alpha_0^*$: constant | –4.394(5.04) | –3.400(5.60) | –4.356(5.54) | –4.034(9.87) |
| $\beta_1^*$: intention 1 | 2.402(5.91) | 2.275(4.85) | 3.112(5.84) | 2.513(9.65) |
| $\beta_2^*$: intention 2 | 1.560(3.68) | 1.265(2.05) | 0.163(0.15) | 1.339(4.37) |
| $\gamma_1^*$: upscale | 1.242(3.37) | 0.702(1.35) | 0.102(0.16) | 0.870(3.28) |
| $\gamma_2^*$: clerical | 0.930(1.29) | 0.341(0.37) | –0.078(0.07) | 0.510(1.10) |
| $\gamma_3^*$: male-head | 1.159(1.43) | 0.149(0.24) | 1.193(1.49) | 0.808(2.03) |
| log-likelihood | –130.232 | –89.552 | –70.282 | –295.042 |

***Table 7.*** Estimation Results of Model 4.

| parameters | Wave 4 (518) | Wave 5 (397) | Wave 6 (384) | Pooled (1299) |
|---|---|---|---|---|
| $\alpha_0^*$: constant | −3.479(15.53) | −5.862(14.20) | −45.267(12.28) | −3.864(27.81) |
| $\alpha_1^*$: intention 1 | 2.067(5.81) | 4.472(9.12) | 3.921(8.89) | 0.734(2.72) |
| $\alpha_2^*$: intention 2 | 1.235(3.21) | 3.344(4.19) | 0.716(0.68) | −2.036(1.26) |
| $\beta_0^*$: constant | −5.724(6.03) | −3.729(6.03) | −5.422(2.97) | −2.610(3.11) |
| $\beta_1^*$: cars | 0.620(4.27) | 0.502(3.22) | −0.041(0.15) | 0.474(1.34) |
| $\beta_2^*$: new-household | 3.633(6.98) | 4.468(8.90) | 0.950(0.87) | 2.660(1.06) |
| $\beta_3^*$: upscale | 2.668(6.61) | 3.195(8.10) | 0.062(0.07) | 1.115(1.14) |
| $\beta_4^*$: working-hours | −0.417(2.70) | −0.675(5.05) | 0.521(1.02) | −0.356(1.23) |
| $\beta_5^*$: male-head | 1.184(1.52) | 0.752(1.32) | 0.859(0.76) | 0.185(0.11) |
| log-likelihood | −138.971 | −133.695 | −70.170 | −346.572 |

1 and intention 2 and statistically significant socio-demographic variables. The coefficients of intention 1 and intention 2 are positive and significant (except intention 2 of wave 6 in Model 2, 3 and 4) indicating that respondents who state intentions of purchasing in the near future are more likely to purchase. And those who show intention to buy within 6 months are more likely to buy than those who show interest to buy within 6 to 12 months. A test of parameter constancy restriction yields a chi-square statistic of 5.0698 with four degrees of freedom for Model 1, 5.5082 with eight degrees of freedom for model 2, 10.685 with twelve degrees of freedom for model 3, and 7.0192 with eighteen degrees of freedom for Model 4. None of them are significant at 15% level. That is, we find that there is a remarkable stability between the relations of actual purchasing behavior and purchase intentions over time and the stated intentions variables are statistically highly significant. Thus, we conclude that intentions predict actual purchase.

With regard to the relations between the stated intentions and true intentions, the likelihood ratio test between pooled Model 1 and 2 yields a chi square value of 16.2 with one degree of freedom which is highly significant. In other words, the true purchase intentions to buy in the next six months is not completely represented by the stated purchase intentions to buy in the next six months. A better representation of the true intentions to buy for the next six months should be a weighted average of timed interest in the next six months and in the next seven to 12 months.

The likelihood ratio test statistics between models 2 and 3 has a chi-square value of 16 with 3 degrees of freedom which is significant at 1% level. That is, we do find complementarities between the intentions variables and socio-demographic variables, namely, having an upscale family, having a clerical job and having a male household head increase the probability of purchase. However, as argued in Section 2, the complementarities between the intentions and socio-demographic variables are because the response to an intentions questions is binary. Respondents will think as a statistician and state "yes" when they feel their intentions of purchasing the personal computer within X months is high enough. Then as shown by (12) intentions are not the single best predictors of actual purchase. FED variables are needed as supplementary information to predict purchase.

Model 4 and Model 1, 2 and 3 are not nested. In Model 4, we are not able to find socio-demographic variables that are highly significant in predicting the future shocks within the time frame of interest.[3] The socio-demographic variables that appear to be related to future shocks between the time of survey and actual purchase are number of cars owned, life cycle 1 (new household), life cycle 4 (upscale families), and number of working hours of householder employment. New households and upscale households are more likely to purchase. Families with cars and/or male household head are also more likely to purchase. The negative coefficient of working hours indicating people who do not spend a lot of time at home are less likely to purchase a PC for home use. However, the average estimated probability $\pi$ is about 0.05145 using wave 6 socio-demographic variables. It appears too high and contributes to an exaggerated projection of the actual purchase percentage in wave 7. If no socio-demographic variable is used, the estimated probability of $\pi$ is only 0.0015, indicating that if there is any shock during the time of frames of interest that had led to a change in behavior, it is extremely small.

We also compare the predictive performance of all four models. We use waves 4 and 5 to estimate the models. Then we use the wave 6 data with the estimated coefficients to predict the market average of actual purchase

$$\frac{1}{N}\sum_{i=1}^{N_6} P(y_i = 1 \mid \tilde{I}_{i,6}\, \underset{\sim}{x}_{i,6},\, \hat{\underset{\sim}{\theta}}). \qquad (14)$$

where $N_6$ denotes the total number of observations in wave 6, and $\hat{\underset{\sim}{\theta}}$ are the estimated values of $\underset{\sim}{\theta}$ based on waves 4 and 5.

The average percentage of households that purchased new PCs between wave 6 and 7 is 5.9896%. Model 1 predicts 7.4292%, model 2 predicts 6.8963%, model 3 predicts 6.7181%, and model 4 predicts 18.832%. Models 1 and model 4 predict poorly. Models 2 and 3 predict the market outcome within one percentage of error, with a slight edge for model 3. These results appear to support the hypothesis that intentions are powerful indicators of future purchase behavior. However, a conversion scale is needed to convert stated intentions to true intentions. Intentions questions formulated in terms of probabilities rather than in terms of yes/no answers are likely to be a more reliable indicator of true intentions.

# PURCHASE INTENTIONS MODELS

In this section, we try to relate observed purchase intentions with socio-demographic variables based on the assumption that the true intention is a function of these variables as postulated in (3).[4] Let $I_j$ denote intentions to buy

***Table 8.*** Estimation Results of Intention Model.

| parameters | Wave 4 (518) | Wave 5 (397) | Wave 6 (384) | Pooled (1299) |
|---|---|---|---|---|
| intention 1 | | | | |
| constant | −2.479(9.50) | −2.245(6.54) | −2.544(7.33) | −2.433(13.69) |
| cars | 0.135(1.30) | −0.028(0.19) | 0.259(2.08) | 0.128(1.84) |
| young | 0.829(2.75) | 0.776(1.92) | −0.007(0.02) | 0.601(2.84) |
| mid age-no kids | 0.314(1.15) | 0.863(2.73) | 0.390(1.20) | 0.505(2.93) |
| intention 2 | | | | |
| constant | −2.022(5.05) | −1.532(3.48) | −2.675(4.85) | −2.055(7.98) |
| cars | −0.097(0.93) | −0.160(1.27) | 0.032(0.28) | −0.078(1.20) |
| young | 0.439(1.70) | 0.068(0.20) | 0.625(2.12) | 0.393(2.35) |
| education | 0.153(2.02) | 0.138(1.51) | 0.252(2.44) | 0.175(3.52) |
| upscale | −2.265(1.15) | −0.616(2.19) | −0.475(1.59) | −0.425(2.79) |
| intention 3 | | | | |
| constant | −1.490(4.61) | −1.467(3.76) | −2.699(5.54) | −1.807(8.28) |
| young | 0.829(3.93) | 1.044(4.15) | 0.712(2.77) | 0.855(6.28) |
| education | 0.053(0.67) | 0.070(0.77) | 0.372(3.41) | 0.146(2.85) |
| upscale | −0.460(2.15) | −0.222(0.91) | −0.135(0.53) | −0.291(2.16) |
| professional | 0.332(1.44) | 0.388(1.42) | −0.182(0.68) | 0.200(1.37) |
| white-collar | −2.273(1.29) | −0.460(1.81) | −0.291(1.17) | −0.330(2.44) |
| intention 4 | | | | |
| constant | −0.820(7.92) | 0.518(5.61) | 0.453(5.00) | 0.093(1.78) |
| young | 0.960(5.19) | 0.704(3.75) | 0.414(2.35) | 0.648(6.35) |
| upscale | −0.587(3.22) | −0.548(3.64) | −0.146(0.99) | −0.380(4.38) |
| log-likelihood | −1527.27 | −1423.92 | −1398.51 | −4484.83 |

within the $j$th six months, 1, 2, 3 and 4. Since $I_j$ if formulated in terms of 'yes' or 'no' format as postulated in (4), we assume a conditional logit model

$$\text{Prob}(I_{ij} = 1 \mid x_i) = \frac{e^{\underline{\omega}_j' \underline{x}_i}}{1 + \sum_{p=1}^{4} e^{\underline{\omega}_j' \underline{x}_i}}, \qquad (15)$$

$p = 1, \cdots, 4$ represents the first four scaled intentions measures.[5] The maximum likelihood estimates of cross-sectional waves 4, 5, 6 and the pooled data are reported in Table 5. The likelihood ratio test for parameter constancy has a chi square value of 270.26 with 36 degrees of freedom, which is significant at the 1% level. We also use waves 4 and 5 data to estimate the coefficients and combine them with the wave 6 socio-demographic variables to predict the intentions response in wave 7. The actual percentages of those responding to purchase within six months, the next 7 to 12 months, 12 to 18 months, 19 to 24 months, sometime in the future and do not intend to buy are 3.2801%, 3.9894%, 4.9645%, 7.6241% and 71.365%, respectively. The predicted percentages are 5.4964%, 8.8652%, 12.81%, 34.885% and 37.493%, respectively. The prediction errors of intentions using socio-demographic variables are much bigger than the prediction errors of actual purchase using intentions data. This lack of stability between purchase intentions and the observed socio demographics variables could be because factors affecting individual purchase intentions are numerous and the observed variables fail to capture all of them. In other words, it is much more difficult to model purchase intentions behavior than to model actual purchase as a function of purchase intentions. The relations between actual purchase and intentions are much more stable and predictable than the relations between actual purchase and socio-demographics variables.

## CONCLUSIONS

In this paper, we develop four models to investigate the link between the stated intentions and purchase at the micro level based on the social psychology view that true intentions determine purchase behavior. We argue that the different strength of association between stated intentions and purchase or the complementarities between stated intentions and FED variables found in the empirical literature can be attributed to the discrepancy between the stated intentions and true intentions. The first model assumes stated intentions perfectly predict purchase. The second model takes into account measurement bias and uses true intentions to predict purchase. The third model is a binary intention response model in which the prediction power of FED variables are examined. The last model allows true intention to change over time due to the shift of FED variables. We then rely on the stability of estimation and good predictive power to select the model that can best describes the discrepancy between stated intentions and purchase.

We use a survey panel data of PC intentions to investigate the relationship between stated intentions and actual purchase at the micro level. (1) We find a remarkably stable relationship between intentions and purchase over time which indicates that intentions are powerful predictor of actual purchase. (2) The true intentions are not accurately represented by stated intentions. A better representation of the true intentions should be a weighted average of stated intentions. Thus, we find support of the psychometric literature that stated intention should be transformed into an estimate of the true intention. A converted stated intentions to true intention remains to be most reliable predictor of actual purchase behavior. (3) In addition, when stated intentions are measured in binary form, FED variables such as upscale family, clerical and male-head are complementary to intentions in predicting purchase. However, the complementarities between the stated intentions and socio-demographic variables can be attributed to a consumer facing an intentions question responding as would a statistician asked to make a best point prediction of a future event. (4) We have not found significant evidence of exogenous events that lead to change intention or behavior within the time frame of interest. In fact, if there are exogenous events that lead to a change in behavior between the time of survey and actual purchase, they cannot be predicted by the observed socio-demographic variables. (5) It is much more difficult to model intentions as function of socio-demographic variables than to model actual purchase as a function of intentions. In summary, we found that intentions are powerful indicators of future purchase behavior. True intention converted from stated intentions together with the complementary FED variables remains to be the most reliable predictor of actual purchase behavior for the data set we use here. Intentions questions would yield more useful information if it is formulated in terms of probabilities rather than in terms of yes/no answers.

When collecting intentions to predict purchase, it is probably advisable for the marketing researchers to formulate the intention questions in terms of probabilities instead of in terms of yes/no answers as this will probably reduce the discrepancy between stated intentions and true intentions. Our empirical analysis appears to confirm that the most powerful predictor of purchase is true intentions converted from stated intentions.

## NOTES

1. We have excluded repeated purchase because we cannot detect the purchase of an additional PC given the available information. The constructed measures obviously contain errors. Therefore, the conclusion we will draw is based on the assumption that measurement errors are independent of explanatory variables.

2. The logit model makes specific assumption about the probability density function $u$. Although non-parametric methods are available to estimate the parameters up to a scale (e.g. Manski, 1985), they cannot be used to generate prediction, which is our main focus. However, empirical analysis comparing parametric vs. non-parametric approaches appear to indicate the difference is minor (e.g. Newey, Powell & Walker, 1993).

3. As pointed out by a referee that the shocks that interfere with the intentions – behavior relation are more likely to come from externalities (e.g. new information regarding the category). However, if they affect all households in the same way, then $\pi$ will be a constant for all households in a given time, though may vary over time.

4. The impact of these variables may contain the impact of excluded socio-demographic variables that are collinear with the included variables. Excluding relevant collinear variables may create the problem of interpretation of the estimated coefficients, but will have negligible impact on prediction (e.g. Intriligator, Bodkin & Hsiao, 1996). Our interest here is in predicting the outcome rather than identifying individual impact of included explanatory variables.

5. As suggested by a referee, an alternative approach to model true intentions as a function of social-demographic variable is to employ a hierarchical Bayes approach and model the response parameters of intentions on behavior ($\beta$ or $\beta^*$) as a function of socio-demographic variable in the second level.

## ACKNOWLEDGMENTS

We wish to thank editors and two anonymous referees for helpful suggestions.

## REFERENCES

Bemmaor, A. C. (1995). Predicting Behavior From Intention-to-Buy Measures: The Parametric Case. *Journal of Marketing Research*, *32*, 176–191.

Gerard, A. F., & Thomas, J. F. (1974). Commentaries on Mcneil, Federal Programs To Measure Consumer Purchase Expectations. *Journal of Consumer Research*, *1*(3), 11–15.

Duncan, G. T. (1974). An Empirical Bayes Approach to Scoring Multiple Choice Tests in the Misinformation Model. *Journal of the American Statistical Association*, *69*, 50–57.

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

Fitzsimons, G. J., & Morwitz, V. G. (1996). The effect of measuring intent on brand-level purchase behavior. *Journal of Consumer Research*, *23*, 1–11.

Friedman, M. (1957). *A Theory of the Consumption Function*. Princeton: Princeton University Press.

Friedman, M., & Schwarz, A. (1991). An Econometric Analysis of U.K. Money Demand. In: M. Friedman & A. J. Schwartz (Eds), *Monetary Trends in the United States and the United Kingdom*; Alternative Approaches to Analyzing Economic Data. *American Economic Review, 81*(1), 8–49.

Geisser, S. (1980). *A Predictivistic Primer, in Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys* (pp. 363–382). Amsterdam: North Holland.

Gormley, R. (1974). A Note on Seven Brand Rating Scales and Subsequent Purchase. *Journal of the Market Research Society, 16*, 242–244.

Hendry, D. F., & Mizon, G. E. (1990). Procrustean Econometrics: Or Stretching and Squeezing Data. In: C. W. J. Granger (Ed.), *Modeling Economic Series: Readings in Econometric Methodology* (pp. 121–136). Oxford: Clarendon Press.

Hendry, D. F., & Richard, J. F. (1982). On the Formulation of Empirical Models in Dynamic Econometrics. *Journal of Econometrics, 20*, 3–33.

Hsiao, C., & Sun, B. (1999). Modeling Response Bias with an Application to High-tech Product Survey Data. *Journal of Econometrics, 89*, 1–2, 15–39.

Infosino, W. J. (1986). Forecasting New Product Sales From Likelihood of Purchase Ratings. *Marketing Science, 5* (Fall), 372–384.

Intriligator, M. D., Bodkin, R. G., & Hsiao, C. (1996). *Econometric Models, Techniques, and Applications* (2nd ed.). Englewood Cliffs: Prentice-Hall.

Juster, T. (1966). Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design. *Journal of the American Statistical Association, 61*, 658–696.

Kalwani, M. U., & Silk, A. J. (1982). On the Reliability and Predictive Validity of Purchase Intention Measures. *Marketing Science, 1*, 243–286.

Klein, L. R. (1988). The Statistical Approach to Economics. *Journal of Econometrics, 37*, 7–26.

Lord, F. M., & Stocking, M. L. (1976). An Interval Estimate for Making Statistical Inferences about True Scores. *Psychometrika, 41*, 79–87.

Manski, C. F. (1985). Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator. *Journal of Econometrics, 27*(3), 313–333.

Manski, C. F. (1990). The Use of Intentions Data to Predict Behavior: A Best-Case Analysis. *Journal of the American Statistical Association, 85*, 934–940.

McNeil, J. M. (1974). Federal Programs to Measure Consumer Purchase Expectations, 1946–1973: A Post Mortem. *Journal of Consumer Research, 1*, 1–10.

Morwitz, V. G. (1997). It Seems Like Only Yesterday: The Nature and Consequences of Telescoping Errors in Marketing Research. *Journal of Consumer Psychology, 6*(1), 1–30.

Morwitz, V. G., & Schmittlein, D. (1992). Using Segmentation to Improve Sale Forecasts Based on Purchase Intent: Which "Intenders" Actually Buy? *Journal of Marketing Research, 29*, 391–405.

Morrison, D. G. (1979). Purchase Intentions and Purchase Behavior. *Journal of Marketing, 43* (Spring), 65–74.

Newey, W. K, Powell, J. L., & Walker, J. R. (1990). Semiparametric Estimation of Selection Models: Some Empirical Results. *American Economic Review, 80*, 324–328.

Penny, J. C., Hunt, I. M., & Twymay, W. A. (1972). Product Testing Methodology in Relation to Marketing Problems. *Journal of the Market Research Society, 43*, 65–74.

Tauber, E. M. (1975). Predictive Validity in Consumer Research. *Journal of Advertising Research, 15*, 171–191.

Tobin, J. (1959). On the Predictive Value of Consumer Intentions and Attitudes. *Review of Economics and Statistics*, *41*, 1–11.

Warshaw, P. R. (1980). Predicting Purchase and Other Behaviors From General and Contextually Specific Intentions. *Journal of Marketing Research*, *17*, 26–33.

Young, M., DeSarbo, W. S., & Morwitz, V. G. (1998). The Stochastic Modeling of Purchase Intentions and Behavior. *Management Science*, *44*(2), 188–202.

Zellner, A. (1988). Bayesian Analysis in Econometrics. *Journal of Econometrics*, *37*, 27–50.

# DISCRETE CHOICE MODELS INCORPORATING REVEALED PREFERENCES AND PSYCHOMETRIC DATA

Taka Morikawa, Moshe Ben-Akiva and Daniel McFadden

## ABSTRACT

*This paper proposes a methodology for incorporating psychometric data such as stated preferences and subjective ratings of service attributes in econometric consumer's discrete choice models. Econometric formulation of the general framework of the methodology is presented, followed by two practical submodels. The first submodel combines revealed preference (RP) and stated preference (SP) data to estimate discrete choice models. The second submodel combines a linear structural equation model with a discrete choice model to incorporate latent attributes into the choice model using attitudinal data as their indicators. Empirical case studies on travel mode choice analysis demonstrate the effectiveness and practicality of the methodology.*

**Econometric Models in Marketing, Volume 16, pages 29–55.**

**ISBN: 0-7623-0857-5**

## 1. INTRODUCTION

Discrete choice models have been extensively used to analyze consumer's choice behavior in market research (e.g. Green et al., 1977; Gensh & Recker, 1979; Guadagni & Little, 1983; Carpenter & Lehmann, 1985; Winter, 1986; Gupta, 1988; Chintagunta, 1993). About 15 years ago we began to work on the idea of combining discrete choice with conjoint analysis and latent variable models (e.g. McFadden, 1986; Ben-Akiva & Boccara, 1987). The underlying idea is that consumer behavior could be analyzed in more detail using subjective data on preferences, perceptions and attitudes. This approach contrasts with the traditional treatment of consumer behavior, which regards the consumer as an "optimizing black box."

One way of describing the consumer decision process is shown in Fig. 1. In this diagram, ovals refer to unobservable or latent variables, while rectangular boxes represent observable variables. The relationship between the actual attributes of alternatives and observed behavior is represented by three groups of intervening factors: perceptions, attitudes and preferences. Perceptions are consumer's perceived values of attributes of alternatives which are usually influenced by his or her socioeconomic characteristics and market information, while attitudes are his or her subjective importance of attributes. Preference is also a latent factor and represents desirability of alternative choices, which is usually expressed by a *utility function*. Traditionally, the latent factors enclosed by the dashed line have been treated as the black box. Recently, Ben-Akiva et al. (1999) proposed an extended framework that includes more psycho-

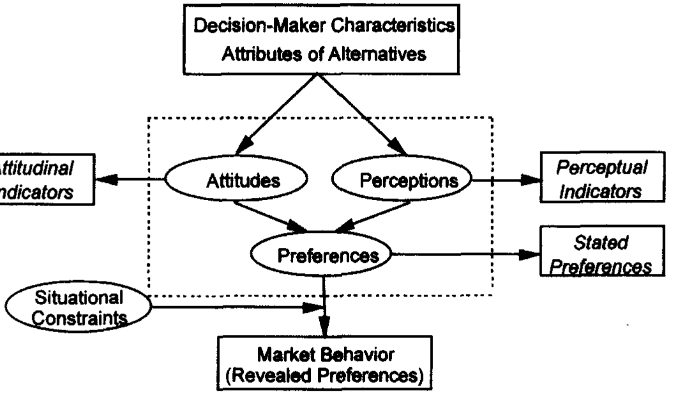


*Fig. 1.* Framework for Analysis of Consumer Behavior.

logical factors such as motivation, perceptions, tastes and attitudes and their indicators.

Market researchers have attempted to analyze explicitly the latent psychological factors and have relied on various indicators of perceptions, attitudes and preferences (Hauser & Koppelman, 1979; Louviere, 1988a; Lichtenstein et al., 1993; Meyer & Johnson, 1995). Attitudinal and perceptual indicators usually represent the level of satisfaction or importance of attributes on a semantic scale. Stated preference (SP) data are collected by presenting hypothetical scenarios to the respondents and asking for their preferences. In contrast to this type of data, measurements based on actual market behavior are termed revealed preference (RP) data.

In econometrics, however, the traditional view has been that valid choice data result only from actual choices having been made. Therefore, most econometric demand models are estimated using revealed preferences, measured attributes of alternatives and objective socio-economic characteristics of the decision maker. SP data, on the other hand, have been extensively used in market research (Green & Rao, 1971; Green & Srinivasan, 1978; Cattin & Wittink, 1982; Louviere, 1988b). These include the applications of conjoint analysis methods and more recently the discrete choice modeling techniques (e.g. Louviere, 1988a; Erlod et al., 1992; Louviere et al., 2001). SP data, which are collected in a fully controlled experimental environment, have the following advantages in contrast with RP data that are generated in natural experiments:

(i) they can elicit preferences for non-existing attributes and alternatives;
(ii) the choice set is prespecified;
(iii) multicollinearity among attributes can be avoided; and
(iv) range of attribute values can be extended.

Despite these advantages, SP data are not always considered to be valid for model estimation due to uncertain reliability of the elicited information under hypothetical scenarios. SP data may contain biases and large random errors if the decision making protocol exercised in a hypothetical situation differs from that exercised in a real choice context. Decision protocols for stating preferences about hypothetical scenarios can be observed in the following contexts:

(i) the respondent considers only the most important attribute of the alternatives (the prominence hypothesis);
(ii) the response is influenced by an "inertia" of the current actual choice (e.g. justification of the current choice);

(iii) the respondent uses the questionnaire as an opinion statement for his or her own benefit (e.g. overstating usage of a new transportation system to promote its construction);
(iv) the respondent does not consider situational constraints; and
(v) the respondent misinterprets or ignores an attribute if the attribute value lacks reality.

In addition, the reliability of SP data also depends on the quality of the questionnaire or the settings of the experiment. The response format such as "rating," "ranking," or "matching" in SP experiments also affects the reliability of the elicited preferences (For a more detailed discussion of these issues, see Ben-Akiva et al., 1991.)

Thus, SP and RP data have complementary characteristics. Simultaneously using both types of data with explicit consideration of unknown reliability of SP data may yield more reliable and useful consumer behavior models as exemplified in the following contexts. It is often the case that the trade-offs among certain attributes cannot be estimated accurately from the available RP data. For instance, high correlation between package size and price per unit in RP data may yield insignificant parameter estimates for their coefficients. However, SP surveys with a design based on low or zero correlation between these attributes may provide additional information on their trade-offs. Although the SP responses may not be valid for forecasting actual behavior due to their unknown bias and error properties, they often contain useful information on trade-offs among attributes. Another context where SP data add critically important information on preferences is the introduction of new attributes and new products. RP data alone cannot provide enough information to assess the impact of those drastic changes in services.

Other types of psychometric data such as attitudinal data have also been used in the choice modeling (Recker & Golob, 1976; Koppelman & Pas, 1980). It has been argued that consumer's choice behavior is determined by latent factors such as "quality" as well as manifest ones such as "size" and "price." Perceptual ratings of quality measures of alternatives, for instance, could be used as explanatory variables instead of their objective values in order to obtain better fit of the observed choice. However, this approach has always been criticized for little predictive validity when it is used for policy analysis.

The study presented in this paper is motivated by a question: How can we benefit from incorporating psychometric data, namely, stated preferences, perceptual and attitudinal data, in economic demand modeling? Our basic strategy is to use those psychometric data as indicators of the latent variables such as utility, attitudes, and perceptions in the behavioral framework depicted in Fig. 1. An answer to the criticism mentioned in the previous paragraph, for instance, is to use such perceptual data only as "indicators" of latent variables which themselves are the function of objective variables. In other words, incorporating such latent qualitative variables in econometric demand models requires some indicators of those variables as .well as the assumed causal relationship among them. In this sense, market behavior, or RP data, can be viewed as an indicator of a latent variable, i.e. utility, but market behavior is also the target variable to be recovered or predicted by the model.

This paper, hence, aims to propose a general framework for incorporating RP, SP, and other psychometric data in discrete choice models and to provide its practical estimation methods. Econometric formulation of the general framework is presented in the next section, which is followed by the two submodels with practical estimation techniques: the combined estimation from RP and SP data and the choice models with latent attributes. The paper also focuses on empirical analyses to assure the practicality of the methodology proposed. Sections 3 and 4 show the empirical works on the methodology developed in Section 2. An integrated model of the two submodels is estimated in Section 5. Concluding remarks are addressed in Section 6.

## 2. FRAMEWORK FOR COMBINING RP, SP, AND PERCEPTUAL DATA

### *2.1. Framework for Incorporating Psychometric Data in a Discrete Choice Model*

This section presents a general framework for incorporating psychometric data such as SP and perceptual data and econometric RP data in a discrete choice model. For the sake of simplicity, we use for presentation a binary choice model in which attributes are measured in terms of the differences of the two alternatives. Different response formats such as multinomial choice, ranking, and pairwise comparison will not change the general framework presented below as long as they are described by utility maximization behavior.

Suppose the following measurements are available from a questionnaire survey and/or an SP experiment:

(i) binary RP choice results;
(ii) binary SP responses;
(iii) perceptual indicators of some latent attributes of alternatives; and

(iv) observed attributes of alternatives and decision-maker's socio-economic characteristics.

The framework consists of two parts: a discrete choice model and a linear structural equation model; each part is composed of structural and measurement equations. RP and SP responses are described by the discrete choice models such as logit and probit while the relationship between perceptual indicators and latent attributes is described by the linear structural equation model.

Structural equations specify relationship between cause-and-effect variables. Since some cause-and-effect variables are not directly observable (e.g. quality and comfort), or latent, identifying these latent variables requires observable indicators. Measurement equations relate latent variables and their indicators. A typical latent variable is the utility in a discrete choice model. The framework proposed in this paper also allows latent attributes or perceptions in the diagram of Fig. 1. Latent attributes, for example, include "brand loyalty" and "quality" in brand choice applications and "convenience" or "comfort" in travel mode choice applications. In the equations below, asterisks (*) are attached to latent variables and superscripts "RP" and "SP" denote the corresponding data.

*Structural Equations*

$$u^{*\mathrm{RP}} = \mathbf{a}'\mathbf{x}^{\mathrm{RP}} + \mathbf{b}'\mathbf{w}^{\mathrm{RP}} + \mathbf{c}'\mathbf{w}^{*\mathrm{RP}} + \nu^{\mathrm{RP}} \tag{1}$$

$$u^{*\mathrm{SP}} = \mathbf{a}'\mathbf{x}^{\mathrm{SP}} + \mathbf{e}'\mathbf{z}^{\mathrm{SP}} + \nu^{\mathrm{SP}} \tag{2}$$

$$\mathbf{w}^{*\mathrm{RP}} = \mathbf{B}\mathbf{s}^{\mathrm{RP}} + \zeta^{\mathrm{RP}} \tag{3}$$

where

$u^*$ = latent utility;

$\mathbf{x}$, $\mathbf{w}$, $\mathbf{z}$ = vectors of observable explanatory variables;

$\mathbf{w}^*$ = vector of latent explanatory variables;

$\mathbf{s}$ = vector of observable variables that influence $\mathbf{w}^*$;

$\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{e}$, $\mathbf{B}$ = arrays of unknown parameters;

$\nu$ = random component of utility; and

$\zeta$ = vector of normally distributed disturbances.

*Measurement Equations*

$$d^{RP} = \begin{cases} 1, \text{ if } u^{*RP} \geq 0 \\ -1, \text{ if } u^{*RP} < 0 \end{cases} \tag{4}$$

$$d^{SP} = \begin{cases} 1, \text{ if } u^{*SP} \geq 0 \\ -1, \text{ if } u^{*SP} < 0 \end{cases} \tag{5}$$

$$\mathbf{y}^{RP} = \boldsymbol{\Lambda}\mathbf{w}^{*RP} + \boldsymbol{\varepsilon}^{RP} \tag{6}$$

where

$\mathbf{y}$ = vector of observed indicators of $\mathbf{w}^*$;

$\boldsymbol{\Lambda}$ = matrix of unknown parameters; and

$\boldsymbol{\varepsilon}$ = vector of normally distributed disturbances.

Equations (1), (3) and (4) construct the RP choice model and (2) and (5) form the SP choice model. The linear structural equation model is composed of (3) and (6).

This framework has two aspects from the viewpoint of statistical estimation. The first one is the combined estimation with RP and SP data (Morikawa, 1989; Ben-Akiva & Morikawa, 1990a, b) and the other aspect is the identification of latent variable $\mathbf{w}^*$ through a covariance structure model (Morikawa, 1989; Morikawa et al., 1990). These estimation methods are described as submodels in the following subsections.

### 2.2. *Submodel 1: Combined Estimation with RP and SP Data*

This submodel shown in Fig. 2 assumes two different data generating processes: The RP model represents actual behavior, while the SP responses are

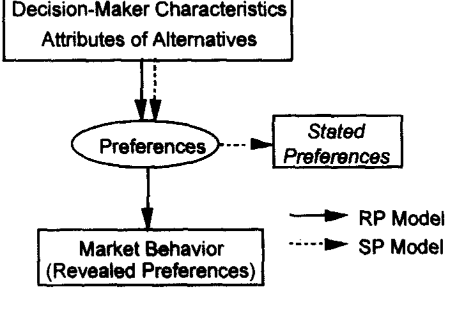


*Fig. 2.* RP/SP Combined Estimation.

modeled by the SP model. SP surveys are often conducted to obtain explicit and clear-cut information of trade-offs among attributes as well as direct preferences of non-existing services. One of the fundamental assumptions in conducting SP surveys is that the trade-off relationship among major attributes is common to both revealed and stated preferences. Otherwise, SP surveys themselves would have little meaning. In travel mode choice analyses, for instance, such attributes usually include line-haul travel time, terminal travel time, travel cost, and the number of transfers. We will denote these common attributes by the attribute vector **x** and its coefficient by the vector **a**.

The other factors affecting revealed and stated preferences are assumed to have different coefficients in RP and SP models. It is found from travel mode choice case studies that alternative-specific constants are likely to have significantly different values in both models (Ben-Akiva & Morikawa, 1990a, b). We denote such attribute vectors by **w** for the RP model and **z** for the SP model and their coefficient vectors **b** and **e**, respectively. The choice in the real market often affects SP as described in the previous section. It is sometimes called the justification bias or inertia effect and is captured by including the RP choice indicator $d^{RP}$ and its coefficient $f$ in the SP utility function.

The models used for the following presentation are also binary choice models and the latent attribute $\mathbf{w}^*$ is omitted for simplicity.

*The RP Model*

$$u^{*RP} = \mathbf{a}'\mathbf{x}^{RP} + \mathbf{b}'\mathbf{w}^{RP} + \nu^{RP} \tag{7}$$

$$d^{RP} = \begin{cases} 1, \text{ if } u^{*RP} \geq 0 \\ -1, \text{ if } u^{*RP} < 0 \end{cases} \tag{8}$$

*The SP Model*

$$u^{*SP} = \mathbf{a}'\mathbf{x}^{SP} + \mathbf{e}'\mathbf{z}^{SP} + fd^{RP} + \nu^{SP} \tag{9}$$

$$d^{SP} = \begin{cases} 1, \text{ if } u^{*SP} \geq 0 \\ -1, \text{ if } u^{*SP} < 0 \end{cases} \tag{10}$$

In the above modeling structure, sharing **a** in both models and estimating it by jointly using RP and SP data provides statistical efficiency. The terms represented by $\mathbf{e}'\mathbf{z}^{SP}$ and $fd^{RP}$ are specific to the SP model and may include SP biases. $\mathbf{e}'\mathbf{z}^{SP}$ also includes effects of hypothetical services that are included only in the SP survey. SP biases can be corrected from prediction by discarding from the fitted utility function the part of $\mathbf{e}'\mathbf{z}^{SP}$ and $fd^{RP}$ that represent the biases. If a part of $\mathbf{e}'\mathbf{z}^{SP}$ includes the effect of the new service included only in the SP questions, that part should be included in the fitted utility function for prediction.

Since the effect of unobserved factors may well be different between revealed and stated preferences, there is no reason for assuming that $\nu^{RP}$ and $\nu^{SP}$ have an identical distribution, or more specifically, have the same variance. Here we introduce a scale parameter $\mu$ that represents the ratio of standard deviations of $\nu^{RP}$ and $\nu^{SP}$, or

$$Var(\nu^{RP}) = \mu^2 Var(\nu^{SP}). \tag{11}$$

If SP data contain more random noise than RP data, $\mu$ will lie between 0 and 1. $\mu$ is also known to represent the "scale" of the model coefficients. The scale of the model is set by arbitrarily fixing the variance of the random utility term in order to identify the coefficients of a discrete choice model. For instance, assuming that $\nu^{RP}$ and $\nu^{SP}$ are normally distributed, the scale of the probit RP model is set to one (i.e. $Var(\nu^{RP}) = 1$), and the RP and SP models are:

$$P(d^{RP} = 1) = \boldsymbol{\Phi}(\mathbf{a}'\mathbf{x}^{RP} + \mathbf{b}'\mathbf{w}^{RP}), \tag{12}$$

and

$$P(d^{SP} = 1) = \boldsymbol{\Phi}(\,\mu(\mathbf{a}'\mathbf{x}^{SP} + \mathbf{e}'\mathbf{z}^{SP} + fd^{SP})), \tag{13}$$

where $\boldsymbol{\Phi}(\,)$ denotes the CDF of the standard normal.

If we can assume that unobserved factors are statistically independent between revealed and stated preferences, the joint estimators of **a**, **b**, **e**, $f$ and $\mu$ are obtained by maximizing the joint log-likelihood:

$$L(\mathbf{a}, \mathbf{b}, \mathbf{e}, f, \mu) = \sum_{n=1}^{N^{RP}} \log\{\boldsymbol{\Phi}[d_n^{RP}(\mathbf{a}'\mathbf{x}_n^{RP} + \mathbf{b}'\mathbf{w}_n^{RP})]\}$$

$$+ \sum_{n=1}^{N^{SP}} \log\{\boldsymbol{\Phi}[d_n^{SP}\mu(\mathbf{a}'\mathbf{x}_n^{SP}\mu(\mathbf{a}'\mathbf{x}_n^{SP} + \mathbf{e}'\mathbf{z}_n^{SP})]\}, \tag{14}$$

where $N^{RP}$ and $N^{SP}$ are the numbers of observations of RP and SP data sets, respectively.

Under the assumption of statistical independence of $\nu^{RP}$ and $\nu^{SP}$, or equivalently:

$$\text{Prob}(d^{RP}, d^{SP}) = \text{Prob}(d^{RP})\,\text{Prob}(d^{SP}), \tag{15}$$

then joint estimation yields consistent and asymptotically efficient and normal estimators. If this assumption does not hold, the joint estimators are still consistent and asymptotically normal but not fully efficient. In this case, a variance-covariance matrix of the maximal likelihood estimates calculated as the inverse of the information matrix will be biased. The joint estimation procedure requires special (but not complicated) programming due to the non-linearity in parameters. Hensher and Bradley (1993) proposed an estimation technique that utilizes a nested logit estimation software by creating an artificial nesting structure between RP and SP.

All the parameters can also be estimated sequentially. The sequential estimation procedure described in Ben-Akiva and Morikawa (1990b) avoids the non-linearity problem and can be carried out by MNL estimation software packages. The sequential estimators are consistent but not fully efficient.

The assumption of independence between the RP and SP error terms within the same individual may often be too strong. Furthermore, if the SP model has the RP choice indicator as an explanatory variable (as shown in (9)) and correlation between the error terms exists, there will be a problem of "state dependence and serial correlation" and, consequently, all the parameter estimates will be inconsistent. Morikawa (1994) proposed two approaches to remedy this problem. The first one is to include in the SP utility function the dummy variable that represents the RP choice. Such dummy captures unobserved preference factors for the specific alternative and, consequently, the remaining error term is less correlated with explanatory variables. The second is to explicitly consider serial correlation between the RP and SP utilities by splitting the error term into the alternative-and-individual-specific error and the white noise. Although this requires integrating the choice probabilities in computing the likelihood, the full information maximum likelihood estimator can be obtained.

### 2.3. *Submodel 2: Discrete Choice Model with Latent Attributes*

The idea of this submodel is to use psychometric data and the choice data as the indicators of some latent constructs. Using psychometric data as the indicators of latent variables is not new in psychology where the factor analysis is the most famous and basic approach (Johnson & Wichern, 1988). Structural equation models later made it possible to represent cause-and-effect relationship by using observable variables as covariates of the latent factors and have been widely applied in social and behavioral sciences (e.g. Goldberger, 1972; Duncun, 1975; Bielby & Hauser, 1977; Jöreskog & Sörbom, 1979; Bentler, 1980). Jöreskog and Sörbom (1984) developed computer software for

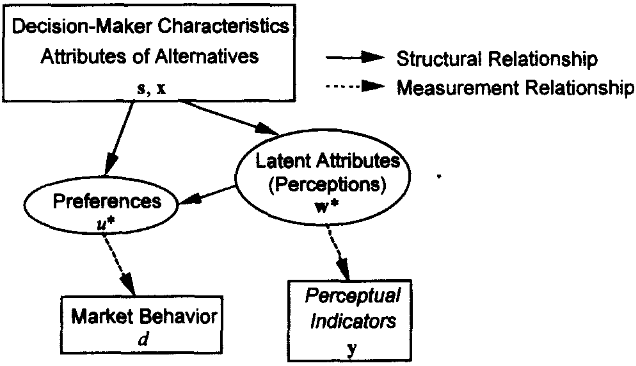


*Fig. 3.* Choice Model with Latent Attributes.

specifying and estimating structural equation models, known as LISREL (Linear Structural Relationships).

Submodel 2 is the combination of the structural equation model and discrete choice model. As shown in Fig. 3, the model system contains two types of latent variables and associated indicators. The first one is the intangible or latent attributes such as "beauty" and "novelty" that affect the choice behavior of interest. The indicators of these latent variables could be subjective answers of perceptual questions about the alternatives in the choice context. The utilities that represent latent preferences of the alternatives are the second type of latent variables. Their indicators are the choice in the real market or stated preferences.

This system can be formulated by combining the two existing modeling schemes: a discrete choice model and a structural equation model. Equations (16)–(19) are extracted for this system from the general framework presented in Section 2. We can see the two models there; one is a binary discrete choice model with latent attributes that consists of a structural Eq. (16) and a measurement Eq. (18), and the other is a linear structural equation model with latent variables that is composed of a structural Eq. (17) and a measurement Eq. (19). In the following presentation the SP model is omitted for simplicity.

*Structural Equations*

$$u^* = \mathbf{a}'\mathbf{x} + \mathbf{c}'\mathbf{w}^* + \nu \quad (16)$$

$$\mathbf{w}^* = \mathbf{B}\mathbf{s} + \boldsymbol{\zeta} \quad (17)$$

*Measurement Equations*

$$d = \begin{cases} 1, \text{ if } u^* \geq 0 \\ -1, \text{ if } u^* < 0 \end{cases} \tag{18}$$

$$\mathbf{y} = \mathbf{\Lambda w}^* + \boldsymbol{\varepsilon} \tag{19}$$

where

$\nu$ = random component of utility where $\nu \sim N(0, 1)$;

$\mathbf{z}$ = vector of normally distributed disturbances where $\boldsymbol{\zeta} \sim MVN(0, \mathbf{\Psi})$; and

$\boldsymbol{\varepsilon}$ = vector of normally distributed disturbances where $\boldsymbol{\varepsilon} \sim MVN(0, \mathbf{\Theta})$.

*2.3.1. Sequential Estimation Method*

Assuming all the variables are normally distributed, the choice probability is derived as follows. The joint distribution of **y**, **w*** and $u^*$ is

$$\begin{vmatrix} \mathbf{y} \\ \mathbf{w}^* \\ u^* \end{vmatrix} \sim \text{MVN}(\mathbf{M}_1, \mathbf{\Omega}_1), \tag{20}$$

where

$$\mathbf{M}_1 = \begin{vmatrix} \mathbf{\Lambda Bs} \\ \mathbf{Bs} \\ \mathbf{a'x} + \mathbf{x'Bs} \end{vmatrix} \quad \text{and} \quad \mathbf{\Omega}_1 = \begin{vmatrix} \mathbf{\Lambda\Psi\Lambda'} + \mathbf{\Theta} & \mathbf{\Lambda\Psi} & \mathbf{\Lambda\Psi c} \\ \mathbf{\Psi\Lambda'} & \mathbf{\Psi} & \mathbf{\Psi c} \\ \mathbf{c'\Psi\Lambda'} & \mathbf{c'\Psi} & 1 + \mathbf{c'\Psi c} \end{vmatrix}.$$

Here, given the observable variables **y**, **x**, **s**, the conditional distribution of **w*** and $u^*$ is

$$\begin{vmatrix} w^* \\ u^* \end{vmatrix} \sim \text{MVN}(\mathbf{M}_2, \mathbf{\Omega}_2), \tag{21}$$

where

$$\mathbf{M}_2 = \begin{vmatrix} \mathbf{Bs} + \mathbf{\Psi\Lambda'}(\mathbf{\Lambda\Psi\Lambda'})^{-1}(\mathbf{\Psi} - \mathbf{\Lambda Bs}) \\ \mathbf{a'x} + \mathbf{c'}\{\mathbf{Bs} + \mathbf{\Psi\Lambda'}(\mathbf{\Lambda\Psi\Lambda'} + \mathbf{\Theta})^{-1}(\mathbf{\Psi} - \mathbf{\Lambda Bs})\} \end{vmatrix},$$

and, defining $\boldsymbol{\omega} = \boldsymbol{\Psi} - \boldsymbol{\Psi}\boldsymbol{\Lambda}'[\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}]^{-1}\boldsymbol{\Lambda}\boldsymbol{\Psi}$,

$$\boldsymbol{\Omega}_2 = \begin{vmatrix} \boldsymbol{\omega} & \boldsymbol{\omega}\mathbf{c} \\ \mathbf{c}'\boldsymbol{\omega} & 1 + \mathbf{c}'\boldsymbol{\omega}\mathbf{c} \end{vmatrix}.$$

Hence, the choice probability of the discrete choice model given **y**, **x** and **s** is

$$\mathrm{P}(d\,|\,\mathbf{y}, \mathbf{x}, \mathbf{s}) = \boldsymbol{\Phi}\left( d\, \frac{\mathbf{a}'\mathbf{x} + \mathbf{c}'\{\mathbf{B}\mathbf{s} + \boldsymbol{\Psi}\boldsymbol{\Lambda}'[\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}]^{-1}(\mathbf{y} - \boldsymbol{\Lambda}\mathbf{B}\mathbf{s})\}}{\sqrt{1 + \mathbf{c}'\boldsymbol{\omega}\mathbf{c}}} \right). \tag{22}$$

Since the measurement equation of the choice model, (18), is non-linear, the whole system of equations (16) – (19) cannot be estimated simultaneously with an existing program such as LISREL (Jöreskog & Sörbom, 1984) but requires programming the likelihood. Instead, the two step estimation method described below will yield consistent but not fully efficient estimators.

*Step 1:* Use a LISREL type estimator to estimate (17) and (19) and calculate the fitted values:

$$\widehat{\mathbf{w}^*} = \hat{\mathbf{B}}\mathbf{s} + \hat{\boldsymbol{\Psi}}\hat{\boldsymbol{\Lambda}}'[\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Psi}}\hat{\boldsymbol{\Lambda}}' + \hat{\boldsymbol{\Theta}}]^{-1}(\mathbf{y} - \hat{\boldsymbol{\Lambda}}\hat{\mathbf{B}}\mathbf{s}), \tag{23}$$

$$\hat{\omega} = \hat{\boldsymbol{\Psi}} - \hat{\boldsymbol{\Psi}}\hat{\boldsymbol{\Lambda}}'[\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Psi}}\hat{\boldsymbol{\Lambda}}' + \hat{\boldsymbol{\Theta}}]^{-1}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Psi}}. \tag{24}$$

*Step 2:* Use a probit MLE to estimate the model of (22) using $\widehat{\mathbf{w}^*}$ and $\hat{\omega}$, namely, estimate **a** and **c** using the following choice probability:

$$\mathrm{P}(d\,|\,\mathbf{y}, \mathbf{x}, \mathbf{s}) = \boldsymbol{\Phi}\left( d\, \frac{\mathbf{a}'\mathbf{x} + \mathbf{c}'\widehat{\mathbf{w}^*}}{\sqrt{1 + \mathbf{c}'\hat{\boldsymbol{\omega}}\mathbf{c}}} \right). \tag{25}$$

### 2.3.2. *Simultaneous Estimation Method*

Assuming again all the variables are normally distributed, the choice probability can be derived as follows. The joint distribution of **y**, **x**, and $u^*$ is the same as (20). Given **w***, the conditional distribution of **y** and $u^*$ is

$$\begin{bmatrix} \mathbf{y} \\ u^* \end{bmatrix} \sim \mathrm{MVN}(\mathbf{M}_3, \boldsymbol{\Omega}_3), \tag{26}$$

where

$$\mathbf{M}_3 = \begin{vmatrix} \boldsymbol{\Lambda}\mathbf{w}^* \\ \mathbf{a}'\mathbf{x} + \mathbf{c}'\mathbf{w}^* \end{vmatrix} \quad \text{and} \quad \boldsymbol{\Omega}_3 = \begin{vmatrix} \boldsymbol{\Theta} & 0 \\ 0 & 1 \end{vmatrix}.$$

Then the joint probability of $\mathbf{y}$ and $u^*$ can be calculated by taking the mathematical expectation with respect to $\mathbf{w}^*$:

$$\Pr(d, \mathbf{y}) = \mathbf{w}^* \Phi\{d(\mathbf{a}'\mathbf{x} + \mathbf{c}'\mathbf{w}^*)\} \cdot \prod_{i=1}^{n} \phi \left| \frac{y_i - (\lambda_i \mathbf{w}^*)}{\theta_i} \right| \cdot \prod_{j=1}^{m} \phi\left( \frac{w_j^* - \mathbf{B}_j \mathbf{s}_j}{\psi_j} \right) d\mathbf{w}^*, \tag{27}$$

where $\phi$ denotes the PDF of the standard normal. In the above equation, we assume that the dimensions of $\mathbf{y}$ and $\mathbf{w}^*$ are $n$ and $m$, respectively, and that the components of $\mathbf{y}$ and $\mathbf{w}^*$ are independent among others. The maximum likelihood method is used to estimate the parameters to obtain consistent and asymptotically efficient estimates.

## 3. APPLICATION OF THE RP/SP COMBINED ESTIMATION

In this section an empirical analysis of the RP/SP combined estimation (submodel 1) is presented. The case is about intercity travel mode choice.

The survey was conducted during 1987 by the Hague Consulting Group for the Netherlands Railways to assess factors which influence the choice between rail and car for intercity travel. The City of Nijmegen, in the eastern part of the Netherlands, was selected as the data collection site. This city has rail connections with the major cities in the western metropolitan area called the Randstad which contains Amsterdam, Rotterdam and The Hague. Traveling from Nijmegen to the Randstad takes approximately two hours by both rail and car.

The home interview survey consisted of three parts:

(1) the characteristics of an intercity trip to the Randstad made within the previous three months (RP data);
(2) SP experiment of a choice between two different rail services (SP1 data); and
(3) SP experiment of a choice between rail and car (SP2 data).

The home interview survey was administered using lap-top micro-computers and the respondents replied to the questions appearing on the computer screen. The main advantage of a computer administered survey is that a desirable SP experimental design can be generated on site based on the service levels of the actual trip.

The RP data have 228 observations each including level-of-service attributes (e.g. travel time and cost), socio-economic characteristics (e.g. age and sex), and subjective ratings of latent travel characteristics (e.g. relaxation, reliability).

The SP experiments were framed in the context of the actual trip observed in the RP data and used the full-profile pairwise comparison method. The respondent was shown two hypothetical alternatives (two different rail services in the SP1 experiment and rail versus car in the SP2 experiment) at a time, each of which was described by the following four attributes: travel cost, travel time, the number of transfers (only for rail), and luxury level of the train (only for rail). Then, the respondent was asked which mode would be chosen *for the particular intercity trip reported in the RP question* in terms of a five point rating scale: (1) definitely choose alternative 1; (2) probably choose alternative 1; (3) not sure; (4) probably choose alternative 2; and (5) definitely choose alternative 2. Each respondent was presented with several pairs in SP1 and SP2 experiments. SP1 data (rail vs. rail) contain 2,875 comparisons (an average of 12–13 comparisons per respondent), while SP2 data (rail vs. car) include 1,577 comparisons (an average of 7 comparisons per respondent).

All the explanatory variables are in terms of differences between rail and car, more specifically, the values for rail minus the values for car. Socio-economic variables are included in the rail utility function.

A binary probit model estimated from the RP data is shown in the first column of Table 1. The second and third columns report the SP1 (rail vs. rail) and the SP2 (rail vs. car) models, respectively. The ordered probit models are applied to the ordered categorical responses described above with two threshold parameters, $\theta_1$ and $\theta_2$, representing four threshold values which are set to be symmetric with respect to zero.

The SP2 experiment was designed to collect information on mode switching behavior (from rail to car, or vice-versa) by presenting to the respondents hypothetical rail and car modes which are described by line-haul travel time and travel cost. However, since the respondent was instructed to refer to the trip reported in the RP questions, he or she may have considered additional attributes such as terminal time and the number of transfers that would have been required for the trip in evaluating the hypothetical alternatives. These additional attributes have the same values as reported in the RP questions. Thus, the model estimated from SP2 data includes these additional trip attribute variables which do not vary in the SP experiment. Characteristics of the traveler and the trip such as sex and purpose are also included. There may also be a bias in the stated preferences toward the mode actually used, reflecting the inertia effect, justification of past behavior, or omitted attributes that are not captured by the included variables. This bias can be estimated by including a dummy variable which indicates the actual choice.

***Table 1.*** Estimation Results of Submodel 1 (*t*-Statistics in Parentheses).

| | RP | SP1 | SP2 | RP + SP1 | RP + SP2 | RP + SP1 + SP2 |
|---|---|---|---|---|---|---|
| Rail constant (RP) | 0.501 | | | 0.455 | 0.702 | 0.718 |
| | (1.8) | | | (1.8) | (3.0) | (3.4) |
| Rail constant (SP) | | | −0.970 | | −3.82 | −3.82 |
| | | | (−9.8) | | (−4.0) | (−4.0) |
| Cost per person | −0.0270 | −0.0828 | −0.0111 | −0.0279 | −0.0338 | −0.0337 |
| | (−4.4) | (−25.4) | (−5.6) | (−5.2) | (−6.5) | (−6.8) |
| Line-haul time | −0.342 | −0.967 | −0.156 | −0.327 | −0.401 | −0.394 |
| | (−1.4) | (−11.6) | (−1.9) | (−4.9) | (−2.1) | (−6.1) |
| Terminal time | −1.61 | | −0.272 | −1.60 | −1.46 | −1.47 |
| | (−4.83) | | (−1.9) | (−4.9) | (−4.63) | (−4.77) |
| Number of transfers | −0.139 | −0.140 | 0.0433 | −0.0478 | −0.0348 | −0.0569 |
| | (−1.0) | (−4.3) | (0.8) | (−3.4) | (−0.3) | (−3.8) |
| Comfort | | 0.493 | | 0.166 | | 0.201 |
| | | (14.4) | | (4.9) | | (6.24) |
| Business trip dummy | 0.902 | | −0.115 | 0.887 | 0.358 | 0.363 |
| | (3.2) | | (−1.2) | (3.2) | (1.74) | (1.78) |
| Female dummy | 0.488 | | −0.102 | 0.488 | 0.230 | 0.232 |
| | (2.4) | | (−1.5) | (2.4) | (1.4) | (1.5) |
| Inertia dummy | | | 1.60 | | 5.68 | 5.70 |
| | | | (18.7) | | (4.7) | (4.8) |
| $\theta_{11}$ | | 0.0176 | | 0.0176 | | 0.0176 |
| | | (5.9) | | (5.9) | | (5.9) |
| $\theta_{12}$ | | 0.271 | | 0.271 | | 0.271 |
| | | (25.3) | | (25.3) | | (25.3) |
| $\theta_{21}$ | | | 0.0829 | | 0.0827 | 0.0827 |
| | | | (8.3) | | (8.4) | (8.4) |
| $\theta_{22}$ | | | 0.485 | | 0.484 | 0.484 |
| | | | (21.3) | | (21.6) | (21.7) |
| $\mu_1$ | | | | 2.97 | | 2.45 |
| | | | | (5.05) | | (6.5) |
| $\mu_2$ | | | | | 0.259 | 0.258 |
| | | | | | (4.9) | (4.9) |
| $\bar{\rho}^2$ | 0.243 | 0.321 | 0.377 | 0.319 | 0.368 | 0.339 |

Now, the RP data are combined with the SP1 data. The likelihood for the RP model is expressed by an ordinary binary probit model, while that for the SP1 model is expressed by the ordered probit model with threshold parameters. The fourth column of Table 1 shows the joint estimation results. All the estimates have the expected signs and have small standard errors. The scale parameter $\mu_2$

is estimated to be greater than 1, which indicates that SP1 data have less random noise than the RP data.

Then, the SP2 data are combined with the RP data. Note that the rail specific constants are separately estimated for each data, and the coefficient of the inertia variable is only estimated for the SP2 data. The fifth column of Table 1 shows the results of the joint estimation. All the coëfficients have the expected signs. The scale parameter $\mu_2$ is estimated between 0 and 1, which indicates a greater variance of the random utilities in the SP2 data.

Lastly, the RP data are combined with the two SP data and all the parameters are jointly estimated. As shown in the sixth column of Table 1, all the parameters are accurately estimated with the expected signs.

The first step in evaluating the usefulness of the combined estimator is to inspect the estimated coefficients of the separate RP, SP1 and SP2 models. A comparison of equivalent coefficients among these three models reveals large differences in the scales of the estimated utilities; the scale of the SP1 model is about 2.5 times greater than the scale of the RP model and the scale of the RP model is about four times greater than the scale of the SP2 model. This observation is verified by the results of the combined estimators. The ratio of the scale parameters of the SP1 and the RP models is given by $\mu_1$ with an estimated value of about 2.5. The ratio of the scales of the SP2 and the RP models is given by $\mu_2$ with an estimated value of about 0.26. These results indicate that the respondents were able to sharply discriminate between alternative rail services in the SP1 experiment. On the other hand, the stated choices between rail and car alternatives in the SP2 experiment were subject to significantly greater unexplained variance. Thus, a simple SP experiment such as SP1, may yield reliable information about trade-offs among attributes.

The most convincing demonstration of the important role that SP data can play in model estimation is provided by the estimated coefficient of the line-haul travel time variable. In the RP model this coefficient is too small and not significantly different from zero. (This is not an unusual occurrence in the estimation of mode choice models from RP data and may be due to the limited variability of the difference between car and train line-haul time.) In the SP models the coefficients of line-haul time have reasonable values and are significantly different from zero. Thus, a combined estimator that controls for the difference in scales yields a usable negative coefficient of approximately –0.4 which can now be used to predict the effects of changes in line-haul travel times.

The preference bias in the SP2 data toward the mode actually chosen was detected by the *inertia* variable. In the RP + SP1 + SP2 model, for example, the rail specific constant estimated from the SP data is –3.82 for car users and 1.88 ( = –3.82 + 5.70) for rail users. Thus, rail users have an SP rail constant of 1.88, which is greater than the RP value of 0.50, while for car users the SP rail constant is –3.82 and this is significantly smaller than the RP value. This indicates that car users have a greater preference bias toward their current mode than rail users. In other words, car users have a greater inertia or exhibit a greater justification bias than rail users.

# 4. APPLICATION OF THE CHOICE MODEL WITH LATENT ATTRIBUTES

This section presents an empirical case study for the second submodel: choice models with latent attributes. The Netherland travel survey data described in the previous section include the following subjective evaluation of trip attributes for both chosen and unchosen modes and they are used as perceptual indicators:

(i) relaxation during the trip (*relax*);
(ii) reliability of the arrival time (*relia*);
(iii) flexibility of choosing departure time (*flex*);
(iv) ease of traveling with children and/or heavy baggage (*ease*);
(v) safety during the trip (*safety*); and
(vi) overall rating of the mode (*overall*).

The first five perceptual indicators, (i)–(v) are described by five point ratings such as: (1) very poor; (2) poor; (3) neutral; (4) good; and (5) very good, and the overall evaluation of the mode is rated by a 10 point scale. These serve as **y** in (5) and are included in terms of the differences between rail and car.

Two latent variables, *ride comfort* and *convenience*, denoted by **w*** in (1) and (16) are specified as follows:

*ride comfort*:

$$w_1^* = \beta_1 aged + \beta_2 first + \beta_3 lhtime + \beta_4 aged \times lhtime + \zeta_1, \quad (comfort) \qquad (28)$$

*convenience*:

$$w_2^* = \beta_5 aged + \beta_6 trmtime + \beta_7 xfern + \beta_8 freepark + \zeta_2, \qquad (convenience) \quad (29)$$

where

*aged* 1 if the traveler is 40 years old or older, 0 otherwise;

*first*: 1 if the traveler uses the first class by rail, 0 otherwise;

*lhtime*: line-haul travel time by rail less line-haul travel time by car (hours);

*trmtime*: terminal time by rail less terminal time by car (hours);

*xfern*: the number of transfers by rail; and

*freepark*: 1 if free parking is available by car, 0 otherwise.

Since all the observable variables are measured in terms of the differences between rail and car, the two latent variables should also be interpreted as the differences between rail and car.

The relationship between these two latent variables and psychometric indicators, **y**, is described by the following measurement equations:

$$\begin{bmatrix} y_1(relax) \\ y_2(relia) \\ y_3(flex) \\ y_4(ease) \\ y_5(safety) \\ y_6(overall) \end{bmatrix} = \begin{bmatrix} 1 & \lambda_{12} \\ \lambda_{21} & 1 \\ 0 & \lambda_{32} \\ 0 & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} & \lambda_{62} \end{bmatrix} \begin{bmatrix} w_1^* \\ w_2^* \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} \tag{30}$$

The utility function that is the structural equation of the choice model part is specified as follows:

$$u^* = a_0 + a_1 costpp + a_2 lhtime + a_3 trmtime + a_4 xfern + a_5 business$$
$$+ a_6 female + c_1 w_1^* + c_2 w_2^* + \nu, \tag{31}$$

where

*costpp*: travel cost per person (Guilder);

*business*: 1 if the trip purpose is on business, 0 otherwise; and

*female*: 1 if female, 0 otherwise.

This model represents the binary RP choice of rail vs. car as described in the previous section.

First conducted is the sequential estimation method described in Section 2 utilizing available software for linear structural equation estimation. Parameter estimates of the linear structural equation model are shown in Table 2. All the

*Table 2.* Estimates of the Linear Structural Equation Model Sequential Estimation (*t*-Statistics in Parentheses).

$$
\hat{\mathbf{B}}' = \begin{bmatrix} (w_1^*) & (w_2^*) & \\ -0.232(-1.4) & 0.406(3.3) & (aged) \\ -0.292(-1.3) & 0 & (lhtime) \\ 0 & -0.522(-2.1) & (trmtime) \\ 0.286(1.0) & 0 & (first) \\ 0 & -0.0471(-0.6) & (xfern) \\ 0 & 0.164(1.6) & (freepark) \\ -0.0405(-0.1) & 0 & (aged \times lhtime) \end{bmatrix}
\qquad
\hat{\Lambda} = \begin{bmatrix} (w_1^*) & (w_2^*) & \\ 1 & 0.170(0.8) & (relax) \\ 0.772(1.8) & 1 & (relia) \\ 0 & 1.49(4.3) & (flex) \\ 0 & 1.16(5.2) & (ease) \\ 0.686(3.1) & 0.329(2.0) & (safe) \\ 1.64(2.6) & 2.43(5.9) & (overall) \end{bmatrix}
$$

parameters in the measurement equations have positive signs as expected and sufficiently large $t$-statistics. Although some estimates are not significant in the structural equations, most of them have expected signs.

Then the fitted values of the latent variables are used as explanatory variables in the rail/car binary choice model with the scale correction as shown in (25). The estimation results of the choice model with and without the latent variables are shown in the first and second columns of Table 4, respectively. Both the latent variables have significantly positive coefficients and raise the goodness-of-fit substantially. Since the variable of line-haul travel time is used in both the structural equation model and the choice model, its coefficient in the choice model becomes insignificant probably due to multicollinearity. The alternative specific constant (rail constant) becomes also less significant because the two latent variables capture much of the intangible factors specific to the travel mode. In other words, the choice model without the latent variables might have suffered from the omitted variable problem.

Simultaneous estimation is then conducted by using the likelihood given by (27). Parameter estimates of the linear structural equation part is shown in Table 3. Most of the parameters have the same tendency in terms of sign and magnitude as in the sequential estimation result. More parameters are observed statistically significant in the simultaneous estimation result. The choice model part shown in the third column of Table 4 demonstrates similar results to the sequential estimation model. Both the latent attributes have significantly positive parameter estimates.

When the policy analysis is conducted with the future values of the explanatory variables, only the structural equations can be used because we usually do not know the future values of the subjective ratings (perceptual indicators) of the latent attributes. In that sense, having more significant

***Table 3.*** Estimates of the Linear Structural Equation Part Simultaneous Estimation ($t$-Statistics in Parentheses).

$$\hat{\mathbf{B}}' = \begin{bmatrix} (w_1^*) & (w_2^*) & \\ -0.427(-2.4) & 0.378(2.4) & (aged) \\ -0.323(-1.7) & 0 & (lhtime) \\ 0 & -1.98(-9.0) & (trmtime) \\ 0.281(0.9) & 0 & (first) \\ 0 & -0.396(-3.7) & (xfern) \\ 0 & 0.482(3.5) & (freepark) \\ -0.339(-1.3) & 0 & (aged \times lhtime) \end{bmatrix} \quad \hat{\Lambda} = \begin{bmatrix} (w_1^*) & (w_2^*) & \\ 0.433(7.6) & 0.280(3.2) & (relax) \\ 0.527(12.5) & 0.661(10.2) & (relia) \\ 0 & 0.815(14.7) & (flex) \\ 0 & 0.794(14.2) & (ease) \\ 0.462(11.6) & 0.311(5.2) & (safe) \\ 0.784(8.5) & 1.76(14.1) & (overall) \end{bmatrix}$$

***Table 4.*** Choice Models with Latent Attributes ($t$-Statistics in Parentheses).

| | Model w/o Latent Attributes | Sequential Estimation Model | Simultaneous Estimation Model |
|---|---|---|---|
| Rail constant | 0.583 | 0.322 | −1.81 |
| | (2.0) | (1.0) | (−0.9) |
| Cost per person | −0.0268 | −0.0338 | −0.0379 |
| | (−4.2) | (−4.1) | (−4.3) |
| Line-haul time | −0.405 | 0.0751 | 0.379 |
| | (−1.6) | (0.2) | (0.9) |
| Terminal time | −1.57 | −1.18 | −0.818 |
| | (−4.2) | (−2.6) | (−2.3) |
| Number of transfers | −0.195 | −0.316 | −0.230 |
| | (−1.3) | (−1.7) | (−1.2) |
| Business trip dummy | 0.942 | 1.33 | 1.28 |
| | (3.6) | (3.6) | (3.3) |
| Female dummy | 0.466 | 0.652 | 0.700 |
| | (2.3) | (2.6) | (2.9) |
| $w_1^*$ *(comfort)* | | 0.882 | 1.29 |
| | | (2.7) | (1.8) |
| $w_2^*$ *(convenience)* | | 1.39 | 1.10 |
| | | (4.1) | (4.7) |
| $\bar{\rho}^2$ | 0.242 | 0.352 | 0.201* |

Note: * goodness-of-fit measure for both the structural equation and choice models.

parameters in the structural equations, the simultaneous estimation result is more useful in this particular case study.

## 5. ESTIMATION OF THE INTEGRATED MODEL

Integrating the two estimation schemes demonstrated in the two submodels, we could estimate a choice model with latent attributes using RP, SP and perceptual data. Table 5 shows an example of such models. Here, the two latent attributes, ride comfort and convenience, are included in the RP utility function, and RP data and the two types of SP data are simultaneously used to estimate coefficients of the utility functions. The linear structural equation model and the choice model are estimated in the sequential way.

The two latent attributes show the significant explanatory power to the RP data as also demonstrated in the previous section. Line-haul travel time in the utility function has a significant coefficient in the utility function. In the previous section this coefficient lost the explanatory power due to the multicollinearity between the latent attributes and line-haul travel time. By

***Table 5.*** Estimation Result of the Integrated Model (RP + SP1 + SP2 + Latent Variables).

| | Coefficient | $t$-statistic |
|---|---|---|
| Rail constant (RP) | 0.526 | 2.3 |
| Rail constant (SP) | −3.97 | −3.8 |
| Cost per person | −0.0352 | −5.7 |
| Line-haul time | −0.407 | −5.3 |
| Terminal time | −1.20 | −3.6 |
| Number of transfers | −0.0590 | −3.5 |
| Business trip dummy | 0.404 | 1.8 |
| Female dummy | 0.262 | 1.5 |
| Inertia dummy | 5.76 | 4.4 |
| $w_1^*$ *(comfort)* | 0.615 | 2.3 |
| $w_2^*$ *(convenience)* | 0.973 | 3.4 |
| $\theta_{11}$ | 0.0176 | 5.9 |
| $\theta_{12}$ | 0.271 | 25.3 |
| $\theta_{21}$ | 0.0827 | 8.4 |
| $\theta_{22}$ | 0.484 | 21.7 |
| $\mu_1$ | 2.35 | 5.6 |
| $\mu_2$ | 0.257 | 4.5 |
| $\bar{\rho}^2$ | 0.341 | |

combining RP and SP data, this key variable recovered significance in the utility function.

Coefficients of the other variables and the scale parameters have similar estimated values to the ones shown in Table 1. This empirical analysis can be seen as a demonstration of efficacy of the methodology presented in the paper.

## 6. CONCLUSIONS

This paper presents a methodology for incorporating psychometric data such as stated preferences and subjective ratings of attributes into the discrete choice modeling framework. The framework is composed of discrete choice models which describe discrete responses of revealed and stated preferences and a linear structural equation or covariance structure model which identifies latent attributes from psychometric perceptual indicators.

Empirical case studies on travel mode choice analysis have demonstrated the effectiveness of this methodology. Combined estimation of RP and SP models helped identify coefficients of important variables such as line-haul travel time and detected SP specific biases. Latent attributes identified by the linear structural equation model significantly improved the goodness-of-fit of the discrete choice model.

In the case study of the RP/SP combined estimation method, three combined models were estimated: RP data combined with SP1 (rail vs. rail) data, RP data combined with SP2 (rail vs. car) data, and RP data combined with both SP1 and SP2 data. These combined models were compared against the three models that were separately estimated from the three data sets. The RP model could not successfully identify an important parameter (the coefficient of line-haul travel time), which is a typical problem encountered in estimating models from RP data. This is usually caused by lack of variation in the data and/or misspecification of the model. However, obtaining an acceptable model specification is often very difficult because the actual behavior is influenced by related attributes while the available data are limited. Furthermore, even if the correct model specification was known, estimation of model parameters could fail because of data limitations. SP experiments present simplified hypothetical choice contexts and, therefore, may provide useful information on trade-offs among attributes.

The case study provided a clear demonstration of the usefulness of the combined estimation method. Specifically, the coefficient of the line-haul travel time variable was successfully estimated by combining RP and SP data. The SP1 experiment for rail vs. rail choice provided information on the trade-offs among attributes with the least random noise. On the other hand, SP data are often not reliable because of the oversimplified hypothetical circumstances. This problem was mitigated by using additional variables from the RP data in estimating the SP model.

A potential bias in the SP data was captured by the introduction of the *inertia* variable. This variable captured the preference bias toward the mode actually chosen. As discussed above, it was found that car users had a greater inertia or habitual effect in choosing a travel mode.

Thus, these case studies successfully demonstrated the key features of the RP/SP combined modeling method (Ben-Akiva & Morikawa, 1990a, b):

(i) efficiency: joint estimation of preference parameters from all the available data;
(ii) bias correction: explicit response models for SP data that include both preference and bias parameters; and
(iii) identification: estimation of trade-offs among attributes and the effects of new services that are not identifiable from RP data.

This methodology of combining different preference data sources has recently been widely applied in various contexts not only in demand forecasting but also in environment valuation (e.g. Hensher & Bradley, 1993; Swait & Louviere, 1993; Adamowicz et al., 1994; Ortuzar & Iacobelli, 1998; Hensher et al., 1999).

The paper also proposed a method for incorporating attitudinal data such as subjective ratings of latent attributes. The framework is composed of discrete choice models which describe discrete responses of revealed choices and a linear structural equation model which identifies latent attributes from psychometric perceptual indicators. It is totally different in concept from the traditional methods in which psychometric indicators are directly used as explanatory variables. The key feature of the proposed method is that we can calculate the latent attributes from the observable variables once parameters are estimated. This implies that the models described in this paper can be used for forecasting demand in conjunction with changes in product attributes, level-of-service, and consumer's characteristics.

The empirical case study demonstrated the effectiveness of this methodology by showing that inclusion of the latent attributes significantly improved goodness-of-fit measure of the discrete choice model. Two estimators were presented: sequential and simultaneous estimation. The sequential method can utilize existing linear structural equation estimation software such as LISREL, but provides not fully efficient estimators. The simultaneous full information maximum likelihood method yields efficient estimators although it requires programming the likelihood. The empirical analysis showed that the two methods yielded similar estimation results both in the choice model part and in the linear structural equation model part. But the more effective result was obtained by the simultaneous estimation in the sense that more significant parameter estimates were found in the structural equations that are used for forecasting.

An estimation result of the model that integrates the two submodels is exhibited in Chapter 5. This particular empirical analysis shows strong explanatory power of the latent attributes that are identified by a structural equation model and significance of key variables such as travel time and cost in the utility function, which demonstrates effectiveness of combining RP and psychometric data in a general and consistent framework.

In the general framework of consumer behavior analysis depicted in Fig. 1, focused on in this paper are incorporating stated preferences and perceptual indicators to better identify latent preferences and perceptions. The methodology proposed in this paper seems to be well supported by the case studies. More empirical works, however, are called for in order to justify it in a more conclusive way. Some of the other aspects of the general framework have also been worked by the authors. Discrete choice models with explicit consideration of situational constraints and choice set formation are proposed by Ben-Akiva and Boccara (1995) and Morikawa (1996). Preliminary work on incorporating attitudinal indicators has been done by Sasaki et al. (1999).

## REFERENCES

Adamowicz, W., Louviere, J., & Williams, M. (1994). Combining stated and revealed preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management, 26*, 271–292.

Ben-Akiva, M., & Morikawa, T. (1990a). Estimation of switching models from revealed preferences and stated intentions. *Transportation Research, 24A*, 485–495.

Ben-Akiva, M., & Morikawa, T. (1990b). Estimation of travel demand models from multiple data sources. In: M. Koshi (Ed.), *Transportation and Traffic* (pp. 461–476). Elsevier.

Ben-Akiva, M., Morikawa, T., & Shiroishi, F. (1991). Analysis of the reliability of preference ranking data. *Journal of Business Research, 23*, 253–268.

Ben-Akiva, M., & Boccara, B. (1987). *Integrated framework for travel behavior analysis.* Presented at IATBR Conference, Aix-en-Provence, France.

Ben-Akiva, M., & Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing, 12*, 9–24.

Ben-Akiva, M., McFadden, D., Garling, T., Gopinath, D., Bolduc, D., Borsch-Supan, A., Delquie, P., Larichev, O., Morikawa, T., Polydoropoulou, A., & Rao, V. (1999). Extended framework for modeling choice behavior. *Marketing Letters, 10*(3), 187–203.

Bentler, P. (1980). Multivariate analysis with latent variables: causal models. *Annual Review of Psychology, 31*, 419–456.

Bielby, W., & Hauser, R. (1977). Structural equation models. *Annual Review of Sociology, 3*, 137–161.

Carpenter, G., & Lehmann, D. (1985). A model of marketing mix, brand switching, and competition. *Journal of Marketing Research, 22*, 318–329.

Cattin, P., & Wittink, D. R. (1982). Commercial use of conjoint analysis: a survey. *Journal of Marketing, 46*, 44–53.

Chintagunta, P. (1993). Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Science, 12*, 184–208.

Duncun, O. (1975). *Introduction to Structural Equation Models*. Academic Press, New York.

Elrod, T., Louviere, J., & Krishnakumar, D. (1992). An empirical comparison of rating-based and choice-based conjoint models. *Journal of Marketing Research, 29*, 368–377.

Gensh, D., & Recker, W. (1979). The multinomial, multiattribute logit choice model. *Journal of Marketing Research, 16*, 124–132.

Goldberger, A. (1972). Structural equation methods in the social sciences. *Econometrica, 40*, 979–1001.

Green, P., & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research, 8*, 355–363.

Green, P., Carmore, F., & Wachpress, D. (1977). On the analysis of qualitative data in marketing research. *Journal of Marketing Research, 14*, 52–59.

Green, P., & Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research, 5*, 103–123.

Guadagni, P., & Little, J. (1983). A logit model of brand choice. *Marketing Science, 2*, 203–238.

Gupta, S. (1988). Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing Research, 25*, 342–355.

Hauser, J., & Koppelman, F. (1979). Alternative perceptual mapping techniques: relative accuracy and usefulness. *Journal of Marketing Research, 16*, 495–506.

Hensher, D., & Bradley, M. (1993). Using stated response data to enrich revealed preference discrete choice models. *Marketing Letters, 4*, 139–152.

Hensher, D., Louviere, J., & Swait, J. (1999). Combining sources of preference data. *Journal of Econometrics, 89*, 197–221.

Johnson, R., & Wichern, D. (1988). *Applied Multivariate Statistical Analysis* (2nd ed.). Prentice-Hall, New Jersey.

Jöreskog, K., & Sörbom, D. (1979). *Advances in Factor Anallysis and Structural Equation Models*. Cambridge, Mass.: Abt Books.

Jöreskog, K., & Sörbom, D. (1984). LISREL VI: Analysis of linear structural relations by maximum likelihood, instrumental variables, and least squares methods, User's guide, Department of Statistics, University of Uppsala, Uppsala, Sweden.

Koppelman, F., & Pas, E. (1980). Travel-choice behavior: models of perceptions, feelings, preference, and choice. *Transportation Research Record, 765*, 26–33.

Lichtenstein, D., Ridgway, N., & Netemeyer, R. (1993). Price perceptions and consumer shopping behavior: a field survey. *Journal of Marketing Research, 30*, 234–245.

Louviere, J. (1988a). Analyzing decision making – metric conjoint analysis. Sage University paper series on quantitative applications in the social science, Sage.

Louviere, J. (1988b). Conjoint analysis modelling of stated preference. *Journal of Transport Economics and Policy, 22*, 93–118.

Louviere, J., Hensher, D., & Swait, J. (2001). *Stated choice methods: Analysis and applications in marketing, Transportation and Environmental Valuation*. Cambridge: Cambridge University Press.

McFadden, D. (1986). The choice theory approach to market research. *Marketing Science, 5*, 275–297.

Meyer, R., & Johnson, E. (1995). Empirical generalizations in the modeling of consumer choice. *Marketing Science, 14*, G180–189.

Morikawa, T. (1989). Incorporating stated preference data in travel demand analysis. Ph.D. dissertation, Department of Civil Engineering, MIT.

Morikawa, T., Ben-Akiva, M., & McFadden, D. (1990). Incorporating psychometric data in economic travel demand models. Banff Invitational Symposium on Consumer Behavior. Banff, Canada.

Morikawa, T. (1994). Correcting state dependence and serial correlation in the RP/SP combined estimation method. *Transportation, 21*, 153–165.

Morikawa, T. (1996). A hybrid probabilistic choice set model with compensatory and non-compensatory choice rules. Proceedings of the 7th WCTR, Sydney 1, 317–325.

Ortuzar, J., & Iacobelli, A. (1998). Mixed modeling of interurban trips by coach and train. *Transportation Research, A32*, 345–357.

Recker, W., & Golob, T. (1976). Attitudinal modal choice model. *Transportation, 4*, 293–310.

Sasaki, K., Morikawa, T., & Kawakami, S. (1999). A discrete choice model with taste heterogeneity using SP, RP and attribute importance ratings. Selected Proc. of 8th World Conf. on Transport Research 3, Elsevier, 39–49.

Swait, J., & Louviere, J. (1993). The role of the scle parameter in the estimation and use of multinomial logit models. *Journal of Marketing Research, 30*, 305–314.

Winter, R. (1986). A reference price model of brand choice for frequently purchased products. *Journal of Consumer Research, 13*, 250–256.

# ANALYSIS OF MULTI-CATEGORY PURCHASE INCIDENCE DECISIONS USING IRI MARKET BASKET DATA

Siddhartha Chib, P. B. Seetharaman and Andrei Strijnev

## ABSTRACT

*Empirical studies in Marketing have typically characterized a household's* purchase incidence *decision, i.e. the household's decision of whether or not to buy a product on a given shopping visit, as being independent of the household's purchase incidence decisions in other product categories. These decisions, however, tend to be related both because product categories serve as complements (e.g. bacon and eggs) or substitutes (e.g. colas and orange juices) in addressing the household's consumption needs, and because product categories vie with each other in attracting the household's limited shopping budget. Existing empirical studies have either ignored such inter-relationships altogether or have accounted for them in a limited way by modeling household purchases in* pairs *of complementary product categories. Given the recent availability of IRI market basket data, which tracks purchases of panelists in several product categories over time, and the new computational Bayesian methods developed in Albert and Chib (1993) and Chib and Greenberg (1998), estimating high-dimensional multi-category models is now possible. This paper exploits these developments to fit an appropriate panel data multivariate probit model to household-level contemporaneous purchases in twelve product categories, with the descriptive goal of isolating*

**Econometric Models in Marketing, Volume 16, pages 57–92.**

**ISBN: 0-7623-0857-5**

*correlations amongst various product categories within the household's shopping basket. We provide an empirical scheme to endogenously determine the degree of complementarity and substitutability among product categories within a household's shopping basket, providing full details of the methodology. Our main findings are that existing purchase incidence models underestimate the magnitude of cross-category correlations and overestimate the effectiveness of the marketing mix, and that ignoring unobserved heterogeneity across households overestimates cross-category correlations and underestimate the effectiveness of the marketing mix.*

## 1. MOTIVATION

Over the past decade, marketing researchers have devoted a lot of attention to the problem of modeling household purchase incidence at the category level (see, for example, Chiang, 1991; Bucklin & Lattin, 1991; Chintagunta, 1993). One reason for modeling category purchase incidence, in addition to brand-choices within the product category, is that such a model provides improved estimates of brand-choice elasticities with respect to marketing mix variables, properly accounting for not just the direct impact but also the indirect impact on brand-choice via category purchase incidence (Chiang, 1991). A second reason stems from the researcher's desire to understand what factors drive category purchase incidence and what impact, if any, marketing-mix variables at the brand level have on category purchase incidence. A third reason is the purely descriptive goal of isolating correlations amongst various product categories within the household's shopping basket, thereby providing a scheme to determine which categories are complements and which are substitutes.

Previous studies have largely focused on the first issue, i.e. obtaining improved estimates of brand-choice elasticities. The second issue, i.e. estimating the impact of brands' marketing variables on category purchase incidence, and the third issue, i.e. estimating cross-category correlations, have been incompletely addressed at best. While the former is in part due to the difficulty of formulating appropriate models of category purchase incidence, the latter is largely due to the computational problems of fitting realistic household-level category purchase incidence models on scanner panel data. For example, if a household buys thirty different product categories during a visit to the store, a model that estimates cross-category correlations must simultaneously model household decisions in thirty different product categories, an onerous task by any standards. The purpose of this paper, which is part of a two-stage research agenda, is to explicitly address the third issue, i.e. estimate cross-category correlations within the household's shopping basket. We study what information is contained in category purchase incidence data when, not just two or three, but a large number of category purchase incidence decisions (twelve in our case) are modeled simultaneously. The success of our fitting enterprise, based on the work of Albert and Chib (1993) and Chib and Greenberg (1998), and summarized in this paper, makes us hopeful that we will be able to scale-up our model to include all the twenty or so categories in the typical shopping basket. The second-stage of our research, described in a companion paper, addresses all three issues simultaneously, i.e. jointly modeling category purchase incidence and brand choice when the number of categories is large.

## 2. OBJECTIVES OF THIS STUDY

Households make purchase decisions in several product categories when they visit the supermarket. For example, a household's regularly scheduled trip to the grocery store may involve the purchase of soft drinks, chips, ketchup, cookies, peanut butter, ice cream, laundry detergents, etc. To the extent that product categories serve different consumption needs of the household, household purchase decisions may appear to be independent across product categories within the household's shopping basket. For example, a household's decision to purchase laundry detergents may be independent of the household's decision to purchase bacon or soft drinks since each product serves a fundamentally different consumption need. On the basis of this independence assumption, empirical researchers typically estimate household purchase incidence decisions separately for each product category, i.e. whether or not a household will buy ketchup during a visit to the store is modeled independently of whether or not it will purchase other products in the store (see, for example, Bucklin & Lattin, 1991; Chiang, 1991). This is also referred to as the *weak separability* assumption.

It is unlikely that the weak separability assumption applies to *all* product categories within a household's shopping basket. For example, some products may serve as consumption complements of each other (say, bacon and eggs) while others may serve as consumption substitutes of each other (say, cola and orange juice).[1] Researchers have accounted for this by identifying pairs of products, a priori, that are obvious complements of each other and estimating bivariate models of household purchase incidence decisions across the two product categories (Chintagunta & Haldar, 1998; Manchanda, Ansari & Gupta, 1999). Such a framework is applicable only when one can identify a priori relationships among product categories. In general, however, one must *endogenously* infer the relationships between product categories within the household's shopping basket using purchase data. For example, one must estimate a high-dimensional model of household purchase incidence decisions across all product categories within the household's shopping basket (also referred to as a *basket-level model* henceforth). Such a basket-level model will endogenously estimate correlations across all pairs of product categories rather than across predefined product categories only. Even if the focus is on estimating correlations among and/or marketing mix elasticities within predefined pairs of product categories (as in Chintagunta & Haldar, 1998; Manchanda, Ansari & Gupta, 1999), it is important to estimate these correlations and elasticities using a basket-level model to eliminate the effects of misspecification bias. This is the first objective of this study, and we summarize it below:

> *Objective 1:* We estimate a basket-level model of household purchase incidence decisions to obtain estimates of pair-wise correlations across all product categories within the household's shopping basket and estimates of marketing mix elasticities in each product category.

Cross-category correlations are of interest to retailers seeking to maximize store profits by jointly coordinating marketing activities across product categories within the store. Cross-category correlations are also of interest to database marketers interested in undertaking cross-selling initiatives across product categories (Berry & Linoff, 1997). A complete basket-level model of household purchase incidence decisions, as proposed in this study, has not been estimated thus far in the marketing literature. We estimate our basket-level model using scanner panel data, which tracks the purchases of a fixed number of households across twelve different product categories in the store over time.

While using scanner panel data, it is important to investigate how sensitive the estimated cross-category correlations are to the panel structure of the data. In other words, one must assess the impact of (ignoring or accommodating) unobserved heterogeneity across households on the estimated cross-category correlations. To the extent that cross-category correlations may proxy for the effects of unobserved heterogeneity if the latter is ignored, it is possible that cross-category correlations may be overstated (and hence "spurious") in the absence of unobserved heterogeneity. Also, the estimated marketing mix elasticities in each product category may be sensitive to the inclusion of unobserved heterogeneity across households. Explicitly investigating this issue is the second objective of this study, and we summarize it below:

*Objective 2:* We estimate the basket-level model of household purchase incidence decisions both with and without accommodating the effects of unobserved heterogeneity across households in order to investigate the consequences of ignoring unobserved heterogeneity on the estimated cross-category correlations and households' responsiveness to marketing variables in each product category.

Disentangling cross-category correlations from unobserved heterogeneity is important to retailers since the two phenomena imply different marketing strategies. For example, if cross-category correlations are observed to be simply proxies for unaccounted-for heterogeneity across households, the marketer could develop marketing programs separately for each product category taking into account the estimated heterogeneity distribution. In such a case, separately maximizing the profits from each category is tantamount to maximizing overall store profits.

To summarize, we propose a basket-level model of household purchase incidence decisions and estimate the proposed model using scanner panel data on household purchases across twelve product categories. The proposed model has a multivariate probit panel structure and is used to estimate pair-wise correlations in households' random utilities across the twelve product categories. We employ an extension of a recently developed Bayesian method (Albert & Chib, 1993; Chib & Greenberg, 1998) to estimate model parameters. Our main findings are that either ignoring or incompletely accounting for cross-category correlations within household shopping baskets overestimates the effectiveness of marketing variables in driving purchase incidence decisions. We also find that ignoring unobserved heterogeneity across households overstates cross-category correlations and understates the effectiveness of marketing variables. The rest of the paper is organized as follows. In the next section we propose the multivariate probit panel model and discuss estimation issues. In Section 4, we provide details of the Markov Chain Monte Carlo sampling scheme. In Section 5, we give a detailed description of the data. In Section 6, we present our empirical results. We conclude with a summary and directions for future research in Section 7.

## 3. MODEL AND ESTIMATION

### *Notation*

Suppose we observe binary responses of $H$ households in $J$ product categories over time. We refer to this collection of responses as $\{y_{htj} \in (0, 1)$: $h = 1, \ldots, H$; $t = 1, \ldots, T_h$; $j = 1, \ldots, J\}$ where subscripts $h$, $t$ and $j$ refer to household, shopping occasion and product category respectively. We define $y_{ht} = (y_{ht1}, y_{ht2}, \ldots, y_{htJ})'$, $y_h = (y'_{h1}, y'_{h2}, \ldots, y'_{hTh})'$ and $y = (y'_1, y'_2, \ldots, y'_H)'$. Note that $y_{htj}$ is a scalar, $y_{ht}$ is a $J$-dimensional vector, $y_h$ is a $J*T_h$-dimensional vector and $y$ is a $\Sigma_h J*T_h$-dimensional vector.

We also observe values of $k$ marketing variables for each product category at each shopping occasion for each household. We refer to this collection of $k$-dimensional covariate vectors as $\{X_{htj}: h = 1, \ldots, H; t = 1, \ldots, T_h; j = 1, \ldots, J\}$. We define $X_{ht}$ as

$$X_{ht} = \begin{pmatrix} X'_{ht1} & 0 & \ldots & 0 \\ 0 & X'_{ht2} & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & X'_{htJ} \end{pmatrix}, \tag{1}$$

and define $X_h = (X'_{h1}, X'_{h2}, \ldots, X'_{hTh})'$ and $X = (X'_1, X'_2, \ldots, X'_H)'$. Note that $X_{ht}$ is a $(J)*(k*J)$-dimensional matrix, $X_h$ is a $(J*Th)*(k*J)$-dimensional matrix and $X$ is a $(\Sigma_h J*T_h)*(k*J)$-dimensional matrix.

We assume that $y_{htj}$ not only depends on $X_{htj}$ but also is correlated with $y_{htk}$ (for $k \neq j$). In other words, a household's response in a product category depends both on category-specific marketing variables and on the household's responses in other product categories. This is a multivariate choice problem for the household. Previous work has either completely ignored dependencies across $y'_{htj}$s, thereby assuming univariate choice problems for the household for each product category (Chiang, 1991; Chintagunta, 1993), or accounted for dependencies across a limited number of obviously related product categories (Chintagunta & Haldar, 1998; Manchanda, Ansari & Gupta, 1999). In our framework we pose the multivariate choice problem in the context of the household's shopping basket, and therefore in its fullest generality. Next we present the model that explains the observed response vector $y$.

### *Multivariate Probit Model with Unobserved Heterogeneity*

Let household $h$'s latent utility at shopping occasion $t$ for product category $j$ be given by

$$Z_{htj} = X'_{htj}\beta_j + b_h + c_{hj} + \varepsilon_{htj}, \tag{2}$$

where $X_{htj}$ is a $k$-dimensional vector of marketing variables pertaining to product category $j$ facing household $h$ at shopping occasion $t$, $\beta_j$ is the corresponding $k$-dimensional parameter vector $(\beta_{j1}, \beta_{j2}, \ldots, \beta_{jk})$, $b_h$ represents a household-specific random effect that is distributed $N(0, d)$, $c_{hj}$ represents a household/category-specific random effect such that $c_h = (c_{h1}, c_{h2}, \ldots, c_{hJ})'$ is distributed $N_J(0, C)$, and $\varepsilon_{htj}$ is a random component such that $\varepsilon_{ht} = (\varepsilon_{ht1}, \ldots, \varepsilon_{htJ})'$ is distributed $N_J(0, \Sigma)$, where $\Sigma$ is a $J*J$ covariance matrix given by

$$\Sigma = \begin{pmatrix} 1 & \sigma_{12} & \cdots & \sigma_{1J} \\ & 1 & \cdots & \sigma_{2J}. \\ & & \cdots & \sigma_{J-1,J} \\ & & & 1 \end{pmatrix}. \tag{3}$$

This covariance matrix is in correlation form for identifiability reasons and contains $p = J*(J-1)/2$ free parameters (see Chib & Greenberg, 1998 for details) given by $\sigma \equiv (\sigma_{12}, \sigma_{13}, \ldots, \sigma_{J-1,J})$.

It is also helpful to rewrite the model in (2) for all $J$ categories as

$$Z_{ht} = X_{ht}\beta + i_J b_h + I_J c_h + \varepsilon_{ht}, \tag{4}$$

where $Z_{ht} = (Z_{ht1}, \ldots, Z_{htJ})'$, $X_{ht}$ is the $(J)*(k*J)$-dimensional matrix of marketing variables facing the household at shopping occasion $t$ (as given by Eq. (1)), $\beta$ is the corresponding $k*J$-dimensional parameter vector $(\beta_1, \beta_2, \ldots, \beta_J)$ where $\beta_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jk})$, $i_j$ is a $J$-dimensional vector of ones, $I_J$ is a $J*J$ identity matrix, $b_h$ is a household-specific (scalar) random effect that is distributed $N(0, d)$, and $c_h$ is a $J$-dimensional household-specific random effect vector that is distributed $N_J(0, C)$. Observed responses $y_{htj}$ are determined by the unobserved latent variables $Z_{htj}$ as:

$$y_{htj} = I[Z_{htj} > 0], \tag{5}$$

where $I$ is the indicator function. This completes the specification of our model. The total number of parameters in the proposed model is equal to $J*[k + (J-1)/2 + 1]$ (i.e. $k*J$ covariate coefficients, plus $J*(J-1)/2$ correlation coefficients, plus $J*1$ random effects parameters).[2] If the number of product categories $J$ is small (say, 2–4) we obtain the cross-category model of Manchanda, Ansari and Gupta (1999). If the random effects are restricted to be the same across product categories, i.e. $c_h$ is ignored, we obtain a restricted version of our proposed model that assumes the unobserved heterogeneity distribution to be common across product categories. If the effects of unobserved heterogeneity are ignored altogether (i.e. $b_h$ and $c_h$ are ignored), we obtain a cross-sectional version (as opposed to a panel version) of our proposed multivariate probit model (as in Chib & Greenberg, 1998). If correlations across product categories are ignored in the common random effects model, i.e. $\Sigma = I$ (a diagonal matrix of ones), we obtain $J$ independent category models with a common unobserved heterogeneity distribution. If the unobserved heterogeneity distributions are assumed to be independent across product categories, we obtain single-category heterogeneous models as in Chiang (1991), Bucklin and Lattin (1991), Chintagunta (1993) etc.

We can estimate marketing mix elasticities for each product category based on our proposed model and compare these elasticities to those obtained using a model that ignores cross-category correlations $\Sigma$. This allows us to understand the effects of ignoring cross-category correlations on measures of managerial relevance such as price elasticities (our research objective no. 1). We can also compare the correlation matrix $\Sigma$ estimated using our proposed model with that estimated using a restricted version of the model that ignores household-specific random effects (i.e. $b_h$ and $c_h$). This allows us to understand the effects of ignoring unobserved heterogeneity across households on the estimated cross-category correlations (our research objective no. 2).

Given $J$ product categories and $T_h$ observations for a given household $h$, likelihood-based estimation of our proposed model requires the computation of the likelihood contribution

$$\Pr(y_h \mid \beta, \sigma, d, C) = \int \left[ \prod_{t=1}^{T_h} \int_{B_{hJt}} \int_{B_{hJ-1,t}} \cdots \int_{B_{h1t}} \phi_J(Z_{ht} \mid X_{ht}\beta + i_J b_h + I_J c_h, \Sigma) dZ_{ht} \Big) \right] \cdot \phi(b_h \mid 0, d) \phi(c_h \mid 0, C) db_h dc_h, \tag{6}$$

for each household $h = 1, \ldots, H$, where $\phi_J(. \mid \mu, \Sigma)$ is the density of a $J$-variate normal distribution with mean $\mu$ and covariance matrix $\Sigma$, $B_{hjt}$ is the interval $(0, \infty)$ if $y_{hjt} = 1$ and the interval $(-\infty, 0)$ if $y_{hjt} = 0$. This likelihood contribution is quite difficult to compute even using simulation techniques. Given the computational intractability of likelihood-based estimation, we adopt a simulation-based Bayesian approach to estimate model parameters.

### *Bayesian Approach to Model Estimation*

Given the response vector $y$, the matrix of covariates $X$, and a prior density on model parameters given by $\pi(\beta, \sigma, d, C)$, Bayes rule yields

$$\pi(\beta, \sigma, d, C \mid y) \propto \pi(\beta, \sigma, d, C) * \Pr(y \mid \beta, \sigma, d, C), \tag{7}$$

where

$$\Pr(y \mid \beta, \sigma, d, C) = \left( \prod_{h=1}^{H} \Pr(y_h \mid \beta, \sigma, d, C) \right) * U[\sigma \in Q], \tag{8}$$

and $\Pr(y_h \mid \beta, \sigma, d, C)$ is given by Eq. (6) and $Q$ is a convex solid body in the hypercube $[-1, 1]^p$ that leads to a proper correlation matrix. This form of the posterior density is not particularly useful for Bayesian estimation since it involves the evaluation of the complicated likelihood function (just as in likelihood-based estimation). Instead of attempting to directly evaluate the joint posterior density we invoke the data augmentation framework of Albert and Chib (1993) and Chib and Greenberg (1998). This framework is based on taking a sampling-based approach, in conjunction with Markov Chain Monte Carlo (MCMC) techniques (Tanner & Wong, 1987; Gelfand & Smith, 1990; Tierney, 1994; Chib & Greenberg, 1995), based on the conditional distributions given by

$$Z_{ht} \mid y_h, \beta, \sigma, d, C; t = 1, \dots, T_h; h = 1, \dots, H$$

$$\beta \mid y_h, Z_h, \sigma, d, C$$

$$b_h \mid y_h, Z_h, \beta, \sigma, C; h = 1, \dots, H$$

$$c_h \mid y_h, Z_h, \beta, \sigma, d; h = 1, \dots, H$$

$$\sigma \mid y_h, Z_h, \beta, d, C$$

$$d^{-1} \mid y_h, Z_h, \beta, \sigma, C$$

$$C^{-1} \mid y_h, Z_h, \beta, \sigma, d$$

Each of these distributions (except that of $\sigma$) is of known form and can be sampled directly. Details are provided in the next section. The key simplification that data augmentation provides in our context is that it allows us to bypass the computation of the likelihood.

# 4. MARKOV CHAIN MONTE CARLO (MCMC) SAMPLING

### *Prior Distributions*

For the purposes of our analysis, we assume that our prior information can be represented by the distributions

$$\beta \sim N_{k*J}(\beta_o, B_o),$$

$$\sigma \sim N_{J*(J-1)/2}(g_o, G_o) * I[\sigma \in Q],$$

$$d^{-1} \sim G(\eta_o, \psi_o),$$

$$C^{-1} \sim \text{Wish}_J(\rho_o, R_o),$$

where the hyperparameters are as follows: $\beta_o$ is a $k*J$-dimensional vector of zeros, $B_o$ is a $(k*J)*(k*J)$ diagonal-matrix, with its diagonal elements equal to 0.1 implying a variance of 10 for each component of $\beta$, $g_o$ is a $p$-dimensional vector with all its elements equal to 0.5, $G_o$ is a $p*p$ identity matrix, $\eta_o = 1$, $\chi_o = 3$, $\rho_o = J + 4$, $R_o = 3*I_{J*J}$. The choice of these priors is intended to represent vague prior information.

*MCMC Algorithm*

We are interested in simulating from the posterior distribution of $(\{Z_h\}, \beta, \sigma, \{b_h\}, \{c_h\}, d^{-1}, C^{-1})$, where $Z_h$ is a $J*T_h$-dimensional vector given by $(Z'_{h1} Z'_{h2} \ldots Z'_{hTh})'$. While it is difficult to sample from the joint posterior, it is possible to simulate from the conditional distributions $f(Z_h \mid \beta, \sigma, d^{-1}, C^{-1})$, $\pi(\beta \mid \{Z_h\}, \sigma, d^{-1}, C^{-1})$, $\pi(b_h, c_h \mid \{Z_h\}, \beta, \sigma, C^{-1})$, $\pi(\sigma \mid \{Z_h\}, \beta, \{b_h\}, \{c_h\}, d^{-1}, C^{-1})$ and $\pi(D^{-1} \mid \{Z_h\}, \beta, \sigma, \{b_h\}, \{c_h\})$. The MCMC sampling algorithm works as follows.

*Step 0:* Initialize $\beta$ to $\beta^{(o)}$, $\sigma$ to $\sigma^{(o)}$, set $g = 1$
*Step 1:* Draw $Z_h^{(g)}$ from $f(Z_h \mid y, \beta^{(g-1)}, \sigma^{(g-1)}, d^{-1(g-1)}, C^{-1(g-1)})$, $h = 1, \ldots, H$.
*Step 2:* Draw $\beta^{(g)}$ from $\pi(\beta \mid y, \{Z_h^{(g)}\}, \sigma^{(g-1)}, d^{-1(g-1)}, C^{-1(g-1)})$.
*Step 3:* Draw $b_h^{(g)}$ from $\pi(b_h \mid y, \{Z_h^{(g)}\}, \beta^{(g)}, c_h^{(g-1)}, \sigma^{(g-1)}, d^{-1(g-1)}, C^{-1(g-1)})$, $h = 1, \ldots, H$.
*Step 4:* Draw $c_h^{(g)}$ from $\pi(c_h \mid y, \{Z_h^{(g)}\}, \beta^{(g)}, b_h^{(g)}, \sigma^{(g-1)}, d^{-1(g-1)}, C^{-1(g-1)})$, $h = 1, \ldots, H$.
*Step 5:* Draw $\sigma^{(g)}$ from $\pi(\sigma \mid y, \{Z_h^{(g)}\}, \beta^{(g)}, \{b_h^{(g)}\}, \{c_h^{(g)}\}, d^{-1(g-1)}, C^{-1(g-1)})$.
*Step 6:* Draw $d^{-1(g)}$ from $\pi(d^{-1} \mid y, \{Z_h^{(g)}\}, \beta^{(g)}, \sigma^{(g)}, \{b_h^{(g)}\}, \{c_h^{(g)}\}, C^{-1(g-1)})$.
*Step 7:* Draw $C^{-1(g)}$ from $\pi(C^{-1} \mid y, \{Z_h^{(g)}\}, \beta^{(g)}, \sigma^{(g)}, \{b_h^{(g)}\}, \{c_h^{(g)}\}, d^{-1(g-1)})$.
*Step 8:* $g = g + 1$.
*Step 9:* Go to step 1.

The above cycle of seven steps is repeated a large number of times (in our example, the entire simulation is run for 10,000 cycles). From the theory of MCMC simulations, it follows that the draws on $\theta = (\{Z_h\}, \beta, \{b_h\}, \{c_h\}, \sigma, d^{-1}, C^{-1})$, beyond a burn-in period of say 500 iterations, may be taken as draws from the posterior distribution of $\theta$. Therefore, on the basis of the simulated sample, we are able to obtain point and interval estimates of the parameters and other summaries of the posterior distribution. Next, we provide the form of each of the five conditional distributions given in steps 1–7.

1. $Z_h \mid y, \beta, \sigma, d^{-1}, C^{-1} \propto N_{J*Th}(Z_h \mid X_h\beta, i_{JTh}d + (i_{JTh} \otimes I_J)C(i_{JTh} \otimes I_J)' + I_{Th} \otimes \Sigma) * \Pi_t \Pi_j \{I(Z_{htj} > 0) * I(y_{htj} = 1) + I(Z_{htj} \leq 0) * I(y = 0)\}$, where $i_{JTh}$ is a $J*T_h$-

dimensional vector of ones, $I_J$ is a $J*J$ identity matrix, and $I_{JTh}$ is a $JT_h*JT_h$ identity matrix. This is a truncated multivariate normal distribution. This distribution is sampled through a Gibbs cycle (see Geweke, 1991). This representation of the conditional posterior of $Z_h$ follows from Albert and Chib (1993).

2. $\beta | y, \{Z_{ht}\}, \sigma, d^{-1}, C^{-1} \sim N_{k*J}(\beta | \hat{\beta}, B)$, where $B = (B_o + X'_h (I_{Th} \otimes \Sigma)^{-1} X_h)^{-1}$, $\hat{\beta} = B(B_o^{-1}\beta_o + \Sigma_h X'_h (I_{Th} \otimes \Sigma)^{-1} Z_h)$.
3. $b_h | y, \{Z_{ht}\}, \beta, \sigma, d^{-1}, C^{-1}, c_h \sim N_J(b_h | \hat{b}_h, B_h)$, where $B_h = ((dI_J)^{-1} + \Sigma_t I'_J I_J)^{-1}$, $\hat{b}_h = B_h (\Sigma_t I'_J (Z_{ht} - X_{ht}\beta - I_J c_h))$, $h = 1, \ldots, H$.
4. $c_h | y, \{Z_{ht}\}, \beta, \sigma, d^{-1}, C^{-1}, b_h \sim N_J(c_h | \hat{b}_h, \hat{B}_c)$, where $\hat{B}_c = (C^{-1} + \Sigma_t I'_J I_J)^{-1}$, $\hat{b}_c = \hat{B}_c (\Sigma_t I'_J (Z_{ht} - X_{ht}\beta - i_J b_h))$, $h = 1, \ldots, H$.
5. $\sigma | y, \{Z_{ht}\}, \beta, \{b_h\}, \{c_h\}, d^{-1}, C^{-1} \sim N_{J*(J-1)/2}(g_o, G_o) * I[\sigma \in Q] * \Pi_h \Pi_t N_J(Z_{ht} | X_{ht}\beta + i_J b_h + I_J c_h, \Sigma)$. We use the Metropolis-Hastings algorithm to sample from this non-standard distribution (details given in the next subsection), following Chib and Greenberg (1998).
6. $d^{-1} | y, \beta, \{b_h\}, \{c_h\}, \sigma \sim IG(\eta_o + H, \chi)$, where $\chi = (\chi_o^{-1} + \Sigma_h b_h b'_h)^{-1}$.
7. $C^{-1} | y, \beta, \{b_h\}, \{c_h\}, \sigma \sim W_J(\rho_o + H, R)$, where $R = (R_o^{-1} + \Sigma_h c_h c'_h)^{-1}$.

### *Metropolis-Hastings (M-H) Algorithm*

The only distribution in the set above that cannot be sampled directly is the distribution of $\sigma$, i.e. $\pi(\sigma | y, \{Z_{ht}\}, \beta, \{b_h\}, \{c_h\}, d^{-1}, C^{-1})$. To sample this distribution we use the M-H algorithm (see Chib & Greenberg, 1995 for a detailed exposition). Suppose $q(\sigma | \sigma', y, \{Z_{ht}\}, \beta, \{b_h\}, \{c_h\}, D^{-1})$ is a candidate generating density. Then to draw $\sigma$ we proceed as follows.

*Step 1:* Sample a proposal value $\sigma'$ given $\sigma$ from $q(\sigma' | y, \{Z_{ht}\}, \beta, \{b_h\}, \{c_h\}, D^{-1})$.

*Step 2:* Move to $\sigma'$ with probability $\alpha(\sigma, \sigma')$ and stay at $\sigma$ with probability $1 - \alpha(\sigma, \sigma')$, where

$$\alpha(\sigma, \sigma') = \min\left\{ \frac{\pi(\sigma' | y, Z_{ht}, \beta, \{b_h\}, \{c_h\}, D^{-1})}{\pi(\sigma | y, Z_{ht}, \beta, \{b_h\}, \{c_h\}, D^{-1})} * \frac{q(\sigma | \sigma', y, Z_{ht}, \beta, \{b_h\}, \{c_h\}, D^{-1})}{q(\sigma' | \sigma, y, Z_{ht}, \beta, \{b_h\}, \{c_h\}, D^{-1})} \right\},$$

We use the *tailored chain* as our choice of candidate generating density, as in Chib and Greenberg (1998). It is specified as

$$\sigma' = \mu + g,$$

where $\mu$ is a $p$-dimensional vector, taken to be the mode of $\log \pi(\sigma | y, \{Z_{ht}\}, \beta, \{b_h\}, \{c_h\}, d^{-1}, C^{-1})$ and $g \sim MVt(0, \tau V, \nu)$, where $V$ is the negative of the second derivative of log $\pi(\sigma | y, \{Z_{ht}\}, \beta, \{b_h\}, \{c_h\}, d^{-1}, C^{-1})$ evaluated at the mode. This approach leads to a well mixing Markov chain.

## 5. DESCRIPTION OF DATA

We employ IRI's scanner panel database on household purchases in twenty-five product categories in a metropolitan market in a large U.S. city. For our analysis, we pick twelve product categories: bacon, butter, coffee, cola, crackers, detergent, hot dogs, ice cream, non-cola beverages, sugar, toilet tissue and paper towels. These product categories[3] have been identified in the literature as being representative of the household's "shopping basket" (see Bell & Lattin, 1998). The dataset covers a period of two years from June 1991 to June 1993 and contains shopping visit information on 494 panelists across four different stores in an urban market. For each product category, the dataset contains information on marketing variables – price, in-store displays, and newspaper feature advertisements – at the SKU-level for each store/week.

Choosing households that bought at the two largest stores in the market (that collectively account for 90% of all shopping visits in the database) yields 488 households. From these households, we pick a random sample of 300 households making a total of 39,276 shopping visits at the two largest stores. This is done to keep the size of the dataset manageable. For those shopping visits when a household visits the store but does not purchase a particular product category, we compute marketing variables as share-weighted average values across all SKUs in the product category, where shares are household-specific and computed using the observed purchases of the household over the study period. Computing marketing variables using such share-weighting has precedence in the empirical marketing literature on category purchase incidence[4] (see, for example, Manchanda, Ansari & Gupta, 1999). Descriptive statistics pertaining to the marketing variables are provided in Table 1.

From Table 1 we can see that average display and feature activity is higher for purchase visits than for non-purchase visits, as expected, for all product categories. In terms of the magnitude of the difference in display and feature activity between purchase and non-purchase visits, the largest magnitude is observed for toilet tissue, suggesting that store merchandising activities strongly influence household purchase incidence for this product category. The smallest magnitudes are observed for ice-cream and non-cola beverages for display and feature respectively. Average prices are lower for purchase visits than for non-purchase visits, as expected, for ten out of the twelve product categories. By and large, these descriptive statistics are consistent with the economic notions of positive own-advertising elasticities, negative own-price

***Table 1.*** Descriptive Statistics on Marketing Variables
Number of households = 300, Number of shopping visits = 39,276.

A. Purchase visits

| Product | Price ($/RP) | Display | Feature | No. of Purchases |
|---|---|---|---|---|
| Bacon | 1.7915 | 0.2078 | 0.5338 | 2473 |
| Butter | 1.0425 | 0.1910 | 0.3079 | 5787 |
| Coffee | 1.9107 | 0.3174 | 0.3439 | 3022 |
| Cola | 0.6033 | 0.3999 | 0.4749 | 5099 |
| Crackers | 2.9236 | 0.2093 | 0.1280 | 4214 |
| Detergent | 0.8991 | 0.3550 | 0.2840 | 3159 |
| Hot dogs | 2.0753 | 0.1564 | 0.3832 | 3847 |
| Ice cream | 0.7196 | 0.0019 | 0.3964 | 4334 |
| Non-cola | 0.6654 | 0.1963 | 0.1340 | 5922 |
| Sugar | 0.4565 | 0.3681 | 0.3820 | 2275 |
| Tissue | 0.3041 | 0.4084 | 0.4457 | 5534 |
| Towels | 0.7386 | 0.3561 | 0.3544 | 4482 |

B. Non-purchase visits

| Product | Price ($/RP) | Display | Feature | No. of Visits |
|---|---|---|---|---|
| Bacon | 2.2949 | 0.0739 | 0.2333 | 36,803 |
| Butter | 1.1089 | 0.0686 | 0.1169 | 33,489 |
| Coffee | 2.0284 | 0.1074 | 0.0998 | 36,254 |
| Cola | 0.7080 | 0.1392 | 0.2306 | 34,177 |
| Crackers | 2.6717 | 0.1003 | 0.0569 | 35,062 |
| Detergent | 1.1150 | 0.0937 | 0.0547 | 36,117 |
| Hot dogs | 2.4145 | 0.0461 | 0.1612 | 35,429 |
| Ice cream | 0.8042 | 0.0008 | 0.1585 | 34,942 |
| Non-cola | 0.6736 | 0.1086 | 0.0779 | 33,354 |
| Sugar | 0.4456 | 0.1193 | 0.1197 | 37,001 |
| Tissue | 0.3369 | 0.1236 | 0.1345 | 33,742 |
| Towels | 0.8081 | 0.1159 | 0.1060 | 34,794 |

elasticities etc. From the last column of Table 1, we can see that the most frequently purchased product category is non-cola beverages (with butter coming second), while the most infrequently purchased product category is sugar (with bacon coming second).

In Table 2a, we report, in matrix form, the purchase frequencies for each product category along the diagonal and pair-wise purchase frequencies for each pair of product categories (i.e. the number of times each pair of product

***Table 2.***

A: Descriptive Statistics – Joint Purchase Frequencies

| | Bacon | Butter | Coffee | Cola | Crackers | Deterg. | Hot dogs | Ice cream | Non-cola | Sugar | Tissue | Towels |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bacon | **2473** | | | | | | | | | | | |
| Butter | 710 | **5787** | | | | | | | | | | |
| Coffee | 324 | 799 | **3022** | | | | | | | | | |
| Cola | 502 | 1316 | 620 | **5099** | | | | | | | | |
| Crackers | 428 | 1198 | 597 | 992 | **4214** | | | | | | | |
| Deterg. | 382 | 879 | 488 | 844 | 752 | **3159** | | | | | | |
| Hot dogs | 653 | 1091 | 468 | 927 | 737 | 608 | **3847** | | | | | |
| Ice cream | 415 | 1046 | 542 | 817 | 797 | 581 | 702 | **4334** | | | | |
| Non-cola | 624 | 1341 | 618 | 1726 | 1104 | 822 | 1018 | 1035 | **5922** | | | |
| Sugar | 338 | 772 | 328 | 478 | 431 | 359 | 453 | 400 | 573 | **2275** | | |
| Tissue | 719 | 1694 | 858 | 1389 | 1180 | 1127 | 1051 | 989 | 1467 | 661 | **5534** | |
| Towels | 490 | 1308 | 751 | 1220 | 1020 | 919 | 791 | 823 | 1221 | 552 | 1897 | **4482** |

B: Descriptive Statistics – Bivariate Correlations

| | Bacon | Butter | Coffee | Cola | Crackers | Deterg. | Hot dogs | Ice cream | Non-cola | Sugar | Tissue | Towels |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bacon | **1** | | | | | | | | | | | |
| Butter | 0.1022 | **1** | | | | | | | | | | |
| Coffee | 0.0526 | 0.0953 | **1** | | | | | | | | | |
| Cola | 0.0564 | 0.1207 | 0.0647 | **1** | | | | | | | | |
| Crackers | 0.0551 | 0.1339 | 0.0842 | 0.1089 | **1** | | | | | | | |
| Deterg. | 0.0706 | 0.1092 | 0.0860 | 0.1208 | 0.1249 | **1** | | | | | | |
| Hot dogs | 0.1449 | 0.1267 | 0.0553 | 0.1090 | 0.0897 | 0.0940 | **1** | | | | | |
| Ice cream | 0.0475 | 0.0934 | 0.0636 | 0.0615 | 0.0872 | 0.0694 | 0.0759 | **1** | | | | |
| Non-cola | 0.0736 | 0.0940 | 0.0433 | 0.2026 | 0.1077 | 0.0904 | 0.1048 | 0.0866 | **1** | | | |
| Sugar | 0.0874 | 0.1343 | 0.0626 | 0.0592 | 0.0658 | 0.0705 | 0.0844 | 0.0518 | 0.0701 | **1** | | |
| Tissue | 0.1116 | 0.1814 | 0.1187 | 0.1460 | 0.1386 | 0.1835 | 0.1253 | 0.0884 | 0.1294 | 0.1066 | **1** | |
| Towels | 0.0685 | 0.1463 | 0.1220 | 0.1520 | 0.1395 | 0.1644 | 0.0948 | 0.0839 | 0.1220 | 0.1002 | 0.2913 | **1** |

categories is purchased together) along the off-diagonal. For example, bacon is purchased on 2473 shopping visits, of which 710 are associated with the joint purchase of butter. This means that 28.7% of all bacon purchases are associated with joint purchase of butter. We report bivariate rank correlations, based on these purchase frequencies, in the lower half (i.e. below the main diagonal) of Table 2b. Cross-category correlations are fairly evident, with high magnitudes observed for two pairs: tissue and towels (0.2913), non-cola and cola beverages (0.2026).

All the observed correlations in Table 2b are positive. The reason for this is the large number of "zeros" that characterizes the vector of purchase outcomes for each product category. For example, among the 39,276 store visit observations in the dataset, only 5922 resulted in the purchase of non-cola beverages, 5534 resulted in the purchase of tissue, etc. This means that product categories appear to be complements for no reason other than the fact neither was purchased on a large number of purchase occasions. One way to "correct" for this is to recompute bivariate correlations for each pair after ignoring observations that resulted in a purchase of neither (let us call these "zero observations"). But this creates a problem of the opposite kind, i.e. all pairs of product categories appear to be substitutes on account of our ignoring a large number of outcomes when neither is purchased. However, the amount of distortion observed in the bivariate correlation for a given pair of product categories when its zero observations are ignored, is almost identical to the distortion observed for any other pair of product categories when their zero observations are ignored. This means that comparing bivariate correlations across pairs of product categories is meaningful, regardless of how we compute the correlations. For example, toilet tissue and towels have a much higher bivariate correlation than bacon and coffee regardless of whether or not we ignore each pair's zero observations. Armed with these preliminary findings, we next estimate our proposed econometric model on the basket data in order to estimate cross-category relationships after accommodating the effects of covariates, panel structure of the data etc. While estimating the proposed model, we include the following variables in the household-specific vector $X_{ht}$ (see Eq. 2) for each of the twelve product categories in the shopping basket.

1. Price
2. Feature
3. Display
4. Inventory

Price is a continuous variable, operationalized in dollars per ounce. Feature and display are indicator variables, that take the value 1 if the product is on feature or display respectively, and 0 otherwise. Inventory is a continuous variable (measured in ounces per week), which is computed using the household's product consumption rate which, in turn, is computed by dividing the total product quantity purchased by the household over the study period by the number of weeks in the data. For the first week in the data, each household is assumed to have enough inventory for that week, i.e. the inventory variable for a household at $t = 1$ is assumed to be the household's weekly product consumption rate. We incorporate random effects in the intercept terms for each product category.

## 6. EMPIRICAL RESULTS

We estimate the proposed basket-level model of purchase incidence decisions as well as five benchmark models, as shown below, in order to investigate the consequences of ignoring either cross-category correlations or unobserved heterogeneity across households.

*Model 1:* Multivariate Probit – Full twelve categories
*Model 2:* Multivariate Probit with unobserved heterogeneity restricted to be common across categories – Full twelve categories
*Model 3:* Multivariate Probit – Four categories only
*Model 4:* Multivariate Probit – Two categories only
*Model 5:* Independent Univariate Probits
*Model 6:* Multivariate Probit without unobserved heterogeneity

Comparing model 1 vs. model 2 allows one to investigate the consequences of restricting the unobserved heterogeneity distribution to be the same across product categories. For models 3 and 4, we retain the assumption of common unobserved heterogeneity distribution across product categories (as in model 2). Comparing model 2 vs. models 3 and 4 will demonstrate the consequences of modeling households' purchase incidence decisions only across subsets of the twelve product categories. For model 5, we assume the unobserved heterogeneity distribution to be different across product categories (as in model 1). Comparing model 1 vs. model 5 will demonstrate the consequences of modeling purchase incidence decisions jointly as opposed to separately across product categories. Comparing models 1 or 2 vs. model 6 will demonstrate the consequences of ignoring unobserved heterogeneity across households in a multivariate probit model.

First we look at the estimated inter-category correlation matrix based on the proposed multivariate probit model, allowing the unobserved heterogeneity distribution to be different across product categories (i.e. model 1). This is summarized in Table 3. The lower triangle reports the posterior means, while the upper triangle reports the 95% posterior credibility intervals (symmetric about the posterior mean). The off-diagonal terms in this table indicate that inter-category correlations are non-zero in general, with the correlations being quite large for specific pairs of product categories. For example, the estimated correlation in purchase incidence outcomes between cola and non-cola beverages has a mean of 0.4216 and a credibility interval of (0.40, 0.45). This indicates that households, rather than viewing cola and non-cola beverages as consumption substitutes, buy them together for complementary consumption needs, i.e. to maintain variety in their "beverage pantry." The estimated correlation is also large for hot dogs and bacon (0.3812), another possible consequence of the household's need for variety in the kitchen, this time among the meat products in their refrigerator. A third pair of product categories for which the estimated correlation is high is tissue and detergents (0.3744). This finding is especially interesting since there is little opportunity for a *sheer coincidence* effect, i.e. the two product categories frequently co-occurring in the household's shopping basket on account of having short inter-purchase cycles. In fact, inter-purchase times in these product categories are much larger, on average, than for other product categories in the data. One possible explanation for the large value of the estimated correlation is that since detergents and tissue are typically shelved close to each other in the grocery store, frequently in the same aisle, households have a propensity to pick up both products at the same time. One managerial implication of this "shelf effect" phenomenon is that the retailer may improve store profitability by shelving high-margin product categories close to products with short inter-purchase cycles so that every time a consumer picks up the latter off store shelves, she faces an opportunity to pick up the nearby high-margin product as well.

In Table 4, we report the estimated cross-category correlations using model 2 that assumes the unobserved heterogeneity distribution to be common across product categories. A comparison of Tables 3 and 4 indicates that cross-category correlations are, by and large, understated in Table 4 (i.e. model 2). To the extent that the common unobserved heterogeneity distribution across product categories captures correlations in households' purchase outcomes across categories,[5] one would indeed expect any remaining cross-category correlations in purchase outcomes to decrease after accounting for such unobserved heterogeneity.

In Table 5, we report the estimated cross-category correlations using model 3 that looks at four product categories at a time (as in Manchanda, Ansari & Gupta, 1999). A comparison of Tables 4 and 5 indicates that ignoring the

***Table 3.*** Estimated Pair-Wise Correlations across Product Categories – MVP on 12 Categories with Different Unobserved Heterogeneity across Categories (Model 1).[10]

| | Bacon | Butter | Coffee | Cola | Crackers | Deterg. | Hot dogs | Ice cream | Non-cola | Sugar | Tissue | Towels |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bacon | 1 | 0.19, 0.24 | 0.15, 0.24 | 0.15, 0.25 | 0.12, 0.21 | 0.13, 0.24 | 0.35, 0.41 | 0.10, 0.19 | 0.15, 0.22 | 0.19, 0.29 | 0.31, 0.38 | 0.20, 0.30 |
| Butter | 0.2177 | 1 | 0.18, 0.23 | 0.20, 0.26 | 0.23, 0.29 | 0.19, 0.26 | 0.27, 0.33 | 0.18, 0.24 | 0.15, 0.21 | 0.31, 0.38 | 0.29, 0.35 | 0.24, 0.31 |
| Coffee | 0.1936 | 0.2068 | 1 | 0.14, 0.19 | 0.18, 0.26 | 0.18, 0.27 | 0.19, 0.26 | 0.19, 0.26 | 0.12, 0.20 | 0.22, 0.32 | 0.27, 0.34 | 0.23, 0.32 |
| Cola | 0.2067 | 0.2291 | 0.1640 | 1 | 0.16, 0.20 | 0.17, 0.26 | 0.19, 0.25 | 0.15, 0.24 | 0.40, 0.45 | 0.10, 0.21 | 0.22, 0.30 | 0.24, 0.31 |
| Crackers | 0.1698 | 0.2612 | 0.2183 | 0.1820 | 1 | 0.16, 0.20 | 0.17, 0.23 | 0.16, 0.23 | 0.20, 0.27 | 0.16, 0.24 | 0.23, 0.30 | 0.24, 0.31 |
| Deterg. | 0.1938 | 0.2223 | 0.2246 | 0.2144 | 0.1791 | 1 | 0.15, 0.22 | 0.11, 0.21 | 0.16, 0.23 | 0.24, 0.31 | 0.34, 0.41 | 0.30, 0.37 |
| Hot dogs | 0.3812 | 0.2964 | 0.2226 | 0.2194 | 0.1992 | 0.1872 | 1 | 0.12, 0.17 | 0.19, 0.26 | 0.22, 0.32 | 0.25, 0.32 | 0.20, 0.29 |
| Ice cream | 0.1414 | 0.2089 | 0.2267 | 0.1930 | 0.1907 | 0.1609 | 0.1447 | 1 | 0.12, 0.16 | 0.10, 0.20 | 0.15, 0.23 | 0.16, 0.25 |
| Non-cola | 0.1816 | 0.1793 | 0.1651 | 0.4216 | 0.2355 | 0.1931 | 0.2253 | 0.1419 | 1 | 0.12, 0.17 | 0.23, 0.29 | 0.20, 0.26 |
| Sugar | 0.2362 | 0.3423 | 0.2709 | 0.1512 | 0.1999 | 0.2700 | 0.2667 | 0.1528 | 0.1476 | 1 | 0.22, 0.27 | 0.28, 0.37 |
| Tissue | 0.3449 | 0.3199 | 0.3064 | 0.2576 | 0.2654 | 0.3744 | 0.2900 | 0.1870 | 0.2592 | 0.2471 | 1 | 0.27, 0.31 |
| Towels | 0.2532 | 0.2742 | 0.2779 | 0.2756 | 0.2695 | 0.3351 | 0.2462 | 0.2004 | 0.2326 | 0.3214 | 0.2915 | 1 |

***Table 4.*** Estimated Pair-Wise Correlations across Product Categories – MVP on 12 Categories with Common Unobserved Heterogeneity across Categories (Model 2).[11]

| | Bacon | Butter | Coffee | Cola | Crackers | Deterg. | Hot dogs | Ice cream | Non-cola | Sugar | Tissue | Towels |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bacon | 1 | 0.19, 0.24 | 0.11, 0.20 | –0.01, 0.07 | 0.04, 0.12 | 0.11, 0.20 | 0.35, 0.41 | 0.09, 0.15 | 0.12, 0.19 | 0.23, 0.34 | 0.20, 0.27 | 0.06, 0.15 |
| Butter | 0.2205 | 1 | 0.18, 0.23 | 0.14, 0.19 | 0.22, 0.26 | 0.17, 0.25 | 0.21, 0.27 | 0.13, 0.18 | 0.08, 0.13 | 0.28, 0.33 | 0.27, 0.32 | 0.22, 0.27 |
| Coffee | 0.1475 | 0.2071 | 1 | 0.03, 0.10 | 0.16, 0.22 | 0.17, 0.23 | 0.12, 0.19 | 0.14, 0.20 | 0.02, 0.08 | 0.14, 0.23 | 0.21, 0.28 | 0.24, 0.33 |
| Cola | 0.0381 | 0.1627 | 0.0665 | 1 | 0.06, 0.14 | 0.12, 0.19 | 0.08, 0.14 | –0.03, 0.05 | 0.29, 0.33 | 0.00, 0.08 | 0.16, 0.23 | 0.21, 0.27 |
| Crackers | 0.0896 | 0.2401 | 0.1922 | 0.1020 | 1 | 0.23, 0.29 | 0.12, 0.17 | 0.14, 0.20 | 0.13, 0.19 | 0.11, 0.18 | 0.18, 0.24 | 0.21, 0.28 |
| Deterg. | 0.1519 | 0.2077 | 0.2055 | 0.1518 | 0.2550 | 1 | 0.12, 0.19 | 0.06, 0.13 | 0.09, 0.15 | 0.18, 0.25 | 0.30, 0.36 | 0.32, 0.37 |
| Hot dogs | 0.3819 | 0.2397 | 0.1486 | 0.1123 | 0.1408 | 0.1529 | 1 | 0.13, 0.19 | 0.14, 0.21 | 0.22, 0.29 | 0.15, 0.22 | 0.12, 0.18 |
| Ice cream | 0.1224 | 0.1536 | 0.1724 | 0.0172 | 0.1709 | 0.0910 | 0.1612 | 1 | 0.12, 0.17 | 0.08, 0.15 | 0.06, 0.12 | 0.08, 0.14 |
| Non-cola | 0.1554 | 0.1058 | 0.0538 | 0.3107 | 0.1582 | 0.1269 | 0.1758 | 0.1479 | 1 | 0.12, 0.18 | -0.15, 0.20 | 0.15, 0.21 |
| Sugar | 0.2821 | 0.3085 | 0.1941 | 0.0379 | 0.1513 | 0.2205 | 0.2510 | 0.1145 | 0.1521 | 1 | 0.22, 0.28 | 0.18, 0.25 |
| Tissue | 0.2340 | 0.2953 | 0.2464 | 0.1936 | 0.2105 | 0.3344 | 0.1875 | 0.0907 | 0.1707 | 0.2490 | 1 | 0.50, 0.54 |
| Towels | 0.1129 | 0.2491 | 0.2845 | 0.2354 | 0.2442 | 0.3467 | 0.1491 | 0.1125 | 0.1727 | 0.2104 | 0.5239 | 1 |

***Table 5.*** Estimated Pair-Wise Correlations across Product Categories – MVP on 4 Categories with Common Unobserved Heterogeneity across Categories (Model 3).[12]

| | Bacon | Butter | Coffee | Cola | Crackers | Deterg. | Hot dogs | Ice cream | Non-cola | Sugar | Tissue | Towels |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bacon | 1 | 0.16, 0.22 | 0.09, 0.18 | –0.07, 0.01 | | | | | | | | |
| Butter | 0.1927 | 1 | 0.14, 0.21 | 0.06, 0.12 | | | | | | | | |
| Coffee | 0.1388 | 0.1729 | 1 | –0.05, 0.03 | | | | | | | | |
| Cola | –0.0289 | 0.0877 | –0.0077 | 1 | | | | | | | | |
| Crackers | | | | | 1 | 0.18, 0.24 | 0.06, 0.13 | 0.06, 0.13 | | | | |
| Deterg. | | | | | 0.2114 | 1 | 0.08, 0.16 | –0.01, 0.06 | | | | |
| Hot dogs | | | | | 0.0985 | 0.1197 | 1 | 0.06, 0.13 | | | | |
| Ice cream | | | | | 0.1002 | 0.0296 | 0.0982 | 1 | | | | |
| Non-cola | | | | | | | | | 1 | 0.07, 0.14 | 0.07, 0.13 | 0.06, 0.13 |
| Sugar | | | | | | | | | 0.1044 | 1 | 0.16, 0.24 | 0.14, 0.21 |
| Tissue | | | | | | | | | 0.0949 | 0.2036 | 1 | 0.45, 0.51 |
| Towels | | | | | | | | | 0.0954 | 0.1761 | 0.4804 | 1 |

remaining eight product categories within the shopping basket understates the estimated correlation in purchase incidence decisions across the included four product categories. In fact, for two pairs of product categories ([cola & bacon] and [cola & coffee]), the estimated correlations are negative in the four-variate probit model even though they are positive in the twelve-variate probit model. For example, the posterior mean and credibility interval of the correlation for the pair [cola & bacon], based on model 3, are –0.0289 and (–0.0683, 0.0100) respectively. The corresponding measures based on model 2 are 0.0381 and (–0.0099, 0.0735) respectively. Similarly, the posterior mean and credibility interval of the correlation for the pair [cola & coffee], based on model 3, are –0.0077 and (–0.0446, 0.0264) respectively. The corresponding measures based on model 2 are 0.0665 and (0.0345, 0.1031) respectively. This indicates that if one were to use model 3, instead of model 2, one may falsely conclude, for example, that cola and coffee substitute for each other within the household's shopping basket when, in fact, they do not!

In Table 6, we report the estimated cross-category correlations using model 4 – that looks at pairs of product categories only (as in Chintagunta & Haldar, 1998) – for nine different pairs of product categories. A comparison of Tables 4 and 6 indicates that ignoring the remaining ten product categories within the shopping basket understates the estimated correlation in purchase incidence decisions for each pair of product categories. In fact, for three pairs of product categories – [cola & sugar], [cola & coffee], [cola & crackers] – the estimated correlations are negative in the bivariate probit model even though they are positive in the twelve-variate probit model. For example, the posterior mean and credibility interval of the correlation for the pair [cola & coffee], based on model 4, are –0.0733 and (–0.1100, –0.0367) respectively. The corresponding measures based on model 2 are 0.0665 and (0.0345, 0.1031) respectively. This indicates that if one were to use model 4, instead of model 2, one may falsely that cola and coffee, substitute each other within the household's shopping basket when, in fact, they do not! We summarize this finding below.

*Empirical Finding 1:* A limited operationalization of the multivariate probit model with panel structure, using a subset of the full set of product categories within the household's shopping basket (as in Chintagunta & Haldar, 1998; Manchanda, Ansari & Gupta, 1999), leads one to underestimate correlations in households' purchase incidence decisions across product categories. The estimated correlations even change signs (from positive to negative) in a few cases.

In Table 7, we report the estimated cross-category correlations using model 6 that ignores unobserved heterogeneity across households, i.e. a cross-sectional

***Table 6.*** Estimated Pair-Wise Correlations across Product Categories – MVP on 2 Categories with Common Unobserved Heterogeneity across Categories (Model 4).

| Model/ Categories | Bacon & Hot dogs | Butter & Sugar | Cola & Sugar | Cola & Non-cola | Detergent & Tissue | Detergent & Towels | Tissue & Towels | Cola & Coffee | Cola & Crackers |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 0.2498 | 0.2228 | –0.1172 | 0.1828 | 0.2166 | 0.2169 | 0.4050 | –0.0733 | –0.0190 |
| C.I.[13] | 0.20, 0.29 | 0.18, 0.26 | –0.18, –0.06 | 0.15, 0.21 | 0.17, 0.26 | 0.17, 0.26 | 0.37, 0.45 | –0.11, –0.04 | –0.05, 0.01 |

***Table 7.*** Estimated Pair-Wise Correlations across Product Categories – MVP on 12 Categories without Unobserved Heterogeneity (Model 6).[14]

| | Bacon | Butter | Coffee | Cola | Crackers | Deterg. | Hot dogs | Ice cream | Non-cola | Sugar | Tissue | Towels |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bacon | 1 | 0.31, 0.36 | 0.20, 0.28 | 0.17, 0.24 | 0.20, 0.26 | 0.26, 0.33 | 0.43, 0.49 | 0.16, 0.24 | 0.24, 0.31 | 0.31, 0.39 | 0.35, 0.40 | 0.24, 0.31 |
| Butter | 0.3360 | 1 | 0.30, 0.36 | 0.31, 0.36 | 0.35, 0.39 | 0.34, 0.39 | 0.33, 0.38 | 0.24, 0.29 | 0.23, 0.28 | 0.37, 0.44 | 0.42, 0.47 | 0.39, 0.43 |
| Coffee | 0.2372 | 0.3302 | 1 | 0.21, 0.26 | 0.27, 0.33 | 0.30, 0.37 | 0.21, 0.28 | 0.22, 0.28 | 0.15, 0.20 | 0.22, 0.30 | 0.35, 0.41 | 0.39, 0.45 |
| Cola | 0.2055 | 0.3344 | 0.2361 | 1 | 0.26, 0.30 | 0.30, 0.37 | 0.25, 0.31 | 0.15, 0.21 | 0.42, 0.46 | 0.16, 0.23 | 0.35, 0.40 | 0.40, 0.44 |
| Crackers | 0.2293 | 0.3715 | 0.3028 | 0.2798 | 1 | 0.36, 0.41 | 0.24, 0.30 | 0.25, 0.31 | 0.26, 0.31 | 0.23, 0.29 | 0.34, 0.40 | 0.39, 0.43 |
| Deterg. | 0.2946 | 0.3673 | 0.3337 | 0.3378 | 0.3873 | 1 | 0.26, 0.32 | 0.18, 0.25 | 0.25, 0.31 | 0.30, 0.37 | 0.46, 0.51 | 0.48, 0.53 |
| Hot dogs | 0.4599 | 0.3605 | 0.2486 | 0.2798 | 0.2733 | 0.2927 | 1 | 0.23, 0.28 | 0.26, 0.32 | 0.30, 0.36 | 0.31, 0.36 | 0.29, 0.35 |
| Ice cream | 0.2058 | 0.2682 | 0.2517 | 0.1781 | 0.2819 | 0.2214 | 0.2571 | 1 | 0.22, 0.27 | 0.15, 0.23 | 0.22, 0.26 | 0.24, 0.30 |
| Non-cola | 0.2710 | 0.2542 | 0.1759 | 0.4384 | 0.2862 | 0.2783 | 0.2955 | 0.2488 | 1 | 0.22, 0.29 | 0.31, 0.35 | 0.32, 0.36 |
| Sugar | 0.3492 | 0.4026 | 0.2624 | 0.1950 | 0.2595 | 0.3364 | 0.3307 | 0.1937 | 0.2483 | 1 | 0.33, 0.39 | 0.31, 0.38 |
| Tissue | 0.3743 | 0.4458 | 0.3804 | 0.3787 | 0.3700 | 0.4830 | 0.3357 | 0.2416 | 0.3260 | 0.3667 | 1 | 0.62, 0.66 |
| Towels | 0.2799 | 0.4104 | 0.4219 | 0.4209 | 0.4077 | 0.5021 | 0.3153 | 0.2730 | 0.3415 | 0.3529 | 0.6397 | 1 |

MVP model. A comparison of either Tables 3 or 4 vs. Table 7 indicates that ignoring unobserved heterogeneity across households overstates the estimated inter-category correlations. This finding is in the same spirit as findings in the brand choice literature that ignoring unobserved heterogeneity across households overstates the estimated serial correlation in the error terms in households' random utilities for brands (Allenby & Lenk, 1994; Keane, 1997). We summarize this finding below.

> *Empirical Finding 2:* Ignoring the effects of unobserved heterogeneity across households in the proposed multivariate probit model leads one to overestimate correlations in households' purchase incidence decisions across product categories.

In Tables 8 and 9 we summarize the estimated covariate effects for the twelve product categories based on the six model specifications. While the posterior means are reported in Table 8, the posterior credibility intervals are reported in Table 9. The second column in each table lists the results based on the proposed model estimated on the full set of twelve product categories (i.e. model 1). The estimates of the marketing mix coefficients and product inventory are signed as expected for all twelve categories. Specifically, the coefficients of price are always negative, the coefficients of display and feature are always positive and the coefficients of inventory are always negative. Among the twelve categories, cola beverages show maximum responsiveness to price (posterior mean of –2.1378), ice cream shows maximum responsiveness to store displays (posterior mean of 1.3978), while coffee shows maximum responsiveness to newspaper feature advertising (posterior mean of 1.0471).

The third column of Tables 8 and 9 lists the results based on the proposed model with the unobserved heterogeneity distribution restricted to be common across the twelve product categories (i.e. model 2). A comparison of the estimates in columns 2 and 3 (i.e. model 1 vs. model 2) indicates that household sensitivity to price and display are, by and large,[6] understated in model 2. In other words, restricting the unobserved heterogeneity distribution to be common across product categories leads one to conclude that households are less responsive to pricing and display activities. The feature coefficient, however, shows mixed results, i.e. it is understated for five categories and overstated for the remaining seven categories.

The fourth column of Tables 8 and 9 lists the results based on the proposed model estimated on three mutually exclusive subsets of four product categories (i.e. model 3). A comparison of the estimates in columns 3 and 4 (i.e. model 2 vs. model 3) indicates that household sensitivity to price, display and feature is, by and large,[7] overstated in model 3. Taken together with our earlier findings

*Table 8.* Posterior Means of Estimates of Covariate Effects.

| Parameter | **Model 1** MVP-12 categories Panel (Flexible) | **Model 2** MVP-12 categories Panel | **Model 3** MVP-4 categories Panel | **Model 4** MVP-2 categories[15] Panel | **Model 5** UVP-1 category Panel | **Model 6** MVP-12 categories Cross-sectional |
|---|---|---|---|---|---|---|
| Bacon Intercept | –1.0017 | –1.0816 | –1.0025 | –1.0890 | –0.9553 | –1.1214 |
| Bacon Price | –1.2178 | –0.9810 | –1.1417 | –1.1282 | –1.3549 | –0.8619 |
| Bacon Display | 0.5153 | 0.4904 | 0.5828 | 0.5062 | 0.5994 | 0.4726 |
| Bacon Feature | 0.3067 | 0.3891 | 0.3440 | 0.3912 | 0.2789 | 0.3746 |
| Bacon Inventory | –0.0398 | –0.0438 | –0.0446 | –0.0372 | –0.0381 | –0.0412 |
| Butter Intercept | –1.2751 | –1.2545 | –1.2835 | –1.2964 | –1.2963 | –1.2292 |
| Butter Price | –0.8621 | –0.2619 | –0.4032 | –0.4232 | –1.0287 | –0.1965 |
| Butter Display | 0.4809 | 0.4697 | 0.5164 | 0.5039 | 0.5141 | 0.4230 |
| Butter Feature | 0.7664 | 0.8624 | 0.8779 | 0.8675 | 0.7763 | 0.7858 |
| Butter Inventory | –0.0442 | –0.0425 | –0.0399 | –0.0408 | –0.0431 | –0.0400 |
| Coffee Intercept | –1.7580 | –1.6988 | –1.7452 | –2.2334 | –1.7828 | –1.6431 |
| Coffee Price | –0.1302 | –0.0056 | –0.0029 | –1.7977 | –0.1683 | –0.0039 |
| Coffee Display | 0.5877 | 0.4991 | 0.5231 | 0.7222 | 0.6104 | 0.4490 |
| Coffee Feature | 1.0471 | 1.0185 | 1.0697 | 0.2439 | 1.0779 | 0.9091 |
| Coffee Inventory | –0.0268 | –0.0308 | –0.0290 | –0.0037 | –0.0266 | –0.0285 |

*Table 8.* (Continued).

| Parameter | Model 1<br>MVP-12 categories<br>Panel (Flexible) | Model 2<br>MVP-12 categories<br>Panel | Model 3<br>MVP-4 categories<br>Panel | Model 4<br>MVP-2 categories[16]<br>Panel | Model 5<br>UVP-1 category<br>Panel | Model 6<br>MVP-12 categories<br>Cross-sectional |
|---|---|---|---|---|---|---|
| Cola Intercept | −2.4828 | −1.9734 | −2.1031 | −2.2340/−2.1801/−1.8149/2.2293 | −2.6756 | −1.8723 |
| Cola Price | −2.1378 | −1.3899 | −1.6094 | −1.8062/−1.5488/0.0022/−1.8067 | −2.4360 | −1.2687 |
| Cola Display | 0.7405 | 0.6995 | 0.7026 | 0.7228/0.7915/0.5408/0.7253 | 0.7826 | 0.6623 |
| Cola Feature | 0.2136 | 0.2341 | 0.2529 | 0.2238/0.2615/1.1369/0.2467 | 0.2081 | 0.1868 |
| Cola Inventory | −0.0043 | −0.0041 | −0.0038 | −0.0036/−0.0036/−0.0281/0.0037 | −0.0041 | −0.0039 |
| Crackers Inter. | −1.4098 | −1.3874 | −1.4256 | −1.3257 | −1.4504 | −1.4210 |
| Crackers Price | −0.0566 | −0.0015 | 0.0140 | −0.1612 | −0.0560 | 0.0674 |
| Crackers Display | 0.8078 | 0.7027 | 0.7464 | 0.8123 | 0.8742 | 0.6522 |
| Crackers Feature | 0.5128 | 0.4576 | 0.4590 | 0.5216 | 0.5618 | 0.4408 |
| Crackers Invent. | −0.0537 | −0.0562 | −0.0514 | −0.0480 | −0.0467 | −0.0562 |
| Deterg. Intercept | −1.8150 | −1.7058 | −1.7350 | −1.8472/−1.8614 | −1.8600 | −1.6292 |
| Deterg. Price | −0.9404 | −0.5361 | −0.5708 | −0.6082/−0.8654 | −1.1270 | −0.4770 |
| Deterg. Display | 0.7598 | 0.7244 | 0.7763 | 0.8224/0.8204 | 0.8165 | 0.6462 |
| Deterg. Feature | 0.8500 | 0.8493 | 0.8740 | 0.9606/0.9132 | 0.8444 | 0.7561 |
| Deterg. Inventory | −0.0118 | −0.0113 | −0.0103 | −0.0106/−0.0106 | −0.0100 | −0.0110 |

*Table 8.* (Continued).

| **Parameter** | **Model 1** MVP-12 categories Panel (Flexible) | **Model 2** MVP-12 categories Panel | **Model 3** MVP-4 categories Panel | **Model 4** MVP-2 categories[17] Panel | **Model 5** UVP-1 category Panel | **Model 6** MVP-12 categories Cross-sectional |
|---|---|---|---|---|---|---|
| Hot dogs Inter. | −0.7690 | −1.0369 | −1.0198 | −0.8788 | −0.7476 | −1.0735 |
| Hot dogs Price | −1.0381 | −0.5991 | −0.6643 | −0.8951 | −1.1282 | −0.5138 |
| Hot dogs Display | 0.6708 | 0.7109 | 0.7254 | 0.7580 | 0.7449 | 0.6808 |
| Hot dogs Feature | 0.5043 | 0.6012 | 0.6539 | 0.5394 | 0.5267 | 0.5591 |
| Hot dogs Invent. | −0.0212 | −0.0218 | −0.0230 | −0.0210 | −0.0221 | −0.0209 |
| Ice cream Inter. | −1.8438 | −1.5724 | −1.6193 | NA | −1.8740 | −1.5335 |
| Ice cream Price | −0.8738 | −0.3546 | −0.4206 | NA | −0.9170 | −0.2945 |
| Ice cream Display | 1.3978 | 1.4967 | 1.5398 | NA | 1.4796 | 1.3689 |
| Ice cream Feature | 0.8459 | 0.9286 | 0.9439 | NA | 0.8576 | 0.8847 |
| Ice cream Invent. | −0.0068 | −0.0085 | −0.0079 | NA | −0.0064 | −0.0087 |
| Non-cola Inter. | −1.4032 | −1.2917 | −1.3256 | −1.4185 | −1.4230 | −1.2653 |
| Non-cola Price | −0.2906 | −0.3977 | −0.3551 | −0.4391 | −0.3162 | −0.3672 |
| Non-cola Display | 0.5337 | 0.4406 | 0.5002 | 0.5240 | 0.5797 | 0.4237 |
| Non-cola Feature | 0.4000 | 0.3188 | 0.3494 | 0.2939 | 0.3829 | 0.2734 |
| Non-cola Invent. | −0.0001 | 0.0000 | −0.0003 | 0.0000 | −0.0002 | 0.0000 |

*Table 8.* (Continued).

| **Parameter** | **Model 1** MVP-12 categories Panel (Flexible) | **Model 2** MVP-12 categories Panel | **Model 3** MVP-4 categories Panel | **Model 4** MVP-2 categories[18] Panel | **Model 5** UVP-1 category Panel | **Model 6** MVP-12 categories Cross-sectional |
|---|---|---|---|---|---|---|
| Sugar Intercept | –2.2659 | –2.1551 | –2.2408 | –2.3497/–2.3294 | –2.3595 | –2.0453 |
| Sugar Price | –0.2970 | –0.3212 | –0.3215 | –0.4646/–0.3783 | –0.4029 | –0.2608 |
| Sugar Display | 0.6097 | 0.5896 | 0.6522 | 0.6049/0.6234 | 0.5411 | 0.5516 |
| Sugar Feature | 0.7451 | 0.6891 | 0.7074 | 0.6870/0.7417 | 0.7984 | 0.6316 |
| Sugar Invent. | –0.0282 | –0.0328 | –0.0323 | –0.0310/–0.0309 | –0.0280 | –0.0319 |
| Tissue Inter. | –2.7379 | –1.7428 | –1.9260 | –2.0770/–2.1892 | –3.0543 | –1.6315 |
| Tissue Price | –0.9703 | –0.2505 | –0.3523 | –0.4274/–0.4962 | –1.1890 | –0.2043 |
| Tissue Display | 0.7121 | 0.6936 | 0.7371 | 0.7819/0.7854 | 0.7749 | 0.6303 |
| Tissue Feature | 0.8078 | 0.8388 | 0.8819 | 0.9358/0.9241 | 0.8428 | 0.7489 |
| Tissue Invent. | –0.0094 | –0.0090 | –0.0087 | –0.0090/–0.0087 | –0.0090 | –0.0085 |
| Towels Inter. | –2.1006 | –1.6328 | –1.6993 | –1.8862/–1.8505 | –2.1975 | –1.5480 |
| Towels Price | –1.4211 | –0.4173 | –0.3848 | –0.8690/–0.5492 | –1.5782 | –0.3431 |
| Towels Display | 0.8045 | 0.7407 | 0.7933 | 0.8524/0.8361 | 0.8771 | 0.6629 |
| Towels Feature | 0.7676 | 0.8020 | 0.8736 | 0.8505/0.9273 | 0.8069 | 0.7168 |
| Towels Invent. | –0.0241 | –0.0296 | –0.0282 | –0.0243/–0.0272 | –0.0207 | –0.0284 |

*Table 9.* Posterior Credibility Intervals of Estimates of Covariate Effects.

| **Parameter** | **Model 1** MVP-12 categories Panel (Flexible) | **Model 2** MVP-12 categories Panel | **Model 3** MVP-4 categories Panel | **Model 4** MVP-2 categories[19] Panel | **Model 5** UVP-1 category Panel | **Model 6** MVP-12 categories Cross-sectional |
|---|---|---|---|---|---|---|
| Bacon Intercept | −1.0936, −0.9041 | −1.1504, −1.0067 | −1.0850, −0.9142 | −1.1815, −0.9959 | −1.0603, −0.8482 | −1.1798, −1.0588 |
| Bacon Price | −1.3062, −1.1244 | −1.0551, −0.9105 | −1.2169, −1.0725 | −1.2156, −1.0493 | −1.4506, −1.2702 | −0.9351, −0.7947 |
| Bacon Display | 0.4311, 0.5975 | 0.4128, 0.5678 | 0.5087, 0.6615 | 0.4076, 0.5887 | 0.5108, 0.6935 | 0.4004, 0.5470 |
| Bacon Feature | 0.2328, 0.3826 | 0.3227, 0.4561 | 0.2734, 0.4174 | 0.3181, 0.4660 | 0.2072, 0.3585 | 0.3100, 0.4327 |
| Bacon Inventory | −0.0488, −0.0312 | −0.0535, −0.0353 | −0.0543, −0.0349 | −0.0456, −0.0281 | −0.0474, −0.0282 | −0.0495, −0.0325 |
| Butter Intercept | −1.3414, −1.2129 | −1.3010, −1.2071 | −1.3361, −1.2304 | −1.3569, −1.2364 | −1.3771, −1.2174 | −1.2479, −1.2104 |
| Butter Price | −0.9395, −0.7822 | −0.3033, −0.2180 | −0.4470, −0.3530 | −0.4864, −0.3623 | −1.1203, −0.9439 | −0.2369, −0.1573 |
| Butter Display | 0.4022, 0.5601 | 0.4007, 0.5402 | 0.4392, 0.5934 | 0.4286, 0.5782 | 0.4280, 0.5990 | 0.3577, 0.4855 |
| Butter Feature | 0.6994, 0.8368 | 0.8046, 0.9198 | 0.8193, 0.9400 | 0.8059, 0.9332 | 0.7077, 0.8432 | 0.7297, 0.8421 |
| Butter Inventory | −0.0501, −0.0383 | −0.0487, −0.0366 | −0.0462, −0.0331 | −0.0475, −0.0345 | −0.0495, −0.0363 | −0.0460, −0.0341 |
| Coffee Intercept | −1.8316, −1.6868 | −1.7509, −1.6433 | −1.8048, −1.6834 | −1.8831, −1.7423 | −1.8692, −1.6986 | −1.6770, −1.6103 |
| Coffee Price | −0.1903, −0.0654 | −0.0423, 0.0287 | −0.0409, 0.0360 | −0.0394, 0.0433 | −0.2320, −0.1062 | −0.0385, 0.0305 |
| Coffee Display | 0.4999, 0.6714 | 0.4225, 0.5732 | 0.4447, 0.6044 | 0.4560, 0.6278 | 0.5260, 0.7011 | 0.3795, 0.5246 |
| Coffee Feature | 0.9643, 1.1311 | 0.9429, 1.0947 | 0.9814, 1.1509 | 1.0565, 1.2234 | 0.9924, 1.1633 | 0.8374, 0.9789 |
| Coffee Inventory | −0.0314, −0.0222 | −0.0352, −0.0256 | −0.0345, −0.0241 | 0.0331, −0.0234 | −0.0312, −0.0218 | −0.0333, −0.0240 |

*Table 9.* (Continued).

| Parameter | Model 1<br>MVP-12 categories<br>Panel (Flexible) | Model 2<br>MVP-12 categories<br>Panel | Model 3<br>MVP-4 categories<br>Panel | Model 4<br>MVP-2 categories[20]<br>Panel | Model 5<br>UVP-1 category<br>Panel | Model 6<br>MVP-12 categories<br>Cross-sectional |
|---|---|---|---|---|---|---|
| Cola Intercept | −2.5836, −2.3924 | −2.0267, −1.9162 | −2.1649, −2.0367 | −2.3, −2.2; −2.3, −2.1;2.3, −2.2;2.3, −2.2 | −2.7976, −2.5577 | −1.9040, −1.8395 |
| Cola Price | −2.2646, −2.0206 | −1.4693, −1.3144 | −1.7015, −1.5141 | −1.9, −1.7; −1.6, −1.4; −1.9, −1.7; −1.9, −1.7 | −2.5667, −2.3019 | −1.3414, −1.1975 |
| Cola Display | 0.6671, 0.8141 | 0.6380, 0.7634 | 0.6408, 0.7673 | 0.6, 0.8; 0.7, 0.9; 0.6, 0.8; 0.7, 0.8 | 0.7095, 0.8606 | 0.6065, 0.7160 |
| Cola Feature | 0.1453, 0.2770 | 0.1753, 0.2908 | 0.1887, 0.3167 | 0.2, 0.3; 0.2, 0.3; 0.2, 0.3; 0.2, 0.3 | 0.1346, 0.2794 | 0.1329, 0.2417 |
| Cola Inventory | −0.0056, −0.0030 | −0.0052, −0.0028 | −0.0050, −0.0025 | −0.0, −0.0; −0.0, −0.0; −0.0, −0.0; −0.0, −0.0 | −0.0054, −0.0028 | −0.0051, −0.0027 |
| Crackers Inter. | −1.5041, −1.3075 | −1.4555, −1.3241 | −1.4994, −1.3534 | −1.4155, −1.2336 | −1.5480, −1.3577 | −1.4656, −1.3743 |
| Crackers Price | −0.1249, 0.0074 | −0.0497, 0.0452 | −0.0403, 0.0671 | −0.2202, −0.1039 | −0.1240, 0.0124 | 0.0240, 0.1095 |
| Crackers Display | 0.7322, 0.8820 | 0.6313, 0.7818 | 0.6762, 0.8219 | 0.7319, 0.8935 | 0.7901, 0.9561 | 0.5877, 0.7177 |
| Crackers Feature | 0.4135, 0.6100 | 0.3728, 0.5411 | 0.3707, 0.5505 | 0.4226, 0.6180 | 0.4641, 0.6604 | 0.3554, 0.5216 |
| Crackers Invent. | −0.0641, −0.0437 | −0.0672, −0.0458 | −0.0618, −0.0414 | −0.0588, −0.0370 | −0.0570, −0.0359 | −0.0658, −0.0465 |
| Deterg. Intercept | −1.8841, −1.7476 | −1.7540, −1.6553 | −1.7859, −1.6813 | (−1.92, −1.78) (−1.94, −1.78) | −1.9389, −1.7821 | −1.6509, −1.6077 |
| Deterg. Price | −1.0473, −0.8332 | −0.6024, −0.4810 | −0.6392, −0.5003 | (−0.69, −0.53) (−0.95, −0.79) | −1.2451, −1.0198 | −0.5380, −0.4157 |
| Deterg. Display | 0.6813, 0.8342 | 0.6503, 0.8098 | 0.6952, 0.8527 | (0.74, 0.90) (0.74, 0.90) | 0.7329, 0.9024 | 0.5777, 0.7114 |
| Deterg. Feature | 0.7578, 0.9449 | 0.7689, 0.9409 | 0.7916, 0.9571 | (0.87, 1.06) (0.82, 1.02) | 0.7488, 0.9419 | 0.6777, 0.8374 |
| Deterg. Inventory | −0.0141, −0.0095 | −0.0133, −0.0092 | −0.0124, −0.0081 | (−0.01, −0.00) (−0.01, −0.00) | −0.0122, −0.0077 | −0.0130, −0.0090 |

***Table 9.*** (Continued).

| Parameter | Model 1 MVP-12 categories Panel (Flexible) | Model 2 MVP-12 categories Panel | Model 3 MVP-4 categories Panel | Model 4 MVP-2 categories[21] Panel | Model 5 UVP-1 category Panel | Model 6 MVP-12 categories Cross-sectional |
|---|---|---|---|---|---|---|
| Hot dogs Inter. | −0.8627, −0.6686 | −1.1039, −0.9672 | −1.0960, −0.9466 | −0.9733, −0.7849 | −0.8546, −0.6407 | −1.1225, −1.0250 |
| Hot dogs Price | −1.1229, −0.9584 | −0.6527, −0.5465 | −0.7220, −0.6048 | −0.9689, −0.8184 | −1.2101, −1.0403 | −0.5641, −0.4636 |
| Hot dogs Display | 0.5719, 0.7660 | 0.6227, 0.8051 | 0.6353, 0.8207 | 0.6574, 0.8602 | 0.6451, 0.8478 | 0.5980, 0.7583 |
| Hot dogs Feature | 0.43467, 0.5739 | 0.5448, 0.6649 | 0.5903, 0.7189 | 0.4763, 0.6062 | 0.4553, 0.6031 | 0.5041, 0.6179 |
| Hot dogs Invent. | −0.0279, −0.0144 | −0.0295, −0.0145 | −0.0299, −0.0156 | −0.0281, −0.0136 | −0.0291, −0.0147 | −0.0287, −0.0133 |
| Ice cream Inter. | −1.9251, −1.7616 | −1.6219, −1.5214 | −1.6722, −1.5667 | NA | −1.9596, −1.7911 | −1.5580, −1.5101 |
| Ice cream Price | −0.9517, −0.7956 | −0.3988, −0.3060 | −0.4739, −0.3715 | NA | −0.9997, −0.8368 | −0.3348, −0.2538 |
| Ice cream Display | 0.8224, 2.0111 | 0.9359, 2.0117 | 0.9918, 2.1033 | NA | 0.8719, 2.0769 | 0.8542, 1.8706 |
| Ice cream Feature | 0.7816, 0.9059 | 0.8730, 0.9851 | 0.8890, 1.0004 | NA | 0.7946, 0.9239 | 0.8280, 0.9378 |
| Ice cream Invent. | −0.0087, −0.0048 | −0.0104, −0.0067 | −0.0098, −0.0058 | NA | −0.0083, −0.0045 | −0.0106, −0.0068 |
| Non-cola Inter. | −1.4893, −1.3179 | −1.3422, −1.2400 | −1.3842, −1.2691 | −1.4994, −1.3454 | −1.5053, −1.3408 | −1.2904, −1.2397 |
| Non-cola Price | −0.3571, −0.2168 | −0.4507, −0.3453 | −0.4132, −0.2961 | −0.5038, −0.3761 | −0.3900, −0.2451 | −0.4156, −0.3202 |
| Non-cola Display | 0.4617, 0.6038 | 0.3789, 0.4996 | 0.4344, 0.5649 | 0.4561, 0.5909 | 0.5076, 0.6502 | 0.3677, 0.4832 |
| Non-cola Feature | 0.3235, 0.4776 | 0.2467, 0.4020 | 0.2683, 0.4302 | 0.2180, 0.3714 | 0.3038, 0.4623 | 0.2047, 0.3441 |
| Non-cola Invent. | −0.0009, 0.0008 | −0.0009, 0.0009 | −0.0012, 0.0007 | −0.0008, 0.0009 | −0.0010, 0.0007 | −0.0009, 0.0008 |

*Table 9.* (Continued).

| Parameter | **Model 1** MVP-12 categories Panel (Flexible) | **Model 2** MVP-12 categories Panel | **Model 3** MVP-4 categories Panel | **Model 4** MVP-2 categories[22] Panel | **Model 5** UVP-1 category Panel | **Model 6** MVP-12 categories Cross-sectional |
|---|---|---|---|---|---|---|
| Sugar Intercept | −2.3604, −2.1714 | −2.2283, −2.0822 | −2.3219, −2.1594 | −2.44, −2.25; −2.42, −2.23 | −2.4612, −2.2670 | −2.1016, −1.9880 |
| Sugar Price | −0.3816, −0.2132 | −0.3950, −0.2515 | −0.3944, −0.2480 | −0.54, −0.38; −0.46, −0.29 | −0.4884, −0.3158 | −0.3241, −0.1961 |
| Sugar Display | 0.5162, 0.6986 | 0.5097, 0.6664 | 0.5723, 0.7352 | 0.51, 0.69; −0.53, 0.71 | 0.4659, 0.6163 | 0.4813, 0.6273 |
| Sugar Feature | 0.6569, 0.8359 | 0.6028, 0.7703 | 0.6206, 0.7916 | 0.59, 0.78; 0.66, 0.84 | 0.7155, 0.8860 | 0.5564, 0.7064 |
| Sugar Invent. | −0.0328, −0.0238 | −0.0373, −0.0283 | −0.0372, −0.0273 | −0.03, −0.02; −0.04, −0.03 | −0.0324, −0.0235 | −0.0368, −0.0272 |
| Tissue Inter. | −2.8706, −2.5903 | −1.8083, −1.6740 | −2.0100, −1.8381 | −2.18, −1.97; −2.29, −2.09 | −3.2222, −2.8977 | −1.6802, −1.5830 |
| Tissue Price | −1.0593, −0.8718 | −0.2998, −0.2002 | −0.4086, −0.2978 | −0.49, −0.36; −0.56, −0.44 | −1.3030, −1.0819 | −0.2468, −0.1616 |
| Tissue Display | 0.6394, 0.7813 | 0.6282, 0.7601 | 0.6712, 0.8020 | 0.71, 0.85; 0.71, 0.86 | 0.7031, 0.8511 | 0.5711, 0.6872 |
| Tissue Feature | 0.7383, 0.8762 | 0.7714, 0.8985 | 0.8165, 0.9501 | 0.86, 1.01; 0.86, 0.99 | 0.7677, 0.9152 | 0.6896, 0.8091 |
| Tissue Invent. | −0.0112, −0.0075 | −0.0107, −0.0071 | −0.0106, −0.0068 | −0.01, −0.00; −0.01, −0.00 | −0.0109, −0.0072 | −0.0100, −0.0069 |
| Towels Inter. | −2.2015, −1.9771 | −1.6806, −1.5804 | −1.7575, −1.6413 | −1.97, −1.81; −1.94, −1.77 | −2.3115, −2.0882 | −1.5722, −1.5223 |
| Towels Price | −1.5576, −1.2715 | −0.4939, −0.3420 | −0.4601, −0.3055 | −0.97, −0.78; −0.63, −0.46 | −1.7236, −1.4464 | −0.4050, −0.2757 |
| Towels Display | 0.7374, 0.8719 | 0.6797, 0.8045 | 0.7285, 0.8593 | 0.78, 0.93; 0.77, 0.91 | 0.8022, 0.9502 | 0.6071, 0.7193 |
| Towels Feature | 0.6947, 0.8425 | 0.7397, 0.8644 | 0.8048, 0.9435 | 0.77, 0.93; 0.85, 1.00 | 0.7287, 0.8851 | 0.6607, 0.7729 |
| Towels Invent. | −0.0298, −0.0183 | −0.0358, −0.0238 | −0.0340, −0.0222 | −0.03, −0.02; −0.03, −0.02 | −0.0266, −0.0148 | −0.0337, −0.0233 |

that model 3 understates correlations across product categories, this means that the marketing mix variables bear the burden of explaining purchase incidence decisions in product categories that are, in part, due to inter-category correlations (that are incompletely accounted for in the model).

The fifth column in Tables 8 and 9 lists the results from a bivariate version of the proposed model estimated separately on seven different pairs of product categories (i.e. model 4). A comparison of the estimates in columns 3 and 5 (i.e. model 2 vs. model 4) indicates that household sensitivity to price, display and feature is, by and large,[8] overstated in model 4. This finding is consistent with that obtained from comparing models 2 and 3, as discussed in the previous paragraph.

The sixth column in Tables 8 and 9 lists the results from univariate binary probit models estimated separately for the twelve product categories (i.e. model 5). A comparison of the estimates in columns 3 and 6 (i.e. model 2 vs. model 5) indicates that household sensitivity to price, display and feature is, by and large,[9] overstated in model 5. This finding, consistent with the findings obtained by comparing either models 2 and 3 or models 2 and 4, is summarized below.

> *Empirical Finding 3:* A limited operationalization of the proposed multivariate probit model, using a subset of the full set of product categories within the household's shopping basket (as in Chintagunta & Haldar, 1998; Manchanda, Ansari & Gupta, 1999), leads one to overestimate the effects of marketing variables on households' purchase incidence decisions within each product category.

The seventh column in Tables 8 and 9 lists the results from a purely cross-sectional version – one that ignores unobserved heterogeneity across households – of the proposed multivariate probit model (i.e. model 6). A comparison of the estimates in columns 3 and 7 (i.e. model 2 vs. model 6) seems to indicate that household sensitivity to price, display and feature are overstated in model 6. However, such an interpretation must be kept in check on account of a scale incompatibility problem while comparing models 2 and 6, since the cross-sectional probit (i.e. model 6) does not accommodate random effects across households.

## 7. SUMMARY

We propose a multivariate probit model with unobserved heterogeneity to explain households' purchase incidence decisions simultaneously across all product categories within their shopping baskets. We estimate the proposed model using basket-level purchase data on a scanner panel of 300 households. We find that a limited operationalization of the proposed model, using a subset of the full set of product categories within the household's shopping basket, leads one to underestimate inter-category correlations and overestimate the effectiveness of marketing variables. We also find that ignoring unobserved heterogeneity across households leads one to overestimate inter-category correlations and underestimate the effectiveness of marketing variables.

One obvious managerial benefit of our proposed model is that retailer can design optimal prices simultaneously across all product categories, taking cross-category correlations into account, in order to maximize store profits. When cross-category correlations exist, ignoring their effects and maximizing category profits independently across product categories will lead to sub-optimal profits. While the findings of this paper are of managerial interest in and of themselves, the implications of these findings on related household decisions, such as brand choice, are of managerial interest. We are currently extending our proposed model to accommodate households' brand choice decisions within each product category. In this framework, we employ a multinomial logit model for households' conditional brand choices within each product category, coupled with a multivariate probit model of households' purchase incidence decisions across product categories. Whether our reported findings about cross-category correlations in purchase incidence decisions in this paper generalize to such a fully specified framework is an area of ongoing investigation.

Last, but not the least, it will be useful to accommodate unobserved heterogeneity along multiple dimensions (instead of in the intercept term only) and model correlations not only in the error terms but also in household response parameters across product categories. This will allow us to investigate whether households exhibit similar sensitivities to the marketing variables in different product categories using a basket-level analysis (Seetharaman, Ainslie & Chintagunta, 1999 investigate this issue using conditional brand choice data on a panel of households in five product categories).

## NOTES

1. To the extent that product categories within a household's shopping basket vie for a limited shopping budget of the household, the budget constraint induces cross-category dependencies as well.
2. In our application, $J = 12$, $k = 5$ which makes the total number of estimated parameters 138.
3. The excluded product categories are barbecue sauce, cat food, cereals, cleansers, cookies, eggs, nuts, pills, pizza, snacks, soap, softener, yogurt.
4. In a companion paper, in which we model both category purchase incidence and brand choice, we explicitly investigate the consequences of such aggregation on model-based inferences.

5. We thank an anonymous reviewer for alerting us to this issue.
6. Except for display coefficients for hot dogs and ice cream, and price coefficients for non-cola and sugar, this holds for the remaining twenty coefficients.
7. Except for bacon's feature coefficient and the price coefficients for coffee, non-cola and towels, this holds for the remaining thirty-two marketing mix coefficients.
8. This holds for forty-eight out of the fifty-four marketing mix variables in tables 8–11.
9. The overstatement holds for twenty-eight out of thirty-six coefficients.
10. Lower and upper halves of the matrix contain posterior means and credibility intervals respectively.
11. Lower and upper halves of the matrix contain posterior means and credibility intervals respectively.
12. Lower and upper halves of the matrix contain posterior means and credibility intervals respectively.
13. Credibility Interval
14. Lower and upper halves of the matrix contain posterior means and credibility intervals respectively.
15. The estimates of bacon and butter are based on hot dogs and sugar as the respective second categories.
16. The four sets of estimates for cola are based on sugar, non-cola, coffee and crackers respectively as the second category. The estimates for crackers are based on cola as the second category. The two sets of estimates for detergents are based on tissue and towels respectively as the second category.
17. The estimates of hot dogs and non-cola are based on bacon and cola as the respective second categories.
18. The two sets of estimates of sugar are based on butter and cola respectively as the second category. The two sets of estimates for tissue are based on detergents and towels respectively as the second category. The two sets of estimates for towels are based on detergents and tissue respectively as the second category.
19. The estimates of bacon and butter are based on hot dogs and sugar as the respective second categories.
20. The four sets of estimates for cola are based on sugar, non-cola, coffee and crackers respectively as the second category. The estimates for crackers are based on cola as the second category. The two sets of estimates for detergents are based on tissue and towels respectively as the second category.
21. The estimates of hot dogs and non-cola are based on bacon and cola as the respective second categories.
22. The two sets estimates of sugar are based on butter and cola respectively as the second category. The two sets of estimates for tissue are based on detergents and towels respectively as the second category. The two sets of estimates for towels are based on detergents and tissue respectively as the second category.

## ACKNOWLEDGMENTS

We thank marketing seminar participants at the John M. Olin School of Business at Washington University for their comments on an earlier version of this paper.

# REFERENCES

Albert, J., & Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association, 88*, 669–679.

Allenby, G. M., & Lenk, P. J. (1994). Modeling Household Purchase Behavior with Logistic Normal Regression. *Journal of the American Statistical Association, 89*(428), 1–14.

Ashford, J. R., & Sowden, R. R. (1970). Multivariate Probit Analysis. *Biometrics, 26*, 535–546.

Bell, D. R., & Lattin, J. (1998). Shopping Behavior and Consumer Preference for Store Price Format: Why 'Large Basket' Shoppers Prefer EDLP. *Marketing Science, 17*(1), 66–88.

Berry, J. A., & Linoff, G. (1997). *Data Mining Techniques*. Wiley and Sons.

Bucklin, R. E., & Lattin, J. (1991). A Two-State Model of Purchase Incidence and Brand Choice. *Marketing Science, 10*(1), 24–39.

Chiang, J. (1991). A Simultaneous Approach to the Whether, What and How Much to Buy Questions. *Marketing Science, 10*(4), 297–315.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician, 49*(4), 327–335.

Chib, S., & Greenberg, E. (1998). Analysis of Multivariate Probit Models. *Biometrika, 85*(2), 347–361.

Chintagunta, P. K., & Haldar, S. (1998). Investigating Purchase Timing Behavior in Two Related Product Categories. *Journal of Marketing Research, 35*(1), 43–53.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association, 85*(410), 398–409.

Gonul, F., & Srinivasan, K. (1993). Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues. *Marketing Science, 12*(3), 213–229.

Keane, M. P. (1997). Modeling Heterogeneity and State Dependence in Consumer Choice Behavior. *Journal of Business and Economic Statistics, 15*(3), 310–327.

Manchanda, P., Ansari, A., & Gupta, S. (1999). The Shopping Basket: A Model for Multicategory Purchase Incidence Decisions. *Marketing Science, 18*(2), 95–114.

Seetharaman, P. B., Ainslie, A. K., & Chintagunta, P. K. (1999). Investigating Household State Dependence Effects Across Product Categories. *Journal of Marketing Research, 36*(4), 488–500.

Tanner, M. A., & Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association, 82*(398), 528–540.

Tierney, L. (1987). Markov Chains for Exploring Posterior Distributions. *Annals of Statistics, 22*, 1701–1762.

# ADVANCES IN OPTIMUM EXPERIMENTAL DESIGN FOR CONJOINT ANALYSIS AND DISCRETE CHOICE MODELS

Heiko Großmann, Heinz Holling and Rainer Schwabe

## ABSTRACT

*The authors review current developments in experimental design for conjoint analysis and discrete choice models emphasizing the issue of design efficiency. Drawing on recently developed optimal paired comparison designs, theoretical as well as empirical evidence is provided that established design strategies can be improved with respect to design efficiency.*

## 1. INTRODUCTION

The modeling of consumer preferences and choice behavior is one of the most prosperous areas of research in marketing (Carroll & Green, 1995; Ben-Akiva et al., 1997). Over the years a wealth of models has emerged for describing the joint effect of multiple attributes on consumers' product evaluations and choices. Among these models the most prominent approaches are conjoint analysis (Green & Srinivasan, 1978, 1990) and discrete choice (Ben-Akiva & Lerman, 1985). The development of conjoint analysis was inspired by the invention of conjoint measurement in psychology (Luce & Tukey, 1964), which

**Econometric Models in Marketing, Volume 16, pages 93–117.**

**ISBN: 0-7623-0857-5**

was introduced to a larger audience in marketing by Green and Rao (1971). The roots of discrete choice models can be traced back to several previous approaches to utility measurement like Thurstone's law of comparative judgment (Thurstone, 1927), the choice model of Luce (1959), and also to random utility theory (Manski, 1977).

Conjoint analysis and discrete choice models differ with respect to the response formats and the statistical models they use for collecting and analyzing data. In a traditional decompositional conjoint analysis task, attribute profiles usually have to be rank ordered with respect to preference or to be rated on a preference scale. Also, as another mode of data collection in conjoint analysis graded paired comparisons are in widespread use. In contrast, choices from sets of attribute profiles are observed in discrete choice models. As far as statistical models are concerned, conjoint analysis draws on the general linear model whereas discrete choice models are non-linear and usually of the logistic type.

When implementing a conjoint analysis or discrete choice study it has to be decided how many and which profiles are to be presented for evaluation. This task of selecting a number of settings from an experimental domain represents a typical problem of experimental design. Traditionally, marketing researchers have primarily been concerned with modeling issues. Much effort has been devoted to the development of realistic and apparently complex discrete choice models. For example, a number of models have been proposed that incorporate consumer segments or cross effects of brands on each other. The primary design concern with these complex models has been to choose the profiles in such a way that the model parameters can be estimated.

If there are many attributes or levels in the conjoint analysis or discrete choice model the number of profiles that have to be evaluated soon becomes large. To reduce the number of evaluations required from respondents, standard principles from linear model design theory like orthogonality and balance have been used routinely to construct designs for conjoint analysis as well as discrete choice models. When the reduction is to be performed in such a way that the model parameters can be estimated in the most efficient way techniques from optimum experimental design theory can be used. From time to time the issue of optimum design has received some attention in the econometric literature (see e.g. Aigner, 1979; Müller & Ponce de Leon, 1996). However, most of the developments in this growing field of statistics are only scarcely recognized by researchers in economics.

The aim of this contribution is to review basic principles of, and recent developments in, optimum experimental design theory that can be applied to conjoint analysis and discrete choice models. In the next section we give an overview of the statistical theory. This will be followed by a review of the literature on conjoint analysis and discrete choice designs in marketing. In the subsequent two sections a theoretical as well as an empirical comparison of designs will be provided.

## 2. OPTIMAL AND EFFICIENT DESIGNS

For the observational outcome of a random event one has to distinguish between active and passive observations. The latter situation occurs in observational studies where usually the investigator has no possibility to take influence on the outcome by adjusting explanatory variables.

In contrast to that, these explanatory variables are often called exogenous quantities for which it seems reasonable that different values or levels may be chosen. These active experimental situations bear the great advantage that the performance of the statistical inference can be substantially improved by a smart design for the settings of the exogenous variables. For example, the slope of a regression line can be estimated with a three times higher precision when an optimal design is used compared to uniform spacing.

The theory of optimal designs has been extensively developed during the last fifty years for various forms of a functional relationship $Y(x)=\eta(x, \theta)+\varepsilon$. In this formula $\eta$ is a known response function describing the structural dependence of the endogenous variable (response) $Y$ on the exogenous quantities $x$. Here $\theta=(\theta_1, \ldots, \theta_p)^{\mathrm{T}}$ is a vector of unknown parameters specifying the shape $\eta(\cdot, \theta)$ of the response and the exogenous quantities $x=(x^{(1)}, \ldots, x^{(k)})$ consist of $k$ different components $x^{(1)}, \ldots, x^{(k)}$. Finally, the observation is disturbed by a random vector $\varepsilon$ whose distribution may depend on both $x$ and $\theta$. The most prominent functional relationship is the general linear model setting $Y(x)=f(x)^{\mathrm{T}}\theta+\varepsilon$ where the response $\eta(x, \theta)=f(x)^{\mathrm{T}}\theta=\Sigma_{i=1}^{p} f_i(x)\theta_i$ is a linear function in the parameters $\theta$ with known regression functions $f=(f_1, \ldots, f_p)^{\mathrm{T}}$. These models cover both regression and analysis of variance models where for the latter some dummy coding is required. Also more complicated models are included like analysis of covariance models in which both discrete and continuous exogenous quantities may be present.

The size $n$ of an experiment is the number of outcomes $Y_1, \ldots, Y_n$ to be observed. The design of an experiment is the corresponding setting of the exogenous quantities $x_1, \ldots, x_n$. Replications are allowed such that these settings are not necessarily all distinct. Usually, homoscedasticity is assumed for the error term $\varepsilon$ in a linear model, $Var(\varepsilon)=\sigma^2$. Then the performance of the statistical inference is measured in terms of the information matrix $M(x_1, \ldots, x_n)=\Sigma_{i=1}^{n} f(x_i)f(x_i)^{\mathrm{T}}$ or, more intuitively, in terms of its inverse

$M(x_1, \ldots, x_n)^{-1}$ which is proportional to the covariance matrix of the least squares estimator $\hat{\theta}$, $Cov(\hat{\theta}) = \sigma^2 M(x_1, \ldots, x_n)^{-1}$.

Optimization of an experiment can be divided into two separate steps: First, optimize the information matrix with respect to a properly chosen criterion for a given size of the experiment. In this context the most popular criterion is the so-called *D*-criterion which aims at minimizing the determinant of the covariance matrix, $\det M(x_1, \ldots, x_n)^{-1}$, a quantity which is often called erroneously the generalized variance. This is equivalent to maximizing the determinant of the information matrix

$$\max_{x_1, \ldots, x_n} \det M(x_1, \ldots, x_n)$$

where the exogenous variables $x_1, \ldots, x_n$ may range independently over a given design region $\mathcal{X}$ of possible settings. For Gaussian errors the determinant of the covariance matrix is proportional to the volume of the confidence ellipsoid for the parameter vector $\theta$. Moreover, the popularity of the *D*-criterion arises from its computational ease and from the fact mentioned below that it is equivalent to the minimization of the prediction variance $f(x)^\top M(x_1, \ldots, x_n)^{-1} f(x)$ under certain regularity conditions. It is worthwhile noting that, in fact, the optimal settings and their corresponding proportions of replications do not vary much if the size of the experiment is changed. In a second step the size is determined in order to meet the needs of the experiment. For example, the size of the experiment will be influenced by the required precision of the estimates, by the power of a statistical test to be performed, but also by cost considerations. The recent monograph by Cox and Reid (2000) may serve well as an introductory text to the topic.

In agreement with the literature on optimal design theory we will focus on the first step of finding optimal settings for the exogenous variables. The first paper which was explicitly devoted to efficient designs was written by Smith (1918) before any general concepts had been developed. In the beginning of the twentieth century practical needs for optimal or efficient designs arose from agricultural experiments. At that time it was the merit of Fisher (1935) to define the basic concepts of experimental design: repeatability, blocking, and randomization. For analysis of variance settings which are typical for agricultural experiments optimization leads mostly to combinatorial problems (for a survey see Shah & Sinha, 1989).

For general settings Kiefer (1959) introduced the notion of generalized designs for which the proportions of the replications are detached from the sample size. According to this concept $\xi$ denotes a generalized design when it is a finitely supported normalized measure on the possible settings $x$. For a design $\xi$ that describes an experiment of size $n$ the quantities $\xi(x_i) = n_i/n$ denote the proportions of replications $n_i$ at the distinct settings $x_i$, $\Sigma\,\xi(x_i)=1$. The corresponding normalized information matrix is defined by

$$M(\xi)=\sum \xi(x_i)f(x_i)f(x_i)^{\mathsf{T}}.$$

Note that $M(\xi)=n^{-1}\,M(x_1,\ldots,x_n)$. If the requirement is dropped that $\xi(x_i)$ is a multiple of $1/n$ then the designs can be embedded into a convex optimization framework (see also Kiefer, 1974). Hence, standard methods of convex optimization can directly be applied to optimal design theory by taking directional derivatives leading to equivalence theorems based on the saddle-points for minimax solutions. For example, the celebrated Kiefer-Wolfowitz equivalence theorem (Kiefer & Wolfowitz, 1960) states that the $D$-optimality of a design $\xi^*$ is equivalent to the minimax optimality of $\xi^*$ with respect to the prediction variance on the design region, i.e.

$$\det M(\xi^*)=\max_\xi \det M(\xi) \quad \text{if and only if}$$

$$\max_\xi f(x)^{\mathsf{T}} M(\xi^*)^{-1}f(x)=\min_\xi \max_\xi f(x)^{\mathsf{T}} M(\xi)^{-1}f(x).$$

For every design its efficiency is the quality of its performance compared to the benchmark of the optimal design, i.e. the quantity $1/\mathit{efficiency}(\xi)\cdot 100\%$ gives the factor for the number of observations required when the design is used to obtain the same amount of information as contained in the optimal design. Accordingly, for the $D$-criterion, the $D$-efficiency is defined as $\mathrm{eff}_D(\xi)=(\det M(\xi)/\det M(\xi^*))^{1/p}$ where $\xi^*$ denotes the $D$-optimal design. Based on the properties of directional derivatives, suitable efficiency bounds can be computed for the performance of arbitrary designs (see Dette, 1996).

In the sequel various concepts of statistics were applied to evolve solutions of the optimization problem like invariance or equivariance with respect to certain natural transformations of the design region (Giovagnoli, Pukelsheim & Wynn, 1987) which, in a way, generalizes the concept of randomization. These topics are treated in full generality in Pukelsheim (1993).

Due to the variety of possible structural dependencies in the general linear model a vast amount of approaches has been developed for solving particular problems. In the present setting special interest lies in multi-factor models (Schwabe, 1996) with a reasonable number of exogenous variables and in the peculiarities of paired comparisons (van Berkum, 1987a, b, 1989).

It should be noted that often, as in paired comparisons, the linear model only serves as a rough approximation to some non-linear relationship. If the non-linearity can be explicitly specified, large sample behavior is available for the performance of a design. Denote by $f_\theta(x)=(f_{\theta,1}(x),\ldots,f_{\theta,p}(x))^{\mathsf{T}}$ the vector of locally linearized regression functions, if $\theta$ is the true value of the parameter. Here the local regression functions $f_{\theta,i} = (\partial/\partial\theta_i)\,\eta(x, \theta)$ are the partial derivatives of the response function $\eta(x, \theta)$ with respect to the components of the parameter $\theta$. Then $M_\theta(\xi) = \Sigma\, \xi(x) f_\theta(x) f_\theta(x)^{\mathrm{T}}$ is the asymptotic information matrix of the design $\xi$ at $\theta$. Unlike in linear models the quality depends on the unknown parameters. Hence, only locally optimal designs can be generated or such which are related to a Bayesian or minimax loss function (for surveys see Chaloner & Verdinelli, 1995; Fedorov & Hackl, 1997). For generalized linear models a promising approach has been proposed by Ford, Torsney and Wu (1992) which is based on a canonical transformation.

In the situation of generalized linear models the response $\eta$ is linked by a mapping, say $g$, to a linear regression approach, i.e. $\eta(x, \theta) = g(f(x)^{\mathrm{T}}\theta)$ (see McCullagh & Nelder, 1989). The inverse $g^{-1}$ of this mapping is traditionally called the link function of the generalized linear model. By the chain rule the linearized regression functions equal $f_\theta(x) = g'(f(x)^{\mathrm{T}}\theta) f(x)$. In the particular case $\theta = 0$, the linearized regression functions $f_\theta = f_0$ and the inherent linear relationship $f$ coincide up to a multiplicative constant $g'(0)$ independent of $x$. Thus the corresponding information matrices are related by $M_0(\xi) = g'(0)^2 M(\xi)$ and the optimization with respect to the generalized linear model reduces to the optimization with respect to the corresponding linear relationship. Hence, the linear model $Y(x) = f(x)^{\mathrm{T}}\theta + \varepsilon$ may serve well as a surrogate for the corresponding generalized linear model $Y(x) = g(f(x)^{\mathrm{T}}\theta) + \varepsilon$ when the hypothesis $\theta = 0$ is to be tested.

As an additional complication the error terms $\varepsilon$ are commonly heteroscedastic in generalized linear models with variance function $\sigma^2(x, \theta) = h(f(x)^{\mathrm{T}}\theta)$. For example, in case of binary response $\eta(x, \theta)$ denotes the probability of success and $\sigma^2(x, \theta) = \eta(x, \theta)(1 - \eta(x, \theta))$, i.e. $h = g(1 - g)$. The variance function has an influence on the performance and must be included in the (asymptotic) information matrix

$$M_\theta(\xi) = \sum \xi(x)\, \sigma^{-2}(x, \theta) f_\theta(x) f_\theta(x)^{\mathrm{T}}.$$

However, for $\theta = 0$ the variance $\sigma^2(x, \theta) = h(0)$ is independent of $x$ and the optimization of the information matrix $M_0(\xi) = g'(0)^2 h(0)^{-1} M(\xi)$ coincides again with the linear case.

In the present setting of choice models one is mainly concerned with multinomial logistic models where $g^{-1}(z) = \ln z - \ln(1 - z)$ is the logit link function. In particular, for paired comparisons the response $\eta(x, \theta)$ is given by the generalized linear model

$$\eta((a_1, a_2), \theta) = g(f(a_1)^{\mathrm{T}}\theta - f(a_2)^{\mathrm{T}}\theta) = g((f(a_1) - f(a_2))^{\mathrm{T}}\theta)$$

where $x = (a_1, a_2)$ is the pair of alternatives presented. The condition $\theta = 0$ is related to the situation of no preference for either of the alternatives, i.e. $\eta(x, 0) = 1/2$ is independent of $x$. Moreover, $\sigma^2(x, 0) = 1/4 = g'(0)$ and the (asymptotic) information matrix becomes

$$M_0(\xi) = \frac{1}{4} \sum \xi(a_1, a_2)(f(a_1) - f(a_2))(f(a_1) - f(a_2))^{\mathrm{T}}.$$

More generally, for the presentation of larger choice sets $(a_1, \ldots, a_m)$ with $m$ alternatives, say, the (asymptotic) information matrix can be derived as

$$M_0(\xi) = \sum \xi(a_1, \ldots, a_m) \times m^{-1} \sum_i \left[ f(a_i) - m^{-1} \sum_j f(a_j) \right] \left[ f(a_i - m^{-1} \sum_j f(a_j) \right]^{\mathrm{T}} \quad (1)$$

(see e.g. Bunch, Louviere & Anderson, 1996).

If no explicit solution of the optimization problem is available, algorithms can be used like the Fedorov-Wynn algorithm (Fedorov, 1972; Wynn, 1970) which are based on a steepest descent approach for the directional derivatives. Some of these algorithms are implemented in the OPTEX module of the SAS statistical software package. For a survey on the whole scope of experimental design we refer to Atkinson (1988, 1996) and Ghosh and Rao (1996).

## 3. DESIGNS FOR CONJOINT AND DISCRETE CHOICE MODELS

When developing an experimental design for a particular conjoint or discrete choice model the investigator has to consider a number of issues. First, the attribute levels for the profiles to be presented have to be chosen in such a way that the corresponding model parameters are estimable. This amounts to ensuring that the model's design matrix is of full rank. Second, it has to be decided whether levels of each attribute are presented for all profiles or if some attributes are left unspecified and profiles are only constructed from a subset of attributes. For example, with a large number of attributes to be presented, the evaluation of, or choice among, profiles that are made up of all attributes, so-called full profiles, is much more demanding than the respective task for profiles that are only described by some of the attributes. Third, the number of profiles to be presented in a single evaluation or choice trial has to be settled. Here, large sets of profiles might cause the respondents to focus on only a subset of the attributes or to employ some other kind of simplifying strategy to arrive at their choices or evaluations.

These considerations are in line with the classification system for choice experiments* proposed by Green (1974). He suggests a number of strategies for dealing with the issues outlined. In particular, for linear main-effects-only models he proposes the use of orthogonal arrays which can be constructed as regular fractions of full factorial designs when the model is symmetric, i.e. when the number of levels is equal for all attributes. To illustrate, we consider a slight modification of Green's original $4 \times 3 \times 2^7$ airline example where trans-Atlantic flights are characterized by nine attributes with two, three, or four levels. Instead of employing different numbers of levels we will only use two levels for each attribute, i.e. we consider a $2^9$ model. A symmetrical orthogonal array for this model is shown in Table 1.

For asymmetric models, orthogonal arrays can be obtained from regular fractions of full factorial designs by collapsing certain columns (Addelman, 1962). If the investigator wants to include selected interactions in the model equation fractional factorial designs can be used.

As strategies for dealing with the second and third issues above Green (1974) proposed two different two-stage design approaches using balanced incomplete block (BIB) designs and partially incomplete block (PBIB) designs, respectively (for exact definitions of these designs, see Green, 1974; Raghavarao, 1971). When the investigator has decided to use only profiles described on four of the nine attributes the two-stage approach proceeds as follows: First, a BIB design is constructed to assign sets of four attributes to profiles and second, a small design, e.g. an orthogonal array if estimation of main effects suffices, with four-component profiles drawn from the $2^9$ full factorial plan is chosen. The set of profiles then consists of

number of blocks in the BIB design $\times$ number of rows in the second design

profiles. As Green (1974) demonstrated for the example a BIB design with eighteen blocks and four-component orthogonal arrays with eight rows exist so that the profile set comprises $18 \times 8 = 144$ profiles in comparison to

* It should be noted that the choice models considered by Green (1974) are actually linear models. That is, they are conjoint analysis models in the terminology used in the present paper.

***Table 1.*** Orthogonal Array for a $2^9$ Model.

| | Attributes and Levels | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Profile | A | B | C | D | E | F | G | H | I |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| 3 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| 4 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| 5 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| 6 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 |
| 7 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 8 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 9 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 |
| 10 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 11 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| 12 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 13 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 14 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| 15 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| 16 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |

$126 \times 8 = 1008$ profiles which would be obtained if an orthogonal array with eight runs was used for each of the 126 attribute combinations involving only four attributes.

The two-stage approach for constructing designs where in every evaluation trial only a subset of the profiles is presented relies on PBIB designs. Again, for purposes of illustration we consider the example. If the investigator wants to present pairs of profiles for evaluation then he chooses a subset of the profiles in the first step from the $2^9$ full factorial design, e.g. by means of an orthogonal fraction. Here, we assume that the 16 profiles in Table 1 have been chosen. In the second step these profiles are arranged in pairs according to a PBIB design. The 16 pairs derived from the PBIB design with block size $k = 2$ given by Green (1974) are shown in Table 2.

By using this design the number of paired comparisons for the 16 profiles can be reduced from 120 to 48 as compared to the round robin design in which each profile is paired once with every other profile. With different values of the block size $k$ choice designs with sets of size $k$ can be generated, e.g. for $k = 3$ we obtain triples of profiles.

In sum, Green's proposed strategies are effective for constructing parsimonious designs. Moreover, the principles that underly his approach, namely: (a) the

*Table 2.* Pairs Derived from PBIB Design.

| Pairs | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1, 2 | 5, 6 | 9, 10 | 13, 14 | 1, 5 | 2, 6 | 3, 7 | 4, 8 |
| 1, 3 | 5, 7 | 9, 11 | 13, 15 | 1, 9 | 2, 10 | 3, 11 | 4, 12 |
| 1, 4 | 5, 8 | 9, 12 | 13, 16 | 1, 13 | 2, 14 | 3, 15 | 4, 16 |
| 2, 3 | 6, 7 | 10, 11 | 14, 15 | 5, 9 | 6, 10 | 7, 11 | 8, 12 |
| 2, 4 | 6, 8 | 10, 12 | 14, 16 | 5, 13 | 6, 14 | 7, 15 | 8, 16 |
| 3, 4 | 7, 8 | 11, 12 | 15, 16 | 9, 13 | 10, 14 | 11, 15 | 12, 16 |

reduction of the set of alternatives by choice of a suitable subset drawn from a full factorial plan; and (b) the combination of these profiles in sets for evaluation by use of an experimental design that neglects the factorial structure of the profiles, have permeated the literature on experimental designs for conjoint analysis and discrete choice models in marketing to a large extent. For example, Louviere and Woodworth (1983), Batsell and Louviere (1991), as well as Bunch et al. (1996) have adhered to these principles. In general, not much is known about the efficiency of these designs (Carson et al., 1994, p. 361). In Section 4 we will demonstrate for Green's paired comparison design that the principles outlined above may produce designs which are far from optimal in terms of efficiency.

Some authors have argued in favor of shifting emphasis from classical design principles such as orthogonality and balance to design efficiency. Kuhfeld, Tobias and Garratt (1994) recommended the use of the algorithms mentioned at the end of Section 2 and showed for a number of conjoint and discrete choice models how efficient designs can be constructed with the SAS procedure OPTEX. For the latter models they assumed that the parameter vector in the model is equal to zero so that the information matrix for the multinomial model coincides with the one of the general linear model up to constant factor. In a similar vein for conjoint analysis experiments Steckel, DeSarbo and Mahajan (1991) presented a combinatorial optimization procedure for maximizing the determinant of the information matrix in situations where there exist natural correlations among the attributes, e.g. maximum speed and mileage when cars are of concern. In this situation the approach of Kuhfeld et al. (1994) is applicable as well.

The design problem for correlated attributes has also been treated by Louviere and Woodworth (1988). They proposed to construct choice sets by first obtaining ratings for a set of profiles on the correlated attributes. These ratings are then arranged in matrix form where each row of the matrix corresponds to one of the profiles. Subsequently, a second matrix is constructed by adding an orthogonal matrix to the initial one. Every row in this second matrix represents the second profile in a choice set. Designs with larger choice sets can be generated by continuing this procedure. The efficiency of such designs as compared to the approaches of Kuhfeld et al. (1994) and Steckel et al. (1991) is not known.

Another issue that has received attention is the construction of designs which enable the estimation of attribute or availability cross effects on the choice probabilities (Louviere & Woodworth, 1983; Anderson & Wiley, 1992). Lazari and Anderson (1994) presented a model where both types of cross effects can be estimated simultaneously for situations where $m$ brands are described by a single attribute. Moreover, in their model as well as in the model considered by Louviere and Woodworth (1983) violations of the independence of irrelevant alternatives assumption of the multinomial logistic model can be tested. To generate designs, Lazari and Anderson applied the technique of Louviere and Woodworth (1983) for constructing choice sets. As an illustration, we consider the situation where the attribute that characterizes the $m$ brands has $S - 1$ levels. An orthogonal main effects plan is drawn from the $S^m$ full factorial design and the levels are coded consecutively $0, \ldots, S - 1$. Every row of this design then represents a choice set. The level 0 in the $i$th position of a row indicates that the $i$th brand is not present in the choice set whereas a larger number indicates that the brand is present in the set with the attribute adjusted to the respective level.

The approaches considered so far all assumed that the vector of parameters $\theta$ equals zero or in other words that the choice probabilities for the alternatives in a choice set are all the same. As was detailed in Section 2, under this assumption, the optimization problem for the determinant of the information matrix in the discrete choice model remains the same as the corresponding problem in a linear model framework. However, Huber and Zwerina (1996) argued that in most practical marketing research situations some kind of prior knowledge on the parameters is available, for example, when a pretest of a survey is conducted. Therefore, they proposed a method for designing multi-attribute choice experiments that incorporates the use of such prior information. The strategy adopted proceeds by first generating a so-called seed design by use of either an orthogonal array or the SAS procedure OPTEX. Each row in this design represents the first profile of a particular choice set, i.e. the first row in the design matrix for that set. The remaining profiles are constructed from the first one by subsequently incrementing the attributes' levels columnwise and cycling back to one when a level's value exceeds the number of levels of that attribute. Every choice set generated in this way is an instance of what is called a cyclic design (see e.g. John & Williams, 1995) in the statistical literature on experimental designs. In a subsequent optimization step which utilizes the available prior information on the vector $\theta$ the efficiency of the complete choice design formed by concatenating the designs for the single sets is further improved. This improvement is achieved by using one out of two techniques, swapping and relabeling, of what Huber and Zwerina have called utility balancing. With swapping, in every choice set transpositions (swaps) of pairs of levels are examined for every attribute, i.e. every column, in turn. Every swap is tested and the information matrix for the multinomial logistic model in Eq. 1 where the prior information is substituted for $\theta$ is computed. If the swap improves the determinant of the information matrix it is performed. With the relabeling technique permutations of the attribute levels that increase efficiency are investigated. If, for example, the assignment of a 1 to the first level of a three level attribute, a 3 to the second level, and a 2 to the third level in all choice sets instead of a consecutive numbering increases the determinant of the information matrix this relabeling is performed. Huber and Zwerina demonstrated for a number of models ranging from simple to complex that substantial efficiency gains can be accomplished. Moreover, these gains are relatively robust with respect to fallible prior information. However, the nature of the utility balancing principle underlying the Huber and Zwerina approach is essentially heuristic. No analytical results are available whether the proposed procedure reaches the global optimum or how close it comes.

Recently, Sándor and Wedel (in press) have amplified the idea of utilizing prior information in the design construction process for the multinomial logit model. They extended the results by Huber and Zwerina (1996) in three main directions. First, they apply Bayesian design techniques (Chaloner & Verdinelli, 1995) and replace the unknown parameters in the $D$-criterion by prior distributions. Second, they determine the prior distribution by eliciting prior information from respondents based on the methods developed by van Lenthe (1993). Third, they develop further the heuristic algorithms relabeling and swapping to an algorithm called cycling that searches in a larger design space and hence potentially yields designs with higher statistical efficiency. Based on Monte Carlo studies as well as an empirical illustration Sándor and Wedel (in press) provide evidence that the Bayesian approach produces designs that are more efficient than those generated according to Huber and Zwerina (1996).

The idea of utility balancing is also part of the design heuristic for paired comparisons implemented in the most popular software ACA (1994) for adaptive conjoint analysis. Adaptive conjoint analysis is a so-called hybrid conjoint analysis technique because it combines compositional and decompositional measures. Compositional part-worth utilities for attribute levels are obtained through direct preference assessments of the levels and attribute importance ratings. In the decompositional phase of an ACA interview, respondents have to perform a number of graded paired comparisons and to state their relative preferences for one or the other profile in a pair. Usually, the profiles are described by only a subset of the attributes that varies from pair to pair. Pairs are chosen subsequently according to an adaptive algorithm. When choosing the next pair to be presented, this algorithm pursues the following objectives: First, attributes are combined that have occurred together fewest before. Second, levels of these attributes are selected by application of a similar logic. Third, levels are arranged in profiles in order to maximize utility balance. To achieve maximum utility balance, the vector of parameters is estimated after each paired comparison and the utilities of the profiles in a pair are computed by summing the respective parameter estimates for every possible arrangement of the chosen levels into pairs. The pair for which the profiles are most equal in utility then is actually presented. From a statistical point of view this adaptive strategy is dispensable because the information matrix does not depend on the true parameter vector. In Section 5 we will report results from an empirical investigation in which the adaptive design heuristic of ACA is compared to an optimal design that has been constructed according to the principles which are described in the next section.

# 4. SOME OPTIMAL PAIRED COMPARISON DESIGNS

Recently, a new approach for constructing multi-factor paired comparison designs has been proposed by Graßhoff, Großmann, Holling and Schwabe (2000). They proved the *D*-optimality of a certain type of designs in a linear model setting where the interest lies in the estimation of main effects. All attributes are assumed discrete with the same number of levels. These designs are also optimal for the corresponding discrete choice models under the assumption $\theta = 0$. Furthermore, Graßhoff et al. proved the optimality of their designs for situations where the so-called profile strength, i.e. the number of attributes which are allowed to vary in every comparison, is restricted. The construction of the optimal designs relies on general principles for multi-factor models (see Schwabe, 1996) as well as on Hadamard matrices (see e.g. Raghavarao, 1971).

For the airline example with nine attributes each at two levels considered in the previous section, the construction of a *D*-optimal paired comparison design is particularly simple. Instead of first choosing a subset of profiles from the $2^9$

***Table 3.*** Design Matrix for Green's Paired Comparison Design.

| | Attributes | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Number of Pair | A | B | C | D | E | F | G | H | I |
| 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| 2 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 |
| 3 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 4 | 0 | 0 | 0 | 1 | –1 | 0 | 1 | –1 | 0 |
| 5 | 0 | 0 | 0 | 1 | 0 | –1 | 1 | 0 | –1 |
| 6 | 0 | 0 | 0 | 0 | 1 | –1 | 0 | 1 | –1 |
| 7 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | –1 | –1 |
| 8 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | –1 |
| 9 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | –1 | 0 |
| 10 | 0 | 0 | 0 | 1 | –1 | 0 | 1 | 1 | 0 |
| 11 | 0 | 0 | 0 | 1 | 0 | –1 | 1 | 0 | 1 |
| 12 | 0 | 0 | 0 | 0 | 1 | –1 | 0 | –1 | 1 |
| 13 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | –1 |
| 14 | 0 | 0 | 0 | 1 | 0 | 1 | –1 | 0 | –1 |
| 15 | 0 | 0 | 0 | 1 | 1 | 0 | –1 | 1 | 0 |
| 16 | 0 | 0 | 0 | 1 | –1 | 0 | –1 | –1 | 0 |
| 17 | 0 | 0 | 0 | 1 | 0 | –1 | –1 | 0 | 1 |
| 18 | 0 | 0 | 0 | 0 | 1 | –1 | 0 | 1 | 1 |
| 19 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | –1 | 1 |
| 20 | 0 | 0 | 0 | 1 | 0 | 1 | –1 | 0 | 1 |
| 21 | 0 | 0 | 0 | 1 | 1 | 0 | –1 | –1 | 0 |
| 22 | 0 | 0 | 0 | 1 | –1 | 0 | –1 | 1 | 0 |
| 23 | 0 | 0 | 0 | 1 | 0 | –1 | –1 | 0 | –1 |
| 24 | 0 | 0 | 0 | 0 | 1 | –1 | 0 | –1 | –1 |
| 25 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 26 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
| 27 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 28 | 1 | –1 | 0 | 0 | 0 | 0 | 1 | –1 | 0 |
| 29 | 1 | 0 | –1 | 0 | 0 | 0 | 1 | 0 | –1 |
| 30 | 0 | 1 | –1 | 0 | 0 | 0 | 0 | 1 | –1 |
| 31 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | –1 | –1 |
| 32 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | –1 |
| 33 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | –1 | 0 |
| 34 | 1 | –1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 35 | 1 | 0 | –1 | 0 | 0 | 0 | 1 | 0 | 1 |
| 36 | 0 | 1 | –1 | 0 | 0 | 0 | 0 | –1 | 1 |
| 37 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | –1 |
| 38 | 1 | 0 | 1 | 0 | 0 | 0 | –1 | 0 | –1 |
| 39 | 1 | 1 | 0 | 0 | 0 | 0 | –1 | 1 | 0 |
| 40 | 1 | –1 | 0 | 0 | 0 | 0 | –1 | –1 | 0 |
| 41 | 1 | 0 | –1 | 0 | 0 | 0 | –1 | 0 | 1 |
| 42 | 0 | 1 | –1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 43 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | –1 | 1 |
| 44 | 1 | 0 | 1 | 0 | 0 | 0 | –1 | 0 | 1 |
| 45 | 1 | 1 | 0 | 0 | 0 | 0 | –1 | –1 | 0 |
| 46 | 1 | –1 | 0 | 0 | 0 | 0 | –1 | 1 | 0 |
| 47 | 1 | 0 | –1 | 0 | 0 | 0 | –1 | 0 | –1 |
| 48 | 0 | 1 | –1 | 0 | 0 | 0 | 0 | –1 | –1 |

***Table 4.*** Optimal Paired Comparison Design.

| Number of Pair | Attributes | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I |
| 1 | 1 | 1 | 1 | 1 | −1 | 1 | 1 | 1 | 1 |
| 2 | −1 | 1 | −1 | −1 | −1 | 1 | −1 | 1 | −1 |
| 3 | −1 | −1 | −1 | 1 | 1 | 1 | −1 | 1 | 1 |
| 4 | 1 | −1 | 1 | 1 | −1 | −1 | −1 | 1 | −1 |
| 5 | 1 | −1 | −1 | 1 | 1 | 1 | 1 | 1 | −1 |
| 6 | −1 | −1 | 1 | 1 | −1 | 1 | −1 | −1 | −1 |
| 7 | −1 | 1 | 1 | 1 | 1 | 1 | 1 | −1 | −1 |
| 8 | 1 | 1 | −1 | 1 | −1 | 1 | −1 | −1 | 1 |
| 9 | −1 | 1 | 1 | 1 | 1 | −1 | −1 | 1 | 1 |
| 10 | −1 | 1 | −1 | 1 | −1 | −1 | 1 | 1 | −1 |
| 11 | 1 | 1 | −1 | 1 | 1 | −1 | −1 | −1 | −1 |
| 12 | −1 | −1 | −1 | 1 | −1 | −1 | 1 | −1 | 1 |
| 13 | 1 | 1 | 1 | −1 | 1 | −1 | −1 | −1 | −1 |
| 14 | −1 | 1 | −1 | 1 | 1 | −1 | 1 | −1 | 1 |
| 15 | −1 | −1 | −1 | −1 | −1 | −1 | 1 | −1 | −1 |
| 16 | 1 | −1 | 1 | −1 | 1 | 1 | 1 | −1 | 1 |
| 17 | 1 | −1 | −1 | −1 | −1 | −1 | −1 | −1 | 1 |
| 18 | −1 | −1 | 1 | −1 | 1 | −1 | 1 | 1 | 1 |
| 19 | −1 | 1 | 1 | −1 | −1 | −1 | −1 | 1 | 1 |
| 20 | 1 | 1 | −1 | −1 | 1 | −1 | 1 | 1 | −1 |
| 21 | −1 | 1 | 1 | −1 | −1 | 1 | 1 | −1 | −1 |
| 22 | −1 | 1 | −1 | −1 | 1 | 1 | −1 | −1 | 1 |
| 23 | 1 | 1 | −1 | −1 | −1 | 1 | 1 | 1 | 1 |
| 24 | −1 | −1 | −1 | −1 | 1 | 1 | −1 | 1 | −1 |

full factorial plan and then assigning these profiles to pairs according to Table 2 which yields the design matrix shown in Table 3, the profiles and pairs are constructed simultaneously.

This is done by choosing nine columns from a suitable Hadamard matrix. Table 4 shows the design matrix obtained in this way from the Hadamard matrix $H_{24}$ of order 24. Every row in the table represents a paired comparison. A one in the $i$th column indicates that the $i$th attribute of the first profile in a pair is at the high level and the second profile is at the low level of that attribute. Similarly, a minus one indicates that the first profile is at the low and the second profile is at the high level.

The determinant of the normalized information matrix of the design $\xi$ constructed according to Green (1974) in Table 3 equals $\det M(\xi) = 0.0004$

***Table 5.*** Optimal Paired Comparison Design with Profile Strength Four.

| Number of Pair | Factors | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I |
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | –1 | 1 | –1 | 0 | 0 | 0 | 0 | 0 |
| 3 | 1 | 1 | –1 | –1 | 0 | 0 | 0 | 0 | 0 |
| 4 | 1 | –1 | –1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 6 | 0 | 1 | –1 | 1 | –1 | 0 | 0 | 0 | 0 |
| 7 | 0 | 1 | 1 | –1 | –1 | 0 | 0 | 0 | 0 |
| 8 | 0 | 1 | –1 | –1 | 1 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 10 | 0 | 0 | 1 | –1 | 1 | –1 | 0 | 0 | 0 |
| 11 | 0 | 0 | 1 | 1 | –1 | –1 | 0 | 0 | 0 |
| 12 | 0 | 0 | 1 | –1 | –1 | 1 | 0 | 0 | 0 |
| 13 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| 14 | 0 | 0 | 0 | 1 | –1 | 1 | –1 | 0 | 0 |
| 15 | 0 | 0 | 0 | 1 | 1 | –1 | –1 | 0 | 0 |
| 16 | 0 | 0 | 0 | 1 | –1 | –1 | 1 | 0 | 0 |
| 17 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| 18 | 0 | 0 | 0 | 0 | 1 | –1 | 1 | –1 | 0 |
| 19 | 0 | 0 | 0 | 0 | 1 | 1 | –1 | –1 | 0 |
| 20 | 0 | 0 | 0 | 0 | 1 | –1 | –1 | 1 | 0 |
| 21 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| 22 | 0 | 0 | 0 | 0 | 0 | 1 | –1 | 1 | –1 |
| 23 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | –1 | –1 |
| 24 | 0 | 0 | 0 | 0 | 0 | 1 | –1 | –1 | 1 |
| 25 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 26 | –1 | 0 | 0 | 0 | 0 | 0 | 1 | –1 | 1 |
| 27 | –1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | –1 |
| 28 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | –1 | –1 |
| 29 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 30 | 1 | –1 | 0 | 0 | 0 | 0 | 0 | 1 | –1 |
| 31 | –1 | –1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 32 | –1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | –1 |
| 33 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 34 | –1 | 1 | –1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 35 | 1 | –1 | –1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 36 | –1 | –1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |

compared to the value of $\det M(\xi^*) = 1$ for the determinant of the normalized information matrix of the optimal design $\xi^*$ in Table 4. As a result, a $D$-efficiency of $\mathrm{eff}_D(\xi) = (\det M(\xi))^{1/9}/(\det M(\xi^*))^{1/9} = 0.42$ obtains. In other words the model parameters can be estimated more precisely with the optimal design using 24 pairs than with the design using twice as many comparisons.

A closer look at Table 3 reveals that the profiles in every pair do actually vary on only four of the nine attributes, i.e. the design employs comparisons with a profile strength of four. A $D$-optimal design for this profile strength which uses only 36 instead of the 48 paired comparisons in Table 3 can be constructed by arranging the columns of a Hadamard matrix of order four in a cyclic manner. The resulting design $\xi'$ is shown in Table 5. Here, the entries 1 and $-1$ are interpreted in the same way as for the optimal design in Table 4. The additional zeros in every row indicate that the corresponding profiles in a pair do not differ with respect to the respective attributes. For example, a zero in the $i$th position of a row means that both profiles in a pair are characterized by the same level of the $i$th attribute.

In the restricted class of designs with a profile strength of four the design $\xi$ in Table 3 performs much better as compared to the class of designs where the profiles in a pair are permitted to differ on all attributes. This is reflected by the $D$-efficiency of $\mathrm{eff}_D(\xi) = (\det M(\xi))^{1/9}/(\det M(\xi'))^{1/9} = 0.94$. In sum, large efficiency gains may be accomplished with an optimal design when the researcher intends to use a high profile strength. In this situation the number of paired comparisons necessary to achieve a certain precision of the parameter estimates can be substantially reduced. However, with a low profile strength efficiency gains may only be marginal.

## 5. AN EMPIRICAL COMPARISON OF DESIGNS

From a statistical viewpoint $D$-optimal designs outperform less efficient designs that use the same number of observations. As has been noted before (see e.g. Bunch et al., 1996) statistical efficiency is only one of many criteria for judging the quality of an experimental design. One of the non-statistical criteria is the cognitive difficulty of an evaluation or choice task that has to be taken into account. For example, the information processing requirements of full profile designs are usually assumed to be too demanding when there are many attributes (but for some contradicting evidence, see Pullman, Dodson & Moore, 1999). Hence, the question remains whether the statistical superiority of optimal designs with respect to efficiency translates into empirical benefits.

In order to investigate this issue we chose the adaptive design heuristic of ACA which was described in Section 3 as a benchmark. To the best of our knowledge, up to now this procedure has never been compared empirically to $D$-optimal designs. Furthermore, we adopted the principal-agent paradigm (see e.g. West, 1996, and the references therein) for our experiment which has been employed successfully in recent research on conjoint analysis (Teichert, 2000; Huber, Ariely & Fischer, 2001). According to this paradigm the participants (agents) have to perform some action on behalf of a principal. Before being exposed to the task they therefore have to learn the criteria the principal would use when performing the action by herself/himself.

### *5.1. Design of the Experiment*

The participants in our experiment had to act as notebook purchasers for a company. Notebooks were described by six attributes with two levels each. The attributes and their levels are given in Table 6. Additionally, the table contains the true part-worth values of the levels the participants had to learn. These part-worth utilities represent a monetary surplus of how much Deutsche Mark (DM) a notebook with the better level of an attribute is valued higher by the principal than a notebook with the alternative level of that attribute given that both notebooks are identical with respect to the other attributes. For example, the principal would be willing to pay 500 DM more for a notebook with a 750 MHz instead of a 500 MHz processor.

The participants were told to purchase notebooks at an online-retailer's. Each participant had to identify the more valuable notebook and to estimate the surplus value on a continuous DM-scale subsequently.

The empirical study took place in two blocks of 45 minutes on two consecutive days. Fifteen undergraduate students were recruited as participants.

***Table 6.*** Attributes, Levels and True Part-Worth Utilities in the Empirical Study.

| Attribute | Level | Part-worth utility |
|---|---|---|
| Processor | 500 MHz | 0 |
| | 750 MHz | 500 |
| Screen size | 12″ Screen | 0 |
| | 14″ Screen | 400 |
| Hard disk | 10 GB Hard disk | 0 |
| | 20 GB Hard disk | 300 |
| Memory | 64 MB Ram | 0 |
| | 128 MB Ram | 250 |
| CD-Rom/DVD | CD-Rom | 0 |
| | DVD | 100 |
| Modem | no modem | 0 |
| | modem | 50 |

In order to provide an adequate consolidation of the prescribed utility structure, the survey was split up in such a way that the training of the part-worth utilities extended over two days. The complete design of the study is summarized in Table 7.

On the first day the scenario was explained to the participants. As a next step, a training phase for the consolidation of the part-worth utilities followed in which the participants worked on different exercises. At the end of Block A the learning success was tested with a paper-pencil test. On the second day the training phase continued. After the last exercise had been completed the success of the training was tested again. After a short break, the data acquisition which consisted of 48 paired comparisons followed.

All exercises and paired comparisons were conducted computerized with the software ALASCA (Holling, Jütting & Großmann, 2000). This program allows to administer paired comparisons according to an adaptive ACA-like as well as a *D*-optimal design. The participants were already acquainted with the handling of this program. A reward was announced for the "best" purchaser to promote a high motivation. The profiles in the paired comparisons were described by three attributes and a DM-scale was used for responses. All tasks were presented under time limitations in order to prevent an exact calculation of

***Table 7.*** Design of the Empirical Study.

| Block | Phase | Task |
|---|---|---|
| A | Introduction to the scenario | – |
| | Learning phase | Exercise 1: 4 utility evaluations of full profiles<br>Exercise 2: 10 paired comparisons with profiles described by two attributes |
| | | Learning test |
| B | Learning phase | Exercise 3: 4 utility evaluations of full profiles<br>Exercise 4: Rank ordering of 10 full profiles<br>Exercise 5: 10 paired comparisons with profiles described by two attributes |
| | | Learning test |
| | Data acquisition | 48 paired comparisons with profiles described by three attributes |

utility differences of the objects presented. The time limitation was fixed to 10 seconds for each paired comparison. The remaining processing time for each task was displayed in the upper half of the screen. Each participant responded to 24 paired comparisons according to a $D$-optimal design and to the same number of comparisons according to an adaptive design. Seven respondents first worked on the $D$-optimal pairs and on the adaptive ones thereafter and vice versa for the other eight persons.

For each of the 15 participants a vector of utilities based on the adaptive design and a vector of utilities based on the $D$-optimal design was estimated separately by multiple regression using differences of dummy coded attributes excluding an intercept. Thus, regression coefficients correspond to the surplus values.

### 5.2. *Results*

Part-worth utilities were learned by all participants quite well. The relative efficiencies of the 15 adaptive designs ranged from 0.84 to 0.96 with mean of 0.88. For each participant the following criteria were computed based on responses under the $D$-optimal as well as the adaptive design:

- root mean squared error $\sqrt{\frac{1}{6}\sum_{i=1}^{6}(\hat{\theta}_i - \theta_i)^2}$ (RMSE) of the part-worth estimates, where $\theta_i$ denotes the true part-worth utility of the second row level of the $i$th attribute in Table 6 and $\hat{\theta}_i$ the corresponding estimate,
- mean absolute difference $\frac{1}{6}\sum_{i=1}^{6}|\hat{\theta}_i - \theta_i|$ (MADP) between true and estimated part-worth values,
- mean absolute difference $\frac{1}{24}\sum_{i=1}^{24}|y_i - y_{i,\theta}|$ (MADR) between actual and true responses on the paired comparison task, where $y_i$ denotes the response of the participant and $y_{i,\theta}$ the response expected to be given by the principal on the $i$th comparison
- standard errors for the six estimated regression coefficients $\hat{\theta}_i$ (SE1 to SE6).

Table 8 reports means and standard deviations for these criteria as well as results of $t$-tests for dependent samples. In order to stabilize the variances the criterion values were log transformed prior to testing.

The $D$-optimal designs perform better with respect to every criterion. The mean absolute difference between actual and true responses amounts to 128.02 DM for the $D$-optimal designs and is 19% smaller in comparison to adaptive designs. Furthermore, the confidence intervals are considerably smaller. Summarizing the above results there is remarkable evidence that the theoretical advantages of $D$-optimal designs also manifest themselves empirically.

**Table 8.** Results of the Empirical Study.

| Criterion | D-optimal | | Adaptive | | | |
|---|---|---|---|---|---|---|
| | M | SD | M | SD | t | p |
| RMSE | 51.35 | 28.22 | 81.94 | 58.57 | 1.82 | 0.05 |
| MADP | 43.70 | 26.62 | 68.82 | 48.01 | 1.73 | 0.05 |
| MADR | 128.02 | 60.04 | 157.73 | 80.93 | 1.69 | 0.06 |
| SE1 | 50.10 | 24.80 | 71.11 | 34.66 | 3.19 | 0.03 |
| SE2 | 50.10 | 24.80 | 67.03 | 34.51 | 2.30 | 0.02 |
| SE3 | 50.10 | 24.80 | 72.40 | 38.39 | 2.78 | 0.01 |
| SE4 | 50.10 | 24.80 | 70.55 | 38.95 | 2.49 | 0.01 |
| SE5 | 50.10 | 24.80 | 73.14 | 35.52 | 3.26 | 0.00 |
| SE6 | 50.10 | 24.80 | 70.50 | 38.3 | 2.47 | 0.01 |

*Note:* $N = 15$, $p$-values according to one-sided test. For abbreviations of criteria, see text.

# 6. SUMMARY AND CONCLUSIONS

We presented a review of the statistical theory of optimum experimental designs and the approaches that have been proposed in the marketing literature for the design of conjoint analysis and discrete choice experiments. Drawing on recently developed optimal paired comparison designs we demonstrated that the well-known approach of Green (1974) to the design of multi-attribute choice experiments may yield to inefficient designs that can be substantially improved.

Furthermore, in an empirical study we compared the adaptive design heuristic employed in adaptive conjoint analysis and an optimal paired comparison design. The results showed that the optimal design performed better with respect to a variety of criteria than adaptive designs. This provides preliminary first empirical evidence for the superiority of optimal designs in the context of conjoint analysis.

Although our findings are limited with respect to the number of attribute levels used and the structure of the model, because only the estimation of main effects was considered, we conjecture that design approaches for discrete choice models that have followed the logic of Green's approach (e.g. Louviere & Woodworth, 1983) can be improved with respect to efficiency by explicitly recognizing the factorial structure of the profiles when choice sets are composed.

In typical applications of conjoint analysis or discrete choice models the attributes are usually taken to be discrete, i.e. only a finite number of levels is used for each attribute. This entails that for continuous attributes some of the infinite possible levels have to be chosen for inclusion in the model while all others have to be neglected. Modeling the influence of such continuous attributes on evaluations or choices by some kind of known functional relationship, e.g. linear, quadratic or logarithmic, seems to be attractive because fewer parameters have to be estimated. Moreover, these models are likely to yield more reliable parameter estimates. Therefore, future research should consider the design problem for conjoint analysis and discrete choice models that incorporate both discrete as well as continuous attributes. Finally, further empirical research is needed to assess the practical benefits that can be achieved by the implementation of efficient designs.

## ACKNOWLEDGMENTS

This work was supported in part by the *Deutsche Forschungsgemeinschaft* under grant Ho 1286/2-1. We thank an anonymous referee for drawing our attention to the work of Sándor and Wedel.

## REFERENCES

ACA system (Version 4.0) [Computer software] (1994). Evanston, IL: Sawtooth Software.

Addelman, S. (1962). Orthogonal main–effect plans for asymmetrical factorial experiments. *Technometrics*, *4*, 21–46.

Aigner, D. J. (1979). A brief introduction to the methodology of optimal experimental design. *Journal of Econometrics*, *11*, 7–26.

Anderson, D. A., & Wiley, J. B. (1992). Efficient choice set designs for estimating availability cross–effects models. *Marketing Letters*, *3*, 357–370.

Atkinson, A. C. (1988). Recent developments in the methods of optimum and related experimental designs. *International Statistical Review*, *56*, 99–115.

Atkinson, A. C. (1996). The usefulness of optimum experimental designs. *Journal of the Royal Statistical Society B*, *58*, 59–76.

Batsell, R. R., & Louviere, J. J. (1991). Experimental analysis of choice. *Marketing Letters*, *2*, 199–214.

Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. Cambridge, MA: MIT Press.

Ben-Akiva, M., McFadden, D., Abe, M., Böckenholt, U., Bolduc, D., Gopinath, D., Morikawa, T., Ramaswamy, V., Rao, V., Revelt, D., & Steinberg, D. (1997). Modeling methods for discrete choice analysis. *Marketing Letters*, *8*, 273–286.

Bunch, D. S., Louviere, J. J., & Anderson, D. (1996). A comparison of experimental design strategies for choice-based conjoint analysis with generic-attribute multinomial logit models. Unpublished manuscript, Graduate School of Management, University of California at Davis.

Carroll, J. D., & Green, P. E. (1995). Psychometric methods in marketing research: Part I, conjoint analysis. *Journal of Marketing Research, 32*, 385–391.

Carson, R. T., Louviere, J. J., Anderson, D. A., Arabie, P., Bunch, D. S., Hensher, D. A., Johnson, R. M., Kuhfeld, W. F., Steinberg, D., Swait, J., Timmermans, H., & Wiley, J. B. (1994). Experimental analysis of choice. *Marketing Letters, 5*, 351–368.

Chaloner, K., & Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science, 10*, 273–304.

Cox, D. R., & Reid, N. (2000). *The theory of the design of experiments*. Boca Raton, FL: Chapman & Hall/CRC.

Dette, H. (1996). Lower bounds for efficiencies with applications. In: E. Brunner & M. Denker (Eds), *Research Developments in Probability and Statistics. Festschrift in Honor of Madan L. Puri on the Occasion of his 65th Birthday* (pp. 111–124). Utrecht: VSP.

Fedorov, V. V. (1972). *Theory of optimal experiments*. New York: Academic Press.

Fedorov, V. V., & Hackl, P. (1997). Model-oriented design of experiments. *Lecture Notes in Statistics* (Vol. 24). New York: Springer.

Fisher, R. A. (1935). *The design of experiments*. London: Oliver and Boyd.

Ford, I., Torsney, B., & Wu, C. F. J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *Journal of the Royal Statistical Society B, 54*, 569–583.

Ghosh, S., & Rao, C. R. (Eds) (1996). *Design and analysis of experiments. Handbook of statistics* (Vol. 13). Amsterdam: Elsevier.

Giovagnoli, A., Pukelsheim, F., & Wynn, H. (1987). Group invariant orderings and experimental designs. *Journal of Statistical Planning and Inference, 17*, 111–135.

Graßhoff, U., Großmann, H., Holling, H., & Schwabe, R. (2000). *Optimal designs for paired comparisons in main effects analysis of variance models* (Preprint, A-16-2000). Fachbereich Mathematik und Informatik, Freie Universität Berlin.

Green, P. E. (1974). On the design of choice experiments involving multifactor alternatives. *Journal of Consumer Research, 1*, 61–68.

Green, P. E., & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research, 8*, 355–363.

Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research, 5*, 103–123.

Green, P. E., & Srinivasan, V. (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing, 54*, 3–19.

Holling, H., Jütting, A., & Großmann, H. (2000). ALASCA: Computergestützte Entscheidungs- und Nutzenanalyse [ALASCA: Computerized decision and utility analysis]. Göttingen: Hogrefe.

Huber, J., Ariely, D., & Fischer, G. (2001). *Expressing preferences in a principal-agent task: A comparison of choice, rating and matching*. Unpublished manuscript, Fuqua School of Business, Duke University, Durham, NC.

Huber, J., & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research, 33*, 307–317.

John, J. A., & Williams, E. R. (1995). *Cyclic and computer generated designs* (2nd ed.). London: Chapman & Hall.

Kiefer, J. (1959). Optimum experimental designs (with discussion). *Journal of the Royal Statistical Society B*, *21*, 272–319.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, *2*, 849–879.

Kiefer, J., & Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, *12*, 363–366.

Kuhfeld, W. F., Tobias, R. D., & Garratt, M. (1994). Efficient experimental design with marketing research applications. *Journal of Marketing Research*, *31*, 545–557.

Lazari, A. G., & Anderson, D. A. (1994). Designs of discrete choice set experiments for estimating both attribute and availability cross effects. *Journal of Marketing Research*, *31*, 375–383.

Louviere, J. J., & Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data. *Journal of Marketing Research*, *20*, 350–367.

Louviere, J. J., & Woodworth, G. G. (1988). On the design and analysis of correlated conjoint experiments using difference designs. *Advances in Consumer Research*, *15*, 510–517.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, *1*, 1–27.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decisions*, *8*, 229–254.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

Müller, W. G., & Ponce de Leon, A. C. M. (1996) Optimal design of an experiment in economics. *Economic Journal*, *106*, 122–127.

Pukelsheim, F. (1993). *Optimal design of experiments*. New York: Wiley.

Pullman, M. E., Dodson, K. J., & Moore, W. L. (1999). A comparison of conjoint methods when there are many attributes. *Marketing Letters*, *10*, 125–138.

Raghavarao, D. (1971). *Constructions and combinatorial problems in design of experiments*. New York: Wiley.

Sándor, Z., & Wedel, M. (in press). Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research.*

Schwabe, R. (1996). Optimum designs for multi-factor models. *Lecture Notes in Statistics* (Vol. 113). New York: Springer.

Shah, K. R., & Sinha, B. K. (1989). Theory of optimal designs. *Lecture Notes in Statistics* (Vol. 54). Berlin: Springer.

Smith, K. (1918). On the standard deviation of adjusted and interpolated values of an observed polynomial function and its constants and the guidance towards a proper choice of the distribution of observations. *Biometrika*, *12*, 1–85.

Steckel, J. H., DeSarbo, W. S., & Mahajan, V. (1991). On the creation of acceptable conjoint analysis experimental designs. *Decision Sciences*, *22*, 435–442.

Teichert, T. (2000). Auswirkungen von Verfahrensalternativen bei der Erhebung von Präferenzurteilen. *Marketing ZFP*, *2*, [*Effects of different methods on the elicitation of preference judgments*] 145–159.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273–286.

van Berkum, E. E. M. (1987a). Optimal paired comparison designs for factorial and quadratic models. *Journal of Statistical Planning and Inference*, *15*, 265–278.

van Berkum, E. E. M. (1987b). *Optimal paired comparison designs for factorial experiments*. CWI Tracts (Vol. 31). Amsterdam.

van Berkum, E. E. M. (1989). Reduction of the number of pairs in paired comparison designs and exact designs for quadratic models. *Computational Statistics & Data Analysis*, *8*, 93–107.

van Lenthe, J. (1993). ELI: An interactive elicitation technique for subjective probability distributions. *Organizational Behavior and Human Decision Processes*, *55*, 379–413.

West, P. M. (1996). Predicting preferences: An examination of agent learning. *Journal of Consumer Research*, *23*, 68–80.

Wynn, H. P. (1970). The sequential generation of *D*-optimum experimental designs. *Annals of Statistics*, *5*, 1655–1664.

# A DECISION THEORETIC FRAMEWORK FOR PROFIT MAXIMIZATION IN DIRECT MARKETING

Lars Muus, Hiek van der Scheer and Tom Wansbeek

## ABSTRACT

*One of the most important issues facing a firm involved in direct marketing is the selection of addresses from a mailing list. When the parameters of the model describing consumers' reaction to a mailing are known, addresses for a future mailing can be selected in a profit-maximizing way. Usually, these parameters are unknown and have to be estimated. These estimates are used to rank the potential addressees and to select the best targets.*

*Several methods for this selection process have been proposed in the recent literature. All of these methods consider the estimation and selection step separately. Since estimation uncertainty is neglected, these methods lead to a suboptimal decision rule and hence not to optimal profits. We derive an optimal Bayes decision rule that follows from the firm's profit function and which explicitly takes estimation uncertainty into account. We show that the integral resulting from the Bayes decision rule can be either approximated through a normal posterior, or numerically evaluated by a Laplace approximation or by Markov chain Monte Carlo integration. An empirical example shows that indeed higher profits result.*

---

**Econometric Models in Marketing, Volume 16, pages 119–140.**

**ISBN: 0-7623-0857-5**

# 1. INTRODUCTION

Consider a firm engaged in direct marketing, which has to decide which households within a large population to send a mailing. In order to decide which households to target, it is of crucial importance for the firm to assess how the household's response probability depends on its characteristics (demographic variables, attitudes, etc.) known to the firm. If the effect of the characteristics on the response probability are known, potential addressees can be ranked and the most promising ones can be selected.

Of course, these effects are unknown and have to be estimated. Typically, a firm specifies and estimates a *response model* based on a test mailing to get to know the effects of the characteristics on the response probability. Given the zero-one nature of the response variable in the simplest case, the logit model (and, to a lesser degree, the probit model) is frequently used for the purpose.

Given the growing importance of direct marketing, there has been an upsurge in research in the field to extend the basic methodology. Roberts and Berger (1999) provide a recent overview of a wide variety of techniques for the purpose. More in particular, recent research includes the following.

One issue is model selection. In the frequently occurring context of databases containing hundreds of variables, model selection is non-trivial. Levin, Zahavi and Olitsky (1995) propose an expert system, called AMOS, for the selection of variables to be optimally included in a model to predict customer behavior.

Many papers have addressed the use of more sophisticated methods beyond logit or probit. Bult and Wansbeek (1995) explore the use of a non-parametric alternative (i.c., the Cosslett estimator) for the usual discrete choice model in order to obviate undue parametric assumptions. Another kind of flexibility is offered by a number of relatively new statistical techniques whose potential for target selection is increasingly explored, like neural networks, e.g. Zahavi and Levin (1995, 1997), genetic algorithms, e.g. Ratner (1998) and Coates, Doherty and French (1999), and fuzzy logic, e.g. Openshaw (1996). Given the high noise-to-signal ratio in the typical direct marketing database, it is not yet clear whether the high level of sophistication offered by these methods will after all pay off in the field of target selection.

An important but methodologically difficult aspect concerns the dynamics of customer relations. In most cases, a firm involved in direct marketing wishes to establish a lasting relation with its customers, calling for methods that somehow optimize over time. The most profound contribution here is by Gönül and Shi (1998), who use dynamic programming methods to optimize mailings to a database over time. Gönül, Kim and Shi (2000) employ a hazard function approach to model intertemporal behavior of customers being sent catalogs, taking observed and unobserved heterogeneity into account.

Often, the outcome of a mailing has a quantity component in addition to the zero-one-component, calling for an extension of selection methods such that some quantity of the order is taken into account, like the number of items ordered or the amount of donation made. This topic has, e.g. been researched by Bult and Wittink (1996), Otter, van der Scheer and Wansbeek (1999), and Jonker, Paap and Franses (2000). By way of other extensions, Bult, van der Scheer and Wansbeek (1997) and Spring, Leeflang and Wansbeek (1999) investigate various aspects of offer design on response. Koning, Spring and Wansbeek (2001) model selection taking secondary action (pay or not, return or not) into account after the primary action of ordering items after being triggered by direct mail.

Anyhow, what most of these approaches have in common is essentially a model with estimates of the effects of variables on behavior. The estimates are used to formulate a *decision rule* to select households from a mailing list. However, this separation of parameter estimation and formulation of decision rules does not, in general, lead to optimal profits since a suboptimal decision rule is specified (Klein et al., 1978).

The reason for this suboptimality is that estimation usually takes place by considering (asymptotic) squared-error loss, which puts equal weight at over- and under-estimating the parameters. However, while a squared-error loss function may be useful when summarizing properties of the response function, it completely ignores the economic objectives of the marketing firm. Rather, the inferential process should be embedded in the firm's decision-making framework, taking explicitly into account the firm's objective of maximizing expected profit. Put differently, the decision maker should take the *estimation risk* into account when formulating a decision rule regarding which households to solicit. The loss resulting structure is, in general, asymmetric in contrast to the traditional squared-error loss structure. Consequently, the traditional methods thus yield suboptimal decision rules.

The purpose of this paper is to formulate a strict decision theoretic framework for a marketing firm engaged in direct marketing. In particular, we derive an optimal Bayes rule deciding when to send a mailing to a household with a given set of characteristics. This formal approach has a number of advantages. First of all, a rigorous decision theoretic framework clarifies the essential ingredients entering the marketing firm's decision problem. By deriving the optimal Bayes rule based on an expected profit loss function, the present framework yields admissible decision rules with respect to the marketing firm's economic objective. Furthermore, the estimation uncertainty resulting from the firm's assessment of the characteristics of the population of potential targets is explicitly taken into account as an integral part of the optimal decision procedure. Thus, the decision theoretic procedure provides a more firm theoretical foundation for optimal decision making on the part of the firm. Equally important, the present framework provides decision rules yielding higher profits to the firm.

Integration of the estimation and decision step has been studied thoroughly in statistics (e.g. Berger 1985, DeGroot 1970). This formal decision theoretic framework has been applied in a number of economic decision-making situations, including portfolio selection (cf. Bawa, Brown & Klein, 1979), real estate assessment (Varian, 1975), and agricultural economics (e.g. Lence & Hayes, 1994). For further economic applications see Cyert and DeGroot (1987). To the best of our knowledge, only one paper on optimal decision making under uncertainty has been applied to marketing questions (Blattberg & George, 1992). These authors consider a firm whose goal it is to maximize profits by determining the optimal price. They conclude that the firm is better off by charging a higher price than the price resulting from traditional methods, which are based on the estimated price sensitivity parameter. However, in contrast with our approach, they consider a loss function that results from a rather ad-hoc specified model, with only one unknown parameter.

The paper is organized as follows. In the next section we formulate the decision theoretic framework and derive the optimal Bayes decision rule. We show that the decision rule crucially depends on the estimation uncertainty facing the firm. The estimation uncertainty can be incorporated through a posterior density. In Section 3 we derive a closed form expression for the integral resulting from the optimal decision rule by approximating the posterior by the asymptotically normal density of the maximum likelihood (probit) estimator. In Section 4 we discuss the Laplace approximation and Markov chain Monte Carlo integration, which can be used to calculate the integral of interest. In Section 5 we discuss an empirical example, using data provided by a charity firm. Applying the formal decision framework appears to generate the higher profits indeed. We conclude in Section 6.

## 2. THE DECISION THEORETIC FRAMEWORK

Consider a direct marketing firm that has the option of mailing or not mailing to potential targets. In case a mail is sent to a given household the profit to the firm, $\pi$, is given by

$$\pi = rR - c,$$

where $r$ is the revenue from a positive reply, $c$ is the mailing cost, and $R$ is a random variable given by

$$R = \begin{cases} 1 & \text{if the household responds} \\ 0 & \text{if the household does not respond.} \end{cases}$$

Clearly, $c < r$ if the firm has to obtain positive profits at all. We assume that the response is driven by a probit model. Hence, the response probability of a household is

$$P(R = 1 \mid x, \beta) = \Phi(x'\beta),$$

where $\Phi(\cdot)$ is the standard normal integral, $x$ is a $k \times 1$ vector of regressors and $\beta$ is a $k \times 1$ vector of regression coefficients ($\beta \in \mathcal{B} \subseteq \mathbb{R}^k$). In case a mail is sent, the expected profit given $x$ and $\beta$ is

$$\mathrm{E}(\pi \mid x, \beta) = r\mathrm{E}(R \mid x, \beta) - c = r\Phi(x'\beta) - c. \quad (1)$$

With an unknown $\beta$, the firm has to make a decision whether to send a mail ($d = 1$) or not ($d = 0$) to a given household. The loss function considered in the following is given by

$$\mathcal{L}(d, \beta \mid x) = \begin{cases} r\Phi(x'\beta) - c & \text{if} \quad d = 1 \\ 0 & \text{if} \quad d = 0. \end{cases} \quad (2)$$

Notice, that the above loss function is naturally induced by the firm's economic profit maximization objective. In this sense, the present decision theoretic framework naturally encompasses the phenomena of estimation uncertainty, without introducing rather ad hoc statistical criteria.

Inference on the parameter vector $\beta$ is obtained through a test mailing, resulting in the sample

$$S_n \equiv \{(x_1, R_1), \ldots, (x_n, R_n)\}.$$

The posterior density, using Bayes' rule, is given by

$$f(\beta \mid S_n, \theta) = \frac{L(\beta \mid S_n) f(\beta \mid \theta)}{f(S_n \mid \theta)}, \quad (3)$$

where $L(\beta \mid S_n)$ is the likelihood function corresponding to the sample,

$$L(\beta \mid S_n) = \prod_{i=1}^{n} \Phi(x_i'\beta)^{R_i} (1 - \Phi(x_i'\beta))^{1 - R_i},$$

and $f(\beta \mid \theta)$ denotes the prior density, $\theta \in \Theta \subseteq \mathbb{R}^p$ is a $p \times 1$ vector of hyperparameters. Finally, $f(S_n \mid \theta)$ denotes the predictive density given by

$$f(S_n \mid \theta) = \int L(\beta \mid S_n) f(\beta \mid \theta) \, \mathrm{d}\beta. \tag{4}$$

The *posterior risk* corresponding to the loss function (2) is then given by

$$\begin{aligned} \mathcal{R}(d \mid x) &\equiv \mathrm{E}(\mathcal{L}(d, \beta \mid x) \mid S_n) \\ &= \begin{cases} r \int \Phi(x'\beta) f(\beta \mid S_n, \theta) \, \mathrm{d}\beta - c & \text{if} \quad d = 1 \\ 0 & \text{if} \quad d = 0. \end{cases} \end{aligned} \tag{5}$$

The *Bayes decision rule* corresponding to the posterior risk (5) is the decision variable $d$ maximizing $\mathcal{R}(d \mid x)$. It is easily seen that this decision rule is given by

$$d = 1 \qquad \text{if and only if} \qquad \int \Phi(x'\beta) f(\beta \mid S_n, \theta) \, \mathrm{d}\beta \geq \frac{c}{r}. \tag{6}$$

Notice that this decision rule explicitly takes into account the estimation uncertainty inherent when the firm does not know the parameter vector $\beta$. In general, we denote by the *mailing region* a subspace of $\mathbb{R}^k$ containing vectors $x$ corresponding with households to whom a mailing should be sent. According to the above, the *Bayes optimal mailing region* is given by

$$\mathcal{M}_{\mathrm{B}} \equiv \left\{ x \in \mathbb{R}^k \,\Big|\, \int \Phi(x'\beta) f(\beta \mid S_n, \theta) \, \mathrm{d}\beta \geq \frac{c}{r} \right\}.$$

The structure of the mailing region may, in general, be quite complicated.

It is often recommended to base the firm's mailing decision on the point estimates obtained from the test mailing. These point estimates are typically derived by implicitly assuming a squared-error loss function, resulting from the use of standard estimation procedures. As this squared-error loss does not reflect the actual loss suffered by the firm, using the point estimate motivated by squared-error loss will be inappropriate. If the firm neglects the estimation uncertainty it would specify a decision rule based on a point estimate of $\beta$, say $\hat{\beta}$, e.g. the probit estimator based on $S_n$. The point estimate then is used as if it is the true parameter value (e.g. Bult and Wansbeek, 1995). The resulting decision rule, which we call the *naive* decision rule, is thus given by

$$d = 1 \qquad \text{if and only if} \qquad \Phi(x'\hat{\beta}) \geq \frac{c}{r}. \tag{7}$$

This rule evidently ignores the estimation uncertainty surrounding $\hat{\beta}$. Indeed, by a second order Taylor series expansion of $\Phi(x'\beta)$, we obtain

$$\Phi(x'\beta) \approx \Phi(x'\hat{\beta}) + (\beta - \hat{\beta})'x\phi(x'\hat{\beta}) - \frac{1}{2}x'\hat{\beta}\phi(x'\hat{\beta})x'(\beta - \hat{\beta})(\beta - \hat{\beta})'x,$$

using the fact that the derivative of $\phi(t)$ is $-t\phi(t)$, where $\phi(\cdot)$ is the standard normal density. Hence, an approximate Bayes decision rule is given by,

$$\Phi(x'\hat{\beta}) - \frac{1}{2}x'\hat{\beta}\phi(x'\hat{\beta})x'Mx \geq \frac{c}{r} \tag{8}$$

where $M \equiv \mathrm{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$ denotes the mean square error matrix of the estimator $\hat{\beta}$. The major difference between the (approximate) Bayes rule (8) and the naive rule (7) is that estimation uncertainty is explicitly taken into account in the former. Evidently, if the estimation uncertainty is small, i.e. $M$ is small, the approximate Bayes rule (8) is adequately approximated by the naive decision rule (8). Notice that the mailing region for the naive rule is the half space given by

$$\mathcal{M}_{\mathrm{N}} \equiv \left\{ x \in \mathbb{R}^k \mid \Phi(x'\hat{\beta}) \geq \frac{c}{r} \right\}$$

The result of applying the naive decision rule is thus to approximate the mailing region $\mathcal{M}_{\mathrm{B}}$ by the halfspace $\mathcal{M}_{\mathrm{N}}$. As will be demonstrated below this approximation may be rather crude, resulting in a suboptimal level of profits.

In order to implement the optimal decision rule (6), we need to evaluate the expectation of $\Phi(x'\beta)$ over the posterior density of $\beta$. If the posterior admits a closed form solution and is of a rather simple analytical form, this expectation can be solved analytically. Otherwise, numerical methods need to be implemented in order to assess the decision rule (6). In Section 4 we explore various numerical strategies for evaluating the decision rule. However, it is instructive to consider the case where the posterior density is normal, in which case we can fully characterize the mailing region.

## 3. THE CASE OF A NORMAL POSTERIOR

If the posterior density is normal with mean $\mu$ and covariance matrix $\Omega$, we can obtain a closed form expression for (6), namely*

$$\int \Phi(x'\beta)f(\beta \mid S_n, \theta)\, \mathrm{d}\beta$$

$$\begin{aligned}
&= \mathrm{E}_\beta(\Phi(x'\beta)) && \text{where } \beta \sim \mathrm{N}(\mu, \Omega) \\
&= \mathrm{E}_b(\Phi(x'\Omega^{1/2}b + x'\mu)) && \text{where } b = \Omega^{-1/2}(\beta - \mu) \sim N(0, I_k) \\
&= \mathrm{E}_b\mathrm{E}_z I_{\{-\infty,\, x'\Omega^{1/2}b + x'\mu\}}(z) && \text{with } z \sim N(0, 1), \text{ independent of } b \\
&= \mathrm{E}_b\mathrm{E}_z I_{-\infty,\, x'\mu\}}(z - x'\Omega^{1/2}b) \\
&= P(z - x'\Omega^{1/2}b < x'\mu) \\
&= \Phi\left(\frac{x'\mu}{(1 + x'\Omega x)^{1/2}}\right). && (9)
\end{aligned}$$

Hence, the mailing region is given by

$$\begin{aligned}
\mathcal{M}_\mathrm{B} &= \left\{ x \in \mathbb{R}^k \mid \Phi\left(\frac{x'\mu}{(1 + x'\Omega x)^{1/2}}\right) \geq \frac{c}{r} \right\} \\
&= \{ x \in \mathbb{R}^k \mid x'\mu \geq \gamma(1 + x'\Omega x)^{1/2} \} \qquad (10)
\end{aligned}$$

where

$$\gamma \equiv \Phi^{-1}\left(\frac{c}{r}\right).$$

Since in any practical situation $c \ll r$, we assume $\gamma < 0$ whenever the sign of $\gamma$ is relevant. Notice that, when $\Omega_1 > \Omega_2$, $1 + x'\Omega_1 x > 1 + x'\Omega_2 x$. Thus, since $\gamma < 0$, greater uncertainty as to $\beta$ implies that the mailing region expands.

---

* We are indebted to Ton Steerneman for bringing the resultto our attention, and for providing this derivation.

Expression (9) enables us to show explicitly that the Bayes decision rule generates higher expected profits than the naive decision rule. The expected profit (cf. (1)), in case mail is sent, is

$$
\begin{aligned}
q(x) &\equiv E_{\beta}(E(\pi \mid x, \beta)) \\
&= r\Phi\left(\frac{x'\mu}{(1+x'\Omega x)^{1/2}}\right) - c.
\end{aligned}
$$

For all $x$ in $\mathcal{M}_B$ there holds, by definition, that $q(x) > 0$. Since $\mathcal{M} \subseteq \mathcal{M}_B$ it follows that the expected profit is lower for the naive decision rule.

We consider this mailing region in somewhat more detail. The boundary of the mailing region $\mathcal{M}_B$ is given by

$$\{x \in \mathbb{R}^k \mid x'\mu = \gamma(1+x'\Omega x)^{1/2}\} \tag{11}$$

We assume that $\Omega > 0$. By squaring and rewriting the argument of (11) we obtain

$$x'(\mu\mu' - \gamma^2\Omega)x = \gamma^2, \tag{12}$$

which can be written as

$$x'\Omega^{1/2}(\Omega^{-1/2}\mu\mu'\Omega^{-1/2} - \gamma^2 I_k)\,\Omega^{1/2}x = \gamma^2. \tag{13}$$

Let

$$
\begin{aligned}
A_1 &\equiv \frac{\Omega^{-1/2}\mu\mu'\Omega^{-1/2}}{\mu'\,\Omega^{-1}\,\mu} \\
A_2 &\equiv I_k - A_1 \\
\lambda &\equiv \mu'\Omega^{-1}\mu - \gamma^2;
\end{aligned}
$$

$A_1$ and $A_2$ are idempotent matrices of rank 1 and $k-1$, respectively, $A_1A_2 = 0$, and $A_1 + A_2 = I_k$. Hence, we can write (13) as

$$x'\Omega^{1/2}(\lambda A_1 - \gamma^2 A_2)\Omega^{1/2}x = \gamma^2.$$

Let $A_1 = z_1z_1'$ and $A_2 = Z_2Z_2'$, so $(z_1, Z_2)$ is orthonormal. Then

$$
\begin{aligned}
G &\equiv \lambda A_1 - \gamma^2 A_2 \\
&= \lambda z_1 z_1' - \gamma^2 Z_2 Z_2' \\
&= (z_1, Z_2)\begin{pmatrix} \lambda & 0 \\ 0 & -\gamma^2 I_{k-1} \end{pmatrix}\begin{pmatrix} z_1' \\ Z_2' \end{pmatrix}
\end{aligned}
$$

Hence, the eigenvalues of $G$ are $-\gamma^2$ with multiplicity $k-1$, and $\lambda$ with multiplicity one. The sign of $\lambda$ depends on $\Omega$. Informally speaking, for small values of $\Omega$, $\lambda > 0$, and for large values, $\lambda < 0$. In the first case $G$ has one positive and $k-1$ negative eigenvalues. Due to 'Sylvester's law of inertia' (e.g. Lancaster & Tismenetsky, 1985, p. 188), the same holds for $\mu\mu' - \gamma^2\Omega$. Hence, the matrix is indefinite and the boundary is a hyperboloid in the $x$-space. When the uncertainty as to $\beta$ is so large that $\lambda < 0$, all eigenvalues of $G$ are negative and (12) does not have a solution. Hence, all households should be included in the mailing campaign.

We illustrate the mailing region is for $k=2$, $\mu' = (1, 1)$, $\gamma = -1$, and

$$\Omega = \begin{pmatrix} \sigma^2 & \sigma_{12} \\ \sigma_{12} & \sigma^2 \end{pmatrix}.$$

Then, from (10), the mailing region is

$$\mathcal{M}_B = \{x_1, x_2 \mid x_1 + x_2 \geq -\sqrt{1 + \sigma^2(x_1 + x_2) + 2\sigma_{12}x_1x_2}\},$$

which reduces to the halfspace $x_1 + x_2 \geq -1$ if $\sigma^2 = \sigma_{12} = 0$. The matrix in (12) becomes

$$\mu\mu' - \gamma^2\Omega = \begin{pmatrix} 1-\sigma^2 & 1-\sigma_{12} \\ 1-\sigma_{12} & 1-\sigma^2 \end{pmatrix}$$

$$= \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 2-\sigma^2-\sigma_{12} & 0 \\ 0 & -(\sigma^2-\sigma_{12}) \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \tag{14}$$

Hence, the matrix $\mu\mu' - \gamma^2\Omega$ has one negative eigenvalue, $-(\sigma^2 - \sigma_{12})$, and one eigenvalue that is positive if $\sigma^2 + \sigma_{12} < 2$. Using (14), (12) can be rewritten as

$$(2 - \sigma^2 - \sigma_{12})(x_1 + x_2)^2 - (\sigma^2 - \sigma_{12})(x_1 - x_2)^2 = 2,$$

which is a hyperbola in $\mathbb{R}^2$. Its asymptotes are found by putting the left-hand side equal to zero. On letting

$$\varphi \equiv \sqrt{\frac{2 - \sigma^2 - \sigma_{12}}{\sigma^2 - \sigma_{12}}},$$

these asymptotes are found to be

$$\varphi(x_1 + x_2) = \pm(x_1 - x_2),$$

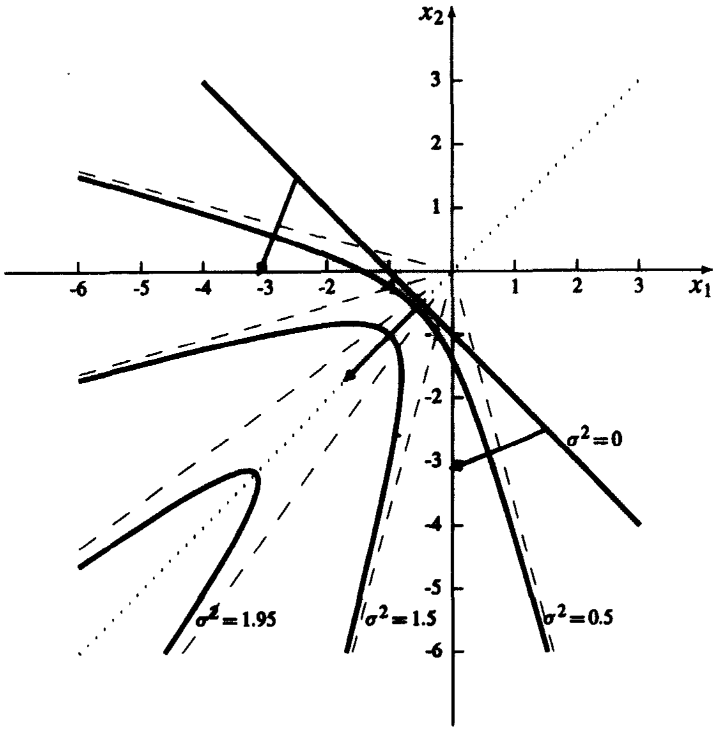


*Fig. 1.* The naive and Bayes optimal mailing region compared. The area to the north-east of the straight line is $\mathcal{M}_N$, and the ellipsoids bound $\mathcal{M}_B$ for various values of $\sigma^2$.

or

$$\frac{x_2}{x_1}=\frac{1-\varphi}{1+\varphi} \quad \text{and} \quad \frac{x_2}{x_1}=\frac{1+\varphi}{1-\varphi}.$$

Figure 1 illustrates the boundary for $\sigma_{12}=0$, and $\sigma^2=0$, 0.5, 1.5, and 1.95, respectively. If $\sigma^2=0$ we have a straight line. This bounds the mailing region of the naive method or the mailing region increases as $\sigma^2$ increases; the arrows indicate the direction of the increase. When $\sigma^2 \geq 2$, the mailing region is simply $\mathbb{R}^2$. The distance between the straight line corresponding with $\sigma^2=0$ and the hyperbola is larger when the $x$-value is larger. This reflects the fact that the uncertainty as to $x'\beta$ increases by the (absolute) value of $x$.

# 4. NUMERICAL EVALUATION OF THE OPTIMAL BAYES RULE

Numerical implementation of the optimal Bayes decision rule (6) requires the evaluation, for each value of $x$, of the integral

$$Q(x) \equiv \int \Phi(x'\beta) f(\beta \mid S_n, \theta)\, d\beta \tag{15}$$

$$= \frac{\int \Phi(x'\beta) L(\beta \mid S_n) f(\beta \mid \theta)\, d\beta}{\int L(\beta \mid S_n) f(\beta \mid \theta)\, d\beta}, \tag{16}$$

using (3) and (4) in the last step. We will now explore various methods for evaluating this integral. Henceforth, we denote the probit estimate of $\beta$, based on $S_n$, by $\hat{\beta}$, and covariance matrix by $\hat{\Omega}$ (e.g. the inverse of the Fisher information matrix evaluated in $\hat{\beta}$).

### *Normal Posterior Approximation*

It is well known that the posterior density converges under suitable regularity conditions to a normal distribution, with mean $\hat{\beta}$ and covariance matrix $\hat{\Omega}$, when the sample size is sufficiently large (Jeffreys, 1967, p. 193, Heyde & Johnstone, 1979). Obviously, the approximation may be rather crude, since it is solely based on the asymptotic equivalence of the Bayes and maximum likelihood estimator. Thus, this approximation completely ignores the prior distribution $f(\beta \mid \theta)$. However, as we showed in Section 3, this property appears to be very valuable since it enables us to obtain a closed form expression for (15), which is given in (9) by substitution of $\hat{\beta}$ for $\mu$ and $\hat{\Omega}$ for $\Omega$. Moreover, Zellner and Rossi (1984) showed that, for moderate sample sizes ($n = 100$), the normal posterior approximation works well for the logit model.

### *Laplace Approximation*

A more refined asymptotic approximation is the Laplace approximation proposed by Tierney and Kadane (1986) (see also Kass et al., 1988, and Tierney et al., 1989). The Laplace approximation of (16) is given by

$$\hat{Q}(x) = \frac{\Psi_1(\hat{\beta}_1) \mid H_1(\hat{\beta}_1) \mid^{-1/2}}{\Psi_0(\hat{\beta}_0) \mid H_0(\hat{\beta}_0) \mid^{-1/2}}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the maximizers of $\Psi_0(\cdot)$ and $\Psi_1(\cdot)$, respectively, and

$$\Psi_0(\beta) \equiv L(\beta \mid S_n) f(\beta \mid \theta)$$
$$\Psi_1(\beta) \equiv \Phi(x'\beta) L(\beta \mid S_n) f(\beta \mid \theta),$$

and

$$H_0(\beta) \equiv -\frac{\partial^2 \ln \Psi_0(\beta)}{\partial\beta\partial\beta'}$$
$$H_1(\beta) \equiv -\frac{\partial^2 \ln \Psi_1(\beta)}{\partial\beta\partial\beta'}.$$

By means of the Laplace approximation, the integral $Q(x)$ is thus evaluated without any need for numerical integration. Instead the Laplace approximation requires maximization, in order to determine $\hat{\beta}_0$ and $\hat{\beta}_1$, and differentiation, in order to find $H_0(\cdot)$ and $H_1(\cdot)$. For $\hat{\beta}_0$ and $\hat{\beta}_1$ we use the values obtained by a single Newton-Raphson step from $\hat{\beta}$ when maximizing $\ln \Psi_0(\beta)$ and $\ln \Psi_1(\beta)$, which does not affect the rate at which the approximation error vanishes. As demonstrated by Tierney and Kadane (1986), Kass et al. (1988), and Tierney et al. (1989), the general error of the approximation vanishes at rate $n^{-2}$. As these authors demonstrate, this approximation is often very accurate.

We apply this approximation for an informative prior and an uninformative prior. As to the former we choose for $f(\beta \mid \theta)$ the normal density with mean $\hat{\beta}$ and covariance matrix $\hat{\Omega}$. Since, $\partial \ln f(\beta \mid \theta)/\partial\beta = -\hat{\Omega}^{-1}(\beta - \hat{\beta})$, we have $\hat{\beta}_0 = \hat{\beta}$, and in the Appendix we show that

$$\hat{\beta}_1 = \hat{\beta} + \xi(\hat{\beta}) H_0(\hat{\beta})^{-1} x, \tag{17}$$

where $\xi(\cdot)$ is a scalar function defined in (18).

For the uninformative prior we use Jeffreys' prior (e.g. Berger, 1985, pp. 82–89, and Zellner 1971, pp. 41–53), given by

$$f(\beta \mid \theta) = \left| -\mathrm{E}\left( \frac{\partial^2 \ln L(\beta \mid S_n)}{\partial\beta\partial\beta'} \right) \right|^{1/2}.$$

Notice that no hyperparameters are involved here. Within the context of binary response models this prior has been examined by, among others, Ibrahim and Laud (1991), and Poirier (1994). These authors support the use of Jeffreys' prior as an uninformative prior but notice that it can be quite cumbersome to work with analytically as well as numerically.

### *Monte Carlo Integration*

The recent development of *Markov chain Monte Carlo* (MCMC) procedures has revolutionized the practice of Bayesian inference. See, for example, Tierney (1994), and Gilks et al. (1995) for expositions of basic Markov chain Monte Carlo procedures. These algorithms are easy to implement and have the advantage that they do not require evaluation of the normalizing constant of the posterior density, given by (4). As a *candidate density* it is natural to select the asymptotic approximation, $q(\beta) \sim N(\hat{\beta}, \hat{\Omega})$. The density of interest, the so-called *target density*, is given by

$$h(\beta) \equiv L(\beta \mid S_n) f(\beta \mid \theta).$$

The independence sampler (Tierney 1994), a special case of the Hastings-Metropolis algorithm, is used to generate random variates $\beta_j$, $j = 1, \ldots, J$, from the (unnormalized) density $h(\beta)$ through the following algorithm, where $\beta_0$ is arbitrarily selected:

(1) draw a *candidate point*, $\beta_j^*$, from $q(\cdot)$
(2) draw $u_j$ from the uniform density on (0, 1)
(3) if $u_j \leq \alpha(\beta_{j-1}, \beta_j^*)$, then $\beta_j = \beta_j^*$ else $\beta_j = \beta_{j-1}$.

Here

$$\alpha(\beta_{j-1}, \beta_j^*) \equiv \begin{cases} \min\left(\dfrac{h(\beta_j^*) q(\beta_{j-1})}{h(\beta_{j-1}) q(\beta_j^*)}\right), & \text{if} \quad h(\beta_{j-1}) q(\beta_j^*) > 0 \\ 1 & \text{else.} \end{cases}$$

The generated $\beta_j$'s, $j = 1, \ldots, J$ are used to evaluate the integral by

$$\hat{Q}(x) = \frac{1}{J} \sum_{j=1}^{J} \Phi(x' \beta_j).$$

We use this algorithm instead of more advanced MCMC procedures, like the Gibbs sampler (e.g. Albert & Chib, 1993), since we have a candidate density that is a good approximation of the target distribution (Roberts, 1995). Again, we apply this algorithm for the (informative) normal prior and for the (uninformative) Jeffreys' prior.

## 5. ILLUSTRATION

We illustrate our approach with an application based on data from a charitable foundation in the Netherlands. This foundation heavily rests on direct mailing. Every year it sends mailings to almost 1.2 million households in the Netherlands. The dependent variable is the response/non-response in 1991. The explanatory variables are the amount of money (in NLG) donated in 1990

(A90) and 1989 (A89), the interaction between these two (INT), the date of entry on the mailing list (ENTRY), the family size (FS), own opinion on charitable behavior in general (CHAR; four categories: donates never, donates sometimes, donates regularly, and donates always). The data set consists of 40,000 observation. All the households on the list donated at least once to the foundation since entry on the mailing list. As a result, the data set does not constitute a random sample. It is not clear whether this induces any distortion in comparing methods.

In order to have a sufficiently large validation sample we used 1,000 observations for estimation. The response rate in the estimation sample is 31.8%. This rather high response rate is not surprising since charitable foundations have in general high response rates (Statistical Fact Book 1994–1995), and the mailing list consists of households that responded to this particular foundation before. The average amount of donation in the estimation sample is NLG 14.56, the cost of a mailing is NLG 3.50. We use the average amount of donation for household selection and to determine the profit implications. Table 1 gives the probit estimates and the average of the coefficients based on the independence sampler with the normal and Jeffreys' prior, respectively. The donation in 1990 and 1989 are, as expected, positively

***Table 1.*** Probit Estimates and Results of the Independence Sampler.

| | Probit Estimates[1] | Independence Sampler[2] | |
|---|---|---|---|
| | | Normal prior | Jeffreys' prior |
| Constant | −0.3938 | −0.3964 | −0.3948 |
| | (0.4511) | (0.3120) | (0.4539) |
| A90 | 0.0052 | 0.0053 | 0.0051 |
| | (0.0014) | (0.0010) | (0.0014) |
| A89 | 0.0074 | 0.0074 | 0.0072 |
| | (0.0030) | (0.0021) | (0.0030) |
| INT | −0.0056 | −0.0057 | −0.0053 |
| | (0.0029) | (0.0019) | (0.0027) |
| ENTRY | −0.0063 | −0.0063 | −0.0063 |
| | (0.0048) | (0.0033) | (0.0048) |
| FS | −0.1526 | −0.1503 | −0.1513 |
| | (0.1408) | (0.1003) | (0.1397) |
| CHAR | 0.0683 | 0.0680 | 0.0685 |
| | (0.0537) | (0.0371) | (0.0530 |

[1] Asymptotic standard errors in parentheses
[2] Standard deviation, based on $J = 10{,}000$, in parentheses

related with the response probability. The negative sign of the interaction term can be interpreted as a correction for overestimation of the response probability if a household responded in 1990 and 1989. The other three coefficients do not significantly differ from zero. As expected, the average value of the coefficients for the independence sampler are similar to the probit estimates. The standard deviations, however, of the normal prior are much smaller.

The basic difficulty in MCMC procedures is the decision when the generated sequence of parameters has converged to a sample of the target distribution. Many diagnostic tools to address this convergence problem have been suggested in the recent literature (see Cowles & Carlin, 1996 for an extensive overview). Following the recommendations of these authors, we generated six parallel sequences of parameters with starting points systematically chosen from a large number of drawings from a distribution that is overdispersed with respect to the target distribution. We inspected the sequences of each parameter by displaying them in a common graph and in separate graphs. We used the Gelman-Rubin statistics (Gelman & Rubin, 1992) to quantitatively analyze the sequences. The results of these diagnostics are satisfying, indicating an almost immediate convergence of the sample.

Table 2 shows the profit implications for the various approaches to determine the posterior risk function and the naive approach for the validation sample. As a benchmark we also give the situation in which the foundation sends all the households a mailing. Of these 39,000 households, 13,274 responded, generating a net profit of NLG 56,784. If the foundation would have used the naive selection approach they would have selected 87.03% (33,946) of the households, with a net profit of NLG 59,345. Using the Bayes decision rule, the foundation would have selected more households, as expected. This ranges

***Table 2.*** Target Selection and Profit Implications.

| | Number selected | Response | Actual profit (NLG) |
|---|---|---|---|
| No selection | 39 000 | 13 274 | 56 784 |
| Naive approach | 33 946 | 12 236 | 59 345 |
| Normal posterior | 34 240 | 12 337 | 59 787 |
| Laplace approximation: | | | |
| Normal prior | 34 018 | 12 250 | 59 297 |
| Jeffreys' prior | 34 256 | 12 341 | 59 789 |
| Independence sampler: | | | |
| Normal prior | 34 153 | 12 310 | 59 698 |
| Jeffreys' prior | 34 271 | 12 347 | 59 824 |

from 34,018 of the Laplace approximation with the normal prior to 34,271 of the independence sampler with Jeffreys' prior. Except for the Laplace approximation with the normal prior, the additional selected households generate sufficient response to increase the net profits, reinforcing the importance of the Bayes decision rule. Net profits increase with 4.5% if the naive selection is used instead of selecting all the households. This percentage increases to 5.3% if we apply the normal posterior approximation, and to 5.4% when using the independence sampler with Jeffreys' prior. Given that the foundation's database contains 1.2 million targets, these increases turn out to be quite substantial. Notice that the figures of the Laplace approximation and independence sampler with the normal prior are much closer to those of the naive approach than those with Jeffreys' prior. This makes intuitive sense since informative priors put more weight to values of $\beta$ near $\hat{\beta}$. In the case of the posterior density degenerating at $\hat{\beta}$, i.e. perfect prior information on $\beta$, the decision rule is equivalent to the naive rule.

## 6. DISCUSSION AND CONCLUSION

In order to select addresses from a list for a direct mailing campaign, a firm can build a response model and uses the (consistently) estimated parameters for selection. The decision rule for selection is often defined on the basis of the estimated parameters taken as the true parameters. This paper shows that this leads to suboptimal results. The reason for this is that the estimation uncertainty resulting from the firm's assessment of the characteristics of the potential targets is not taken into account. Put differently, both steps of a target selection process, estimation and selection, should be considered simultaneously. We formulated a rigorous theoretic framework, based on the firm's profit maximizing behavior, to derive an optimal Bayes decision rule. We demonstrated theoretically as well as empirically that this approach generates higher profits.

An important aspect of our approach is the evaluation of the integral resulting from the Bayes decision rule. We used a normal posterior, Laplace approximation, and Monte Carlo integration to evaluate the Bayes rule numerically. Although the normal posterior approach may be rather crude it has the advantage that a closed form expression can be derived, and, moreover, it performs quite well in the empirical illustration. As a consequence of the former, we do not need the computational intensive methods. Moreover, we obtain a transparent expression for the expected profit, which explicitly shows the effect of estimation risk. It has to be realized, however, that the empirical results indicate that the decision rule is affected by the chosen prior density.

Since the normal posterior approximation ignores the prior density, it has to be used with caution when prior information is available.

This paper has some limitations. First, we considered only the question of selecting households for one direct mailing campaign. That is, we did not consider the long-term impact of the selection process. Second, we solely considered the binary response choice to the mailing and not the amount of money donated. Third, we made the implicit assumption that the parameters are constant across households. This assumption may be unrealistic in practice. It runs, for example, counter to the idea of trying to customize promotions through direct marketing. A company could deal with this kind of heterogeneity by using, for example, latent class analysis (DeSarbo & Ramaswamy, 1994; Wedel et al., 1993). We want to stress, however, that these assumptions are commonly made in direct marketing research. Furthermore, our method results from a general decision theoretic framework that can be extended, in principle, to situations that do suffer from these limitations, in a straightforward manner.

## REFERENCES

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association, 88*, 669–679.

Balestra, P. (1976). *La dérivation matricielle*. Paris: Sirey.

Bawa, V. S., Brown, S. J., & Klein, R. W. (Eds) (1979). *Estimation Risk and Optimal Portfolio Choice*. Amsterdam: North-Holland.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York: Springer-Verlag.

Blattberg, R. C., & George, E. I. (1992). Estimation under profit-driven loss function. *Journal of Business & Economic Statistics, 10*, 437–444.

Bult, J. R., van der Scheer, H. R., & Wansbeek, T. J. (1997). Interaction between target and mailing characteristics in direct marketing, with an application to health care fund raising. *International Journal of Research in Marketing, 14*, 301–308.

Bult, J. R., & Wansbeek, T. J. (1995). Optimal selection for direct mail. *Marketing Science, 14*, 378–394.

Bult, J. R., & Wittink, D. R. (1996). Estimating and validating asymmetric heterogeneous loss functions applied to health care fund raising. *International Journal of Research in Marketing, 13*, 215–226.

Coates, D. S., Doherty, N. F., & French, A. P. (1999). The new multivariate jungle: Computer intensive methods in database marketing. In: G. J. Hooley & M. K. Hussey (Eds), *Quantitative Methods in Marketing* (pp. 404–420). London: Internation Thomson Business Press.

Cowles, M. K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association, 91*, 883–904.

Cyert, R. M., & DeGroot, M. H. (1987). *Bayesian Analysis and Uncertainty in Economic Theory*. London: Chapman and Hall.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

DeSarbo, W. S., & Ramaswamy, V. (1994). CRISP: Customer response based iterative segmentation procedures for response modeling in direct marketing. *Journal of Direct Marketing*, *8*(3), 7–20.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–511.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds) (1995). *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

Gönül, F., Kim, B.-D., & Shi, M. (2000). Mailing smarter to catalog customers. *Journal of Interactive Marketing*, *14*(2). 2–16.

Gönül, F., & Shi, M. (1998). Optimal mailing of catalogs: a new methodology using estimable structural dynamic programming models. *Management Science*, *44*, 1249–1262.

Heyde, C. C., & Johnstone, I. M. (1979). On asymptotic posterior normality for stochastic processes. *Journal of the Royal Statistical Society B*, *41*, 184–189.

Ibrahim, J. G., & Laud, P. W. (1991). On Bayesian analysis of generalized linear models using Jeffreys's prior. *Journal of the American Statistical Association*, *86*, 981–986.

Jeffreys, H. (1967). *Theory of Probability* (3rd ed.). London: Oxford University Press.

Jonker, J. J., Paap, R., & Franses, Ph. H. (2000). Modeling charity donations: Target selection, response time and gift size. Working paper, EI2000-07, Erasmus University, Rotterdam.

Kass, R. E., Tierney, L., & Kadane, J. B. (1988). Asymptotics in Bayesian computation. In: J. M. Bernardo, M. H. DeGroot, D. V. Lindley & A.F. M. Smith (Eds), *Bayesian Statistics* (Vol. 3). Oxford: Oxford University Press.

Klein, R. W., Rafsky, L. C., Sibley, D. S., & Willig, R. D. (1978). Decisions with estimation uncertainty. *Econometrica*, *46*, 1363–1387.

Koning, R. H., Spring, P. N., & Wansbeek, T. J. (2001). Primary and secondary action in direct marketing. Working paper, Department of Economics, University of Groningen.

Lancaster, P., & Tismenetsky, M. (1985). *The theory of matrices* (2nd ed.). Orlando: Academic Press.

Lence, S. H., & Hayes, D. J. (1994). The empirical minimum-variance hedge. *American Journal of Agricultural Economics*, *76*, 94–104.

Levin, N., Zahavi, J., & Olitsky, M. (1995). AMOS – A probabilty-driven, customer-oriented decision support system for target marketing of solo mailings. *European Journal of Operational Research*, *87*, 708–721.

Openshaw, S. (1996). Fuzzy logic models for targeting customers. *Journal of Targeting, Measurement and Analysis for Marketing*, *5*, 151–164.

Otter, P. W., van der Scheer, H. R., & Wansbeek, T. J. (1999). Direct mail selection by joint modeling of the probability and quantity of response. In: H. F. Chen (Ed.), *The Proceedings of the 14th World Congress of IFAC* (pp. 459–464). Oxford: Pergamon-Elsevier.

Poirier, D. (1994). Jeffreys' prior for logit models. *Journal of Econometrics*, *63*, 327–339.

Ratner, B. (1998). Direct marketing models using genetic algorithms. *Journal of Targeting, Measurement and Analysis for Marketing*, *6*, 328–343.

Roberts, G. O. (1995). Markov chain concepts related to sampling algorithms. In: W. R. Gilks, S. Richardson & D. J. Spiegelhalter (Eds), *Markov Chain Monte Carlo in Practice* (pp. 45–57). London: Chapman & Hall.

Roberts, M. L., & Berger, P. D. (1999). *Direct marketing management*. Englewood Cliffs, NJ: Prentice-Hall.

Spring, P. N., Leeflang, P. S. H., & Wansbeek, T. J. (1999). The combination strategy to optimal target selection and offer segmentation in direct mail. *Journal of Market Focused Management*, *4*, 187–203.

Statistical Fact Book (1994–1995). Direct Marketing Association, Inc., New York, NY.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, *22*, 1701–1762.

Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, *81*, 82–86.

Tierney, L., Kass, R. E., & Kadane, J. B. (1989). Fully exponential Laplace approximations to expectations and variances of non-positive functions. *Journal of the American Statistical Association*, *84*, 710–716.

Varian, H. R. (1975). A Bayesian approach to real estate assessment. In: S. E. Fienberg & A. Zellner (Eds), *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*. Amsterdam: North-Holland.

Wedel, M., DeSarbo, W. S., Bult, J. R., & Ramaswamy, V. (1993). A latent class Poisson regression model for heterogeneous count data. *Journal of Applied Econometrics*, *8*, 397–411.

Zahavi, J., & Levin, N. (1995). Issues and problems in applying neural computing to target marketing. *Journal of Direct Marketing*, *9*(3), 33–45.

Zahavi, J., & Levin, N. (1997). Applying neural computing to target marketing. *Journal of Direct Marketing*, *11*(1). 5–22.

Zellner, A. (1971). *An Introduction to Bayesian inference in econometrics*. New York: Wiley.

Zellner, A., & Rossi, P. E. (1984). Bayesian analysis of dichotomous quantal response models. *Journal of Econometrics*, *25*, 365–393.

# APPENDIX: ON THE LAPLACE APPROXIMATION

We first prove (17), then we give the derivatives of Jeffreys' prior. Let

$$g_0(\beta) \equiv \frac{\partial \ln \Psi_0(\beta)}{\partial \beta} = \frac{\partial \ln L(\beta \mid S_n)}{\partial \beta} - \hat{\Omega}^{-1}(\beta - \hat{\beta})$$

$$g_1(\beta) \equiv \frac{\partial \ln \Psi_1(\beta)}{\partial \beta} = \frac{\phi}{\Phi} x + g_0(\beta) = \zeta x + g_0(\beta),$$

where $\phi \equiv \phi(x'\beta)$, $\Phi \equiv \Phi(x'\beta)$, and $\zeta \equiv \dfrac{\phi}{\Phi}$ is the inverse of Mills' ratio. Notice that $g_0(\hat{\beta}) = 0$. Further,

$$H_0(\beta) \equiv -\frac{\partial^2 \ln L(\beta \mid S_n)}{\partial \beta \partial \beta'} + \hat{\Omega}^{-1}$$

$$H_1(\beta) \equiv \frac{\phi(\phi + x'\beta\Phi)}{\Phi^2} xx' + H_0(\beta) = \zeta(\zeta + x'\beta)xx' + H_0(\beta).$$

Then $\hat{\beta}_1$ follows from the Newton-Raphson step

$$\begin{aligned}\hat{\beta}_1 &= \hat{\beta} + H_1(\hat{\beta})^{-1} g_1(\hat{\beta})\\ &= \hat{\beta} + (\hat{\zeta}(\hat{\zeta} + x'\hat{\beta})xx' + H_0(\hat{\beta}))^{-1} g_1(\hat{\beta})\\ &= \hat{\beta} + \frac{1}{1 + \hat{\zeta}(\hat{\zeta} + x'\hat{\beta})x'H_0(\hat{\beta})^{-1}x} H_0(\hat{\beta})^{-1} g_1(\hat{\beta})\\ &= \hat{\beta} + \xi(\hat{\beta}) H_0(\hat{\beta})^{-1} x,\end{aligned}$$

where $\hat{\zeta}$ denotes $\zeta$ evaluated in $\hat{\beta}$, and

$$\xi(\beta) \equiv \frac{\zeta}{1 + \zeta(\zeta + x'\beta)x'H_0(\beta)^{-1}x}. \tag{18}$$

We will now derive the first and second derivative of Jeffreys' prior, given by

$$f(\beta \mid \theta) = \left| -\mathrm{E}\left( \frac{\partial^2 \ln L(\beta \mid S_n)}{\partial\beta\partial\beta'} \right) \right|^{1/2} = |A|^{1/2}$$

where

$$A \equiv \sum_{i=1}^{n} \frac{\phi_i^2}{D_i} x_i x_i',$$

with $\phi_i \equiv \phi(x_i'\beta)$ and $D_i \equiv \Phi_i(1 - \Phi_i)$, where $\Phi_i \equiv \Phi(x_i'\beta)$. Using some well known properties of matrix differentiation (e.g., Balestra, 1976), we obtain the logarithmic first derivative

$$\frac{\partial \ln |A|^{1/2}}{\partial\beta} = \frac{1}{2}((\operatorname{vec} A^{-1})' \otimes I_k)\operatorname{vec}\left(\frac{\partial A}{\partial\beta}\right).$$

Let

$$M \equiv \operatorname{vec} I_k \otimes I_k,$$

then we can write, using the product rule for matrices, the second derivative as

$$\begin{aligned}\frac{\partial^2 \ln |A|^{1/2}}{\partial\beta\partial\beta'} = \frac{1}{2}\Bigg[ &((\operatorname{vec} A^{-1})' \otimes I_k)\left( I_k \otimes \frac{\partial^2 A}{\partial\beta\partial\beta'} \right.\\ &\left.\left.\left. - M'(I_k \otimes ((A^{-1} \otimes I_k) \frac{\partial A}{\partial\beta} A^{-1} \frac{\partial A}{\partial\beta'} \right)\right)\right) \Bigg] M.\end{aligned}$$

Finally, to complete the derivatives we need an expression for $\partial A/\partial\beta$ and $\partial^2 A/\partial\beta\partial\beta'$, which are given by

$$\frac{\partial A}{\partial\beta} = -\sum_{i=1}^{n}\left(\frac{2x_i'\beta\phi_i^2}{D_i} + \frac{\phi_i^3(1-2\Phi_i)}{D_i^2}\right)(x_i x_i' \otimes x_i)$$

$$\frac{\partial^2 A}{\partial\beta\partial\beta'} = \sum_{i=1}^{n}\left(\frac{2\phi_i^2(2(x_i'\beta)^2-1)}{D_i} + \frac{5x_i'\beta\phi_i^3(1-2\Phi_i)+2\phi_i^4}{D_i^2} + \frac{2\phi_i^4(1-2\Phi_i)^2}{D_i^3}\right).$$

$$\cdot (x_i x_i' \otimes x_i x_i'),$$

which enables us to calculate the derivatives of Jeffreys' prior.

# 'NEW AND IMPROVED' DIRECT MARKETING: A NON-PARAMETRIC APPROACH

Jeffrey S. Racine

## ABSTRACT

*In this paper we consider a recently developed non-parametric econometric method which is ideally suited to a wide range of marketing applications. We demonstrate the usefulness of this method via an application to direct marketing using data obtained from the Direct Marketing Association. Using independent hold-out data, the benchmark parametric model (Logit) correctly predicts 8% of purchases by those who actually make a purchase, while the non-parametric method correctly predicts 39% of purchases. A variety of competing estimators are considered, with the next best models being semiparametric index and Neural Network models both of which turn in 36% correct prediction rates.*

## 1. INTRODUCTION

Direct marketing is one of the myriad of ways in which firms attempt to get the highest return from their marketing dollar. This is achieved by targeting individuals who, on the basis of observable characteristics such as demographics and their past purchase decisions, are most likely to be repeat customers. For example, one might think of mailing catalogs only to those who are highly

**Econometric Models in Marketing, Volume 16, pages 141–164.**

**ISBN: 0-7623-0857-5**

likely to be repeat customers or who most 'closely resemble' repeat customers.[1] The success or failure of direct marketing, however, hinges directly upon the ability to identify those consumers who are most likely to make a purchase. This 'identification' typically takes the form of the statistical modeling of the 'likelihood' or, strictly speaking, the *probability* that an individual makes a positive purchase decision.

*Parametric* methods constitute the traditional statistical approach towards modeling the likelihood of a purchase decision. Simply put, parametric methods require one to specify the functional form of the model *prior to estimation*. However, the functional form which generated the observed data is unknown. If the parametric model you use is 'close to the truth' (i.e. is functionally close to the *unknown* process that generated the observed data) then your predictions will be good; however, if you choose an inappropriate parametric model then your predictions may be no better than an unconditional guess. Common parametric approaches towards modeling a purchase decision would include the Logit and Probit models, for example, and these models are often referred to as 'parametric index models'.

Recent *semiparametric* developments such as the semiparametric single-index model of Ichimura (1993) (see also the related papers by Ichimura & Lee, 1991 and Ichimura & Thompson, 1998) extend parametric index models by combining a parametric index function with a data-driven estimator of the probability function, while earlier semiparametric index-based methods such as Manski's (1975) *Maximum Score* method take a robust 'least-absolute-deviation' approach to this problem. While more flexible than their fully parametric counterparts, the need to specify a parametric component leaves semiparametric approaches susceptible to the same critique that is levied against fully parametric approaches.

*Non-parametric* methods, on the other hand, allow the data themselves to fully determine the model. They make fewer assumptions and are more complex than their parametric counterparts, while they typically require more data if they are to attain the same degree of precision as a *correctly specified* parametric model. However, when non-parametric methods are applied to a number of datasets they often yield better predictions than those obtained from commonly applied parametric models which is simply a reflection of the presence of some degree of parametric misspecification.

Unfortunately, traditional non-parametric methods do not handle categorical variables in a satisfactory manner, and marketing databases frequently contain a mix of categorical and continuous variables.[2] Fortunately, recent developments in non-parametric methods permit one to handle categorical variables in a natural manner and frequently beat common parametric models when gauged by their predictive ability again reflecting the presence of some degree of parametric misspecification.

In this paper we assess the predictive performance of a variety of models which are commonly used to predict consumer choice relative to the recently developed non-parametric method of Racine and Li (2001). This latter non-parametric approach allows one to *directly* model the purchase decision probability without requiring specification of a parametric model, while it admits mixed data types and is fully data-driven. The comparison group of models includes parametric, semiparametric, and also Neural Network models.

We use an industry-standard database obtained from the Direct Marketing Association.[3] This database contains data on a reproduction gift catalog company, "an upscale gift business that mails general and specialized catalogs to its customer base several times each year". The base time period covers the period December 1971 through June 1992. Data collected included orders, purchases in each of fourteen product groups, time of purchase, and purchasing methods. Then a three month 'gap' occurs in the data after which customers in the existing database are sent at least one catalog in early Fall 1992. Then from September 1992 through December 1992 the database was updated. This provides an ideal database on which models can be constructed for the base time period and then evaluated on the later time period. We randomly select 4,500 individuals from the first time period, and we focus on predicting the likelihood of a consumer purchase using a variety of modeling strategies. We then evaluate the predictions of the various models on the *independent* hold-out sample consisting of 1,500 randomly selected individuals drawn from the later time period. The use of separate estimation and evaluation datasets permits us to gauge the predictive ability of each of the various approaches. We demonstrate how the new non-parametric econometric method is capable of outperforming a variety of methods which have been used to model consumer choice thereby enabling firms to get the highest possible return from their marketing dollar.

Parametric models for the prediction of binary outcomes have been applied in the marketing literature for such things as the prediction of brand choice (Bunch & Batsell, 1989) and the testing for market structure (Kannan & Wright, 1991). In line with the application considered in this paper, Bult (1993) has considered semiparametric classification models using Manski's (1975) approach and has assessed their performance in the context of direct marketing, while related issues such as profit maximization in a direct marketing framework are addressed in Bult and Wansbeek (1995), Gönül and Shi (1998) and Muus, van der Scheer and Wansbeek (2001), and the references therein.

The success or failure of any approach towards profit maximization depends on ones' ability to identify those most likely to make purchases, hence the approach considered herein has important implications for such lines of inquiry.

The remainder of the paper proceeds as follows. Section 2 outlines the conventional parametric models which are often used to predict consumer choice, while Sections 3–6 outline the semiparametric, Neural Network, and nonparametric estimators applied in this paper. Section 7 contains a discussion of the pros and cons of parametric vs. non-parametric approaches. Section 8 presents the details of the application including descriptions of the two datasets, summaries of model performance, and a discussion of the results, while Section 9 concludes with a brief discussion of the broader utility of the non-parametric approach considered herein. Estimation summaries for all models appear in the appendices.

## 2. PARAMETRIC MODELS

We briefly summarize two common parametric models which are frequently used to predict consumer choice. For an excellent survey on this literature we direct the interested reader to Ameniya (1981) and McFadden (1984) and the references therein. Of course, those familiar with these models may wish to skip this section and proceed directly to Section 6.

Let $Y \in \mathbb{R}$ be a random variable whose outcome will be conditioned on the random variables $X' = (X_1, \ldots, X_p) \in \mathbb{R}^p$. For the present case $Y$ represents the purchase decision and $X$ the observed factors that may influence this decision. Interest lies in the conditional prediction of $Y$ where

$$Y = \begin{cases} 1 & \text{if a purchase is made,} \\ 0 & \text{otherwise.} \end{cases}$$

We define the conditional probabilities associated with $Y$ as

$$Pr[Y=1 \mid X] = F(X, \beta)$$
$$Pr[Y=0 \mid X] = 1 - F(X, \beta),$$

where $F(\cdot)$ is a particular *parametric* distribution function and $\beta$ a vector of unknown parameters. This model is often summarized by the probability function

$$f(y) = F(x, \beta)^y (1 - F(x, \beta))^{1-y}, \qquad y \in \{0, 1\},$$

where $y$ and $x$ are realizations of the random vectors $Y$ and $X$ respectively.

The standard model for this setting is the binomial probability model which is often called a *binary choice model* since the dependent variable takes on only two values (non-purchase/purchase). The binomial probability model is usually written as

$$\begin{aligned} y_i &= E[y_i \mid x_i] + (y_i - E[y_i \mid x_i]) \\ &= F(x_i, \beta) + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \end{aligned} \tag{1}$$

where the subscript $i$ denotes a particular individual of which we observe $n$ and where $E[\cdot \mid \cdot]$ denotes a conditional expectation.

The parametric approach to estimating such probability models requires that one specify both the probability function $F(\cdot)$ and the nature of the relationship between the variables $X$, the parameters $\beta$, and the variable being predicted, $Y$. The typical parametric specification is linear and additive in $X$, hence we express this as the linear combination given by the scalar 'index' $X'\beta \in \mathbb{R}$, while the most popular probability functions are the 'Gaussian' ('Normal') leading to the 'Probit' model and the 'Logistic' leading to the 'Logit' model.

The Probit model uses the normal cumulative distribution function defined as

$$F(x_i'\beta) = \int_{-\infty}^{x_i'\beta} f(t)\, dt,$$

where $F(x_i'\beta)$ represents the normal cumulative distribution function (CDF) and where

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2},$$

which is the standard normal density function.

The Logit model, on the other hand, uses the Logistic distribution given by

$$F(x_i'\beta) = \frac{1}{1 + e^{-x_i'\beta}}. \tag{2}$$

These models are typically estimated via the method of 'maximum likelihood'. Having estimated the model's parameters which we denote by $\hat{\beta}$, we can then generate a prediction of the probability that an unseen individual with characteristics $x_i^0$ will make a purchase by simply computing $F(x_i^{0\prime}\hat{\beta})$ which represents the *estimated probability* that individual $i$ makes a purchase. If $F(x_i^{0\prime}\hat{\beta}) > 0.5$ then the model predicts that it is more likely than not that a purchase will be made, and of course one may well use a higher decision threshold if so desired.

## 3. SEMIPARAMETRIC INDEX MODELS

As was the case for the Probit and Logit models, Ichimura's (1993) semiparametric single-index estimator is also based on the model outlined in Eq. (1), though at first glance the estimator models a *conditional expectation* and not a *conditional probability*. However, for the application at hand where $Y \in \{0, 1\}$, the conditional expectation ($E[y_i | x_i, \beta]$) is also *equal* to the conditional probability ($F(x_i, \beta)$) since, by definition, for discrete outcomes

$$\begin{aligned} E[y_i | x_i, \beta] &= 0 \times Pr[y_i = 0 | x_i] + 1 \times Pr[y_i = 1 | x_i] \\ &= 0 \times (1 - F(x_i, \beta)) + 1 \times F(x_i, \beta) \\ &= F(x_i, \beta). \end{aligned} \tag{3}$$

Motivated in part by the desire to reduce the dimensionality of the conditioning information, Ichimura (1993) proceeds by assuming that $F(x_i, \beta) = F(x_i'\beta)$ as is done for linear index parametric models.[4] Motivated also by the non-parametric literature on the estimation of conditional expectations (see Pagan and Ullah (1999, Ch. 6)), Ichimura (1993) proposes estimating $E[y_i | x_i'\beta]$ ($F(x_i'\beta)$) by kernel methods. The kernel estimator is given by

$$\hat{E}[y | x'\beta] = \frac{\sum_{i=1}^{n} y_i K(x_i'\beta, x'\beta, h)}{\sum_{i=1}^{n} K(x_i'\beta, x\beta, h)} \tag{4}$$

where $K(\cdot)$ is simply a function that satisfies particular 'regularity' conditions such as that in Eq. (9) below.

Having estimated the semiparametric index model defined in Eq. (4), we observe that, in this setting, $\hat{E}[y | x^{0\prime}\beta]$ is a non-parametric probability estimate which is again analogous to $F(x_i^{0\prime}\hat{\beta})$ outlined in Section 2, and again, as in Section 2, if $\hat{E}[y | x^{0\prime}\beta] > 0.5$ then the model predicts that it is more likely than not that a purchase will be made.

For the application in Section 8, we shall use the cross-validation approach outlined below to select both the smoothing parameter $h$ and the index coefficients $\beta$ (see Section 6 for details).

## 4. SEMIPARAMETRIC MAXIMUM SCORE MODELS

Bult (1993) considered the performance of a parametric Logit model relative to Manski's (1975) *Maximum Score* method, and is the first article in direct marketing that uses semiparametric methods. The maximum score approach considered by Bult (1993) uses a robust 'least-absolute-deviation' semiparametric index estimator. Assuming the existence of a linear index, $x_i'\beta$, the model assumes that

$$x_i'\beta = \begin{cases} >0 & \text{if the probability of a purchase } >0.5, \\ <0 & \text{otherwise.} \end{cases}$$

(see Bult (1993, p. 382)). The method involves first creating a new variable taking on the value $y_i^* = 1$ if a purchase is made and $y_i^* = -1$ otherwise, and then maximizing the score function defined as

$$S(\beta) = \sum_{i=1}^{N} S_i(\beta) = \sum_{i=1}^{N} y_i^* \operatorname{sgn}(x_i'\beta) \tag{5}$$

where $\operatorname{sgn}(z)$ takes on the value 1 if $z > 0$ and $-1$ otherwise. It can be seen that incorrectly classified predictions will yield $y_i^* \operatorname{sgn}(x_i'\beta) = -1$ while correctly classified ones will yield $y_i^* \operatorname{sgn}(x_i'\beta) = 1$, hence this score function explicitly attempts to maximize the number of correctly classified cases in the training sample and minimize the number of incorrectly classified ones subject to the limitations of the parametric index function $x_i'\tilde{\beta}$.

Having estimated the parameter vector $\tilde{\beta}$, we can then generate a prediction that an unseen individual with characteristics $x_i^0$ will make a purchase by simply computing $x_i^{0'}\tilde{\beta}$. If $\operatorname{sgn}(x_i^{0'}\tilde{\beta}) > 0$ then the model predicts that it is more likely than not that a purchase will be made.

Due to the well-known potential for this approach to become ensnared by the presence of local minima, $10^5$ restarts of the search algorithm were conducted based on different random parameter values in an attempt to avoid their presence.

## 5. NEURAL NETWORK MODELS

Neural Network models have been used to successfully model a wide range of phenomena (see, for example, White & Racine, 2001) for an application to modeling exchange rates and the references therein). As was the case for the semiparametric index model, we shall use the Neural Network to estimate the conditional mean defined in Eq. (3) which, as noted, coincides with the conditional probability in this instance. We consider a single hidden layer feedforward network with $u$ hidden neurons in the 'middle' layer, $p$ inputs in the 'input' layer, and one output in the output layer which can be expressed as

$$f(x, \alpha, \beta) = \beta_u + \sum_{j=1}^{u} \beta_j \psi\left(\sum_{i=1}^{p} X_i \alpha_{ij} + \alpha_{pj}\right)$$

where $\psi(\cdot)$ is known as a 'transfer function' while $\alpha$ and $\beta$ are known as the 'network weights'.

For our application we use the Logistic CDF transfer function frequently found in applied work which is defined in Eq. (2). We refer the reader to Chen, Racine & Swanson (2001) and the references therein for examples of other transfer functions which are used in applied work. Letting $\omega = (\alpha, \beta)$, we train the network using a least-squares method by solving the supervised learning problem

$$\min_{\omega} \sum_{x} (y - f(x, \omega))^2, \tag{7}$$

and the weights which solve this problem are denoted by $\hat{\omega}$ (see White (1989) for further details). We select the appropriate number of hidden units $u$ using the Schwarz Information Criterion (SIC) (Schwarz (1978), Judge, Hill, Griffiths, Lütkepohl & Lee (1988, p. 848–849)).

Having determined the appropriate network architecture (i.e. number of hidden units, $u$) and having solved Eq. (7) to obtain the weights for this network, we can interpret $f(x^0, \hat{\omega})$ in this setting as a non-parametric probability estimate which is again analogous to $F(x_i^{0'}\hat{\beta})$ outlined in Section 2, and again, as in Section 2, if $f(x^0, \hat{\omega}) > 0.5$ then the model predicts that it is more likely than not that a purchase will be made.

## 6. NON-PARAMETRIC MODELS

The appeal of non-parametric estimation methods stems from the fact that they allow the data to model the relationships among variables, are robust to functional form specification, and have the ability to detect structure which sometimes remains undetected by traditional parametric estimation techniques. We briefly present an outline of the nonparametric estimator of a conditional density with mixed data types in a general framework as is done in Racine & Li (2001). The difference between the general estimator described below and that used in this paper will lie in the dimensionality of $Y$ which, for the application considered in this paper is a scalar, and in the specific data types found in the Direct Marketing Association database.

Let $Y'=(Y_1, \ldots, Y_q)\in\mathbb{R}^q$ be random variables whose outcomes will be conditioned on the random variables $X'=(X_1, \ldots, X_p)\in\mathbb{R}^p$. We let $q_d$ and $q_c$ denote the number of categorical and continuous variables in $Y$ respectively with $q_d+q_c=q$. We arrange the data with the $q_d$ categorical data types appearing first followed by the $q_c$ continuous ones so that $y_i'=(y_{i1}, \ldots, y_{iq})=(y_{i1}, \ldots, y_{iq_d}, y_{iq_d+1}, \ldots, y_{iq_d+q_c})$, with corresponding smoothing parameters $h_y'=(h_1^y, \ldots, h_q^y)=(h_1^y, \ldots, h_{q_d}^y, h_{q_d+1}^y, \ldots, h_{q_d+q_c}^y)$. Again we let $y'=(y_1, \ldots, y_q)=(y_1, \ldots, y_{q_d}, y_{q_d+1}, \ldots, y_{q_d+q_c})$ denote a vector-valued point at which an object is to be estimated. Finally, let $h_x'=(h_1^x, \ldots, h_p^x)=(h_1^x, \ldots, h_{p_d}^x, h_{p_d+1}^x, \ldots, h_{p_d+p_c}^x)$ be the smoothing parameters associated with $X$ and let $h'=(h_y', h_x')=(h_1, \ldots, h_{p_d+p_c+q_d+q_c})$.

A multivariate product kernel for the random variables $(Y, X)'=(Y_1, \ldots, Y_q, X_1, \ldots, X_p)$ consisting of mixed categorical and continuous data types would be given by

$$K(y_i, x_i, y, x, h)=K(y_i, x_i, y, x, h_y, h_x)$$

$$=\prod_{j=1}^{q_d} K(y_{ij}, y_j, h_j^y)\prod_{j=q_d+1}^{q_d+q_c} K(y_{ij}, y_j, h_j^y)\prod_{j=1}^{p_d} K(x_{ij}, x_j, h_j^x)\prod_{j=p_d+1}^{p_d+p_c} K(x_{ij}, x_j, h_j^x),$$

where the kernel functions appearing in the first and third products are categorical and those in the second and fourth products are continuous, while one for the random variables $X'=(X_1, \ldots, X_p)$ consisting of mixed categorical and continuous data types would be given by

$$K(x_i, x, h_x)=\prod_{j=1}^{p_d} K(x_{ij}, x_j, h_j^x)\prod_{j=p_d+1}^{p_d+p_c} K(x_{ij}, x_j, h_j^x), \tag{8}$$

where the kernel functions appearing in the first product are categorical and those in the second are continuous.

These kernel functions are simply functions that satisfy particular 'regularity' conditions. For unordered categorical variables we use the kernel function of Aitchison and Aitken (1976) given by

$$K(x_i, x_j, h)=\begin{cases} 1-h & \text{if } |x_i-x_j|=0, \\ \dfrac{h}{c-1} & \text{if } |x_i-x_j|\geq 1, \end{cases}$$

where $c$ is the number of 'categories' that $X$ can assume, while for ordered categorical variables we use the kernel that can be found in Ahmad and Cerrito (1994) given by

$$K(x_i, x_j, h) = \begin{cases} 1-h & \text{if } |x_i - x_j| = 0, \\ \dfrac{(1-h)}{2} h^{|x_i - x_j|} & \text{if } |x_i - x_j| \geq 1. \end{cases}$$

For continuous variables we use the Epanechnikov kernel function[5] given by

$$K(x_i, x_j, h) = \begin{cases} \dfrac{3}{h4\sqrt{5}}\left(1 - \dfrac{1}{5}\left(\dfrac{x_i - x_j}{h}\right)^2\right) & \text{if } \left(\dfrac{x_i - x_j}{h}\right)^2 < 5.0 \\ 0 & \text{otherwise.} \end{cases}$$

Letting $K(\cdot)$ be the kernel function defined in Eq. (8), the kernel estimator of the *conditional* probability density function (PDF) of $Y$ given $X$ denoted $f(y|x)$ is given by

$$\hat{f}(y|x) = \frac{\sum_{i=1}^{n} K(y_i, x_i, y, x, h_y, h_x)}{\sum_{i=1}^{n} K(x_i, x, h_x)},$$

with the same vector of smoothing parameters $h_x$ used in both the numerator and denominator. Properties of this estimator including rates of convergence and asymptotic normality can be found in Racine and Li (2001), while for a general recent treatment of a host of issues concerning non-parametric kernel estimators we highly recommend Pagan and Ullah (1999).

When $Y$ is a univariate binary variable as is the case in this paper, $\hat{f}(y|x)$ is simply an estimate of the probability associated with the realization $y$ in light of the observable characteristics $x$. That is, $\hat{f}(y|x)$ is a non-parametric probability estimate which is analogous to $F(x_i^{0\prime}\hat{\beta})$ outlined in Section 2. One difference is that, by construction, the parametric estimate $F(x_i^{0\prime}\hat{\beta})$ represents the probability that $Y=1$, while the non-parametric estimates $\hat{f}(0|x^0)$ measure the probability that $Y=0$ and $\hat{f}(1|x)$ the probability that $y=1$. But again, as was the case in Section 2, if $\hat{f}(1|x^0) > 0.5$ then it is more likely than not that a purchase will be made.

It is well known in the non-parametric literature that one can use *any* kernel function satisfying the required regularity conditions, while the choice of the smoothing parameters is *the* crucial factor underlying the estimator's performance.

### *6.1. Smoothing Parameter Selection*

The judicious selection of the smoothing parameters is the most important factor underlying the estimator's performance. We elect to use a fully automatic method of smoothing parameter selection that has a number of desirable features. This is the so-called 'cross-validation' method. Essentially, cross-validation methods are used to select models which should perform well on unseen data. For the case where all variables are continuous, the interested reader is referred to Silverman (1986, p. 52) and the references therein, while for mixed data types the interested reader is referred to Racine and Li (2001) and the references therein.

## 7. PARAMETRIC vs. NON-PARAMETRIC METHODS: DISCUSSION

Non-parametric kernel-based techniques do not presume that one knows or can correctly guess the unknown functional form of the object being estimated, in this case a conditional probability. Rather than presuming that the functional form of this object is known up to a few unknown parameters, we instead substitute less restrictive assumptions such as existence and 'smoothness' for the assumption that the parametric form of, say, a density function is known and equal to, say, $(2\pi\sigma^2)^{-1/2}\exp(-(x-\mu)^2/2\sigma^2)$ which happens to be one of the parametric assumptions underlying the Probit model given in Section 2. The advantage of non-parametric methods is that they are 'consistent', meaning that, as your 'information' (data) grows, you will continue to get closer to the true model. This feature is not shared by common parametric models, that is, if your parametric model is *incorrect* then no amount of data will overcome this deficiency. Of course, if you know the functional form up to a few unknown parameters (say, $\mu$ and $\sigma^2$) then you will *always* do better by using parametric techniques. However, in practice these forms are rarely if ever known, and the unforgiving consequences of parametric misspecification are well known having been mentioned in Section 1.

Since non-parametric techniques assume that less is known about the object of interest being estimated, they are therefore slower to converge to the unknown object being estimated than a *correctly specified* parametric model. However, as is the case here, it is often surprising how non-parametric approaches reveal structure in the data which is missed when one uses popular parametric specifications found in the applied literature. Non-parametric kernel methods are therefore best suited to situations in which one knows nothing about the functional form of the object being estimated, and the researcher is not constrained by a limited number of data points, i.e. an unreasonably small sample. Both of these conditions are met by a variety of marketing databases, hence this would appear to be an almost ideal setting for the successful application of non-parametric methods.

## 8. 'NEW AND IMPROVED' DIRECT MARKETING

### 8.1. Data Description

We have two *independent* estimation and evaluation datasets of sizes $n_1 = 4{,}500$ and $n_2 = 1{,}500$ respectively having one record per customer. We restrict our attention to one product group and thereby select the middle of the fourteen product groups, group eight. The variables involved in the study are listed below, while their properties are summarized in Tables 1 and 2.

***Table 1.*** Summary of the Estimation Dataset ($n_1 = 4{,}500$).

| Variable | Mean | Std Dev | Min | Max |
|---|---|---|---|---|
| Response | 0.09 | 0.28 | 0 | 1 |
| LTDFallOrders | 1.36 | 1.38 | 0 | 15 |
| LastPurchSeason | 1.62 | 0.53 | −1 | 2 |
| Orders4YrsAgo | 0.26 | 0.55 | 0 | 5 |
| LTDPurchGrp8 | 0.09 | 0.31 | 0 | 4 |
| DateLastPurch | 37.31 | 27.34 | 0 | 117 |

***Table 2.*** Summary of the Evaluation Dataset ($n_2 = 1{,}500$).

| Variable | Mean | Std Dev | Min | Max |
|---|---|---|---|---|
| Response | 0.08 | 0.27 | 0 | 1 |
| LTDFallOrders | 1.32 | 1.38 | 0 | 14 |
| LastPurchSeason | 1.63 | 0.51 | −1 | 2 |
| Orders4YrsAgo | 0.25 | 0.52 | 0 | 4 |
| LTDPurchGrp8 | 0.08 | 0.29 | 0 | 3 |
| DateLastPurch | 36.44 | 26.95 | 0 | 116 |

(1) Response – whether or not a purchase was made
(2) LTDFallOrders – life-to-date Fall orders
(3) LastPurchSeason – the last season in which a purchase was made[6]
(4) Orders4YrsAgo – orders made in the latest five years
(5) LTDPurchGrp8 – life-to-date purchases
(6) DateLastPurch – the date of the last purchase[7]

A quick look at Tables 1 and 2 reveals that both the estimation and evaluation datasets are similar in terms of their moments. For example, 8% of individuals in the evaluation dataset make a purchase which is similar to the 9% for those in the estimation dataset. It is important to note that the data summarized in Table 2 is not used to estimate models, therefore a comparison of the performance of various models on the evaluation data is *exactly* the comparison that is of interest to practitioners.

### *8.2. An Unconditional 'Benchmark' Model*

We begin with an 'unconditional' model in which we simply examine predictions based on the unconditional purchase probabilities for the estimation data. This model will serve as a benchmark by which we can assess the value-added by the 'conditional' approaches outlined in Sections 2–6.

Unconditionally, the likelihood that an individual makes a purchase is 8.8%, so the unconditional prediction for an individual drawn at random from the evaluation dataset would be that they *would not make a purchase* which would yield a correct prediction 92.2% of the time. The 'confusion matrix'[8] for this unconditional model is given in Table 3. We also report the measure of predictive performance suggested by McFadden et al. (1977) which was also

***Table 3.*** Confusion Matrix and Classification Rates for the Unconditional Model.

| | Predicted Non-purchase | Predicted Purchase |
|---|---|---|
| Actual Non-purchase | 1383 | 0 |
| Actual Purchase | 117 | 0 |
| Predictive Performance: (McFadden et al. (1977)) | | 91.59% |
| CCR: Overall correct classification rate | | 92.20% |
| CCR(N): Correct non-purchase classification rate | | 100.00% |
| CCR(P): Correct purchase classification rate | | 0.00% |

analyzed by Veall and Zimmermann (1992} and found to have good predictive performance defined as $p_{11}+p_{22}-p_{21}^2-p_{12}^2$ where $p_{ij}$ is the $ij$th entry in the $2\times 2$ confusion matrix expressed as a fraction of the sum of all entries.

As can be seen from Table 3, though the model might appear to be doing well according to the overall classification rate (CCR = 92.2%), this is not what a direct marketer would be interested in. Unconditionally (i.e. ignoring the explanatory variables 6-5), if you were to make a guess about whether or not a given individual would make a purchase, you would guess that they would not, that is, you would predict *zero sales!* Of course, a direct marketer would be primarily interested in the accurate prediction of those *who actually make a purchase* which is given by the 'Actual Purchase' row (117 0). That is, in addition to the overall correct classification rate, one is interested in the question "how many of our actual customers did we foresee making purchases?" Given that the diagonal element of this row (0) tells us that we do not *correctly* predict a single purchase and the off-diagonal element (117) tells us that we *incorrectly* predicted every actual purchase, then this conditional model is of limited practical use. That is, though this unconditional model has a 92.2% overall classification rate, it has a 'correct classification rate' of 0% for purchases (CCR(P) = 0%).

Conditional models, on the other hand, make use of variables 6-5 when forming their predictions. We expect, if the conditional models are adding value, that they would not only have higher overall correct classification rates, but would of course also have higher correct classification rates. We therefore begin with standard *parametric* models which have often been applied to the prediction of consumer purchases.

### *8.3. Logit and Probit Parametric Models*

The Logit and Probit models outlined in Section 2 are perhaps the most widely used models for the prediction of categorical outcomes such as consumer purchases. The within-sample parameter estimates and summary information for each model can be found in Appendix A, while their confusion matrices are presented in Tables 4 and 5.

As can be seen from examining Tables 4 and 5, these models fare better than the unconditional model in terms of their overall correct classification rates and their correct purchase classification rates, but perhaps not by as much as one might have expected. In particular, the Probit model correctly predicts only 3 out of 117 purchases (2.6%) while the Logit model correctly predicts 9 out of

***Table 4.*** Confusion Matrix and Classification Rates for the Probit Model.

| | Predicted Non-purchase | Predicted Purchase |
|---|---|---|
| Actual Non-purchase | 1383 | 0 |
| Actual Purchase | 114 | 3 |
| Predictive Performance: (McFadden et al. (1977)) | | 91.82% |
| CCR: Overall correct classification rate | | 92.40% |
| CCR(N): Correct non-purchase classification rate | | 100.00% |
| CCR(P): Correct purchase classification rate | | 2.56% |

117 purchases (7.7%). We now examine the predictive ability of a semiparametric index model.

### *8.4. Semiparametric Index Models*

We apply the approach of Ichimura (1993) outlined in Section 3 and present the confusion matrix in Table 6. Summary estimation information can be found in Appendix A.3.

As can be seen from examining Table 6, this model fares much better than the Probit and Logit models in terms of its overall and correct purchase classification rates correctly predicting 42 out of 117 purchases (35.9%). model.

### *8.5. Semiparametric Maximum Score Models*

We apply the maximum score model outlined in Section 3. Summary estimation information can be found in Appendix A.4, while Table 7 presents the confusion matrix.

***Table 5.*** Confusion Matrix and Classification Rates for the Logit Model.

| | Predicted Non-purchase | Predicted Purchase |
|---|---|---|
| Actual Non-purchase | 1378 | 1 |
| Actual Purchase | 108 | 9 |
| Predictive Performance: (McFadden et al. (1977)) | | 91.95% |
| CCR: Overall correct classification rate | | 92.47% |
| CCR(N): Correct non-purchase classification rate | | 99.64% |
| CCR(P): Correct purchase classification rate | | 7.69% |

***Table 6.*** Confusion Matrix and Classification Rates for the Semiparametric Index Model.

| | Predicted Non-purchase | Predicted Purchase |
|---|---|---|
| Actual Non-purchase | 1361 | 22 |
| Actual Purchase | 75 | 42 |
| Predictive Performance: ((McFadden et al. (1977)) | | 93.26% |
| CCR: Overall correct classification rate | | 93.53% |
| CCR(N): Correct non-purchase classification rate | | 98.41% |
| CCR(P): Correct purchase classification rate | | 35.90% |

As can be seen from examining Table 7, this semiparametric maximum score model turns in a rather mixed performance. Its overall classification rate is the lowest among all models considered; however, it does a decent job of correctly predicting purchases.

### *8.6. Neural Network Models*

We apply the Neural Network outlined in Section 5. Summary estimation information can be found in Appendix A.5, while Table 8 presents the confusion matrix.

As can be seen from examining Table 8, this model fares a bit worse than the semiparametric index model summarized in Table 6 as it has a lower overall correct classification rate, though it clearly performs much better than the

***Table 7.*** Confusion Matrix and Classification Rates for the Maximum Score Model.

| | Predicted Non-purchase | Predicted Purchase |
|---|---|---|
| Actual Non-purchase | 1342 | 41 |
| Actual Purchase | 77 | 40 |
| Predictive Performance: ((McFadden et al. (1977)) | | 91.80% |
| CCR: Overall correct classification rate | | 92.13% |
| CCR(N): Correct non-purchase classification rate | | 97.04% |
| CCR(P): Correct purchase classification rate | | 34.19% |

**Table 8.** Confusion Matrix and Classification Rates for the Neural Network Model.

| | Predicted Non-purchase | Predicted Purchase |
|---|---|---|
| Actual Non-purchase | 1356 | 27 |
| Actual Purchase | 75 | 42 |
| Predictive Performance: (McFadden et al. (1977)) | | 92.92% |
| CCR: Overall correct classification rate | | 93.20% |
| CCR(N): Correct non-purchase classification rate | | 98.05% |
| CCR(P): Correct purchase classification rate | | 35.90% |

parametric models. Finally, we turn to a non-parametric kernel model to see whether or not moving to a fully non-parametric framework can further improve upon our ability to predict consumer purchases.

### 8.7. Non-parametric Models

We apply the non-parametric estimator outlined in Section 6. Table 9 presents the confusion matrix, and summary estimation information can be found in Appendix A.6.

As can be seen from examining Table 9, this new non-parametric method has a higher overall and correct purchase classification rate than any of the competing approaches, and it correctly predicts 46 out of 117 (39.3%) of purchases.

**Table 9.** Confusion Matrix and Classification Rates for the Non-parametric Model.

| | Predicted Non-purchase | Predicted Purchase |
|---|---|---|
| Actual Non-purchase | 1358 | 25 |
| Actual Purchase | 71 | 46 |
| Predictive Performance: (McFadden et al. (1977)) | | 93.35% |
| CCR: Overall correct classification rate | | 93.60% |
| CCR(N): Correct non-purchase classification rate | | 98.19% |
| CCR(P): Correct purchase classification rate | | 39.32% |

### 8.8. *Discussion*

An examination of the performance of the parametric, semiparametric, and non-parametric approaches (Tables 4–9) reveals the following ranking of models in terms of their out-of-sample performance based upon the measure of McFadden et al. (1977) arranged from highest to lowest (correct purchase classification rates appear in parentheses);

(1) Non-parametric (39.3%)
(2) Semiparametric Index Model (35.9%)
(3) Neural Network Model (35.9%)
(4) Logit Model (7.7%)
(5) Probit Model (2.6%)
(6) Semiparametric Maximum Score Model (34.2%)

For this application it is clear that the parametric models lag far behind the semiparametric index and non-parametric models in terms of their out-of-sample performance, thus a few words on the specification of parametric models are in order. In Section 1 we emphasized the fact that no model can outperform a *correctly specified* parametric model. However, locating the *correct* parametric model for a given dataset remains an unsettled art, and we refer the interested reader to Leamer (1978) and Manski and McFadden (1986) for further discussion. We wish to be clear that *in no way are we claiming that non-parametric estimators will always outperform parametric models*. However, as we demonstrate here, non-parametric estimators may often outperform *common* parametric specifications which clearly reflects the presence of some degree of misspecification of the parametric models. Does a better parametric model exist? Almost certainly! How does one select the correct parametric model? Suffice it to say at this point that this issue remains unsettled, and in this light the appeal of non-parametric approaches is clear.

The semiparametric index model performs better than the Logit and Probit models as expected as they all share a common linear index function, but the semiparametric model is more flexible in terms of the the assumptions that it makes regarding $Pr[Y=1|X]$ (the conditional expectation in this case). The semiparametric index model performs better than the Neural Network model as it has a higher overall classification rate but the same correct classification rate. The non-parametric model, however, turns in the strongest performance in terms of its predictive performance (McFadden et al., 1977), overall correct classification rate, and its correct purchase classification rate, all of which exceed those for *all* models considered. The semiparametric maximum score model, which has been used to model direct marketing (Bult, 1993) does not turn in a strong performance in this setting.

Linear-index parametric approaches such as the Logit and Probit models remain the most commonly applied statistical methods in this setting, but we hope that the reader is convinced that the performance of a wide range of estimators including semiparametric index, Neural Network, and non-parametric approaches admits them as appealing alternatives for the prediction of consumer choice.

We emphasize that many marketing application are ideally suited to non-parametric analysis. This is so since, first, we generally have no priors on functional forms that have generated consumer choice which places parametric methods at a bit of a disadvantage. Second, marketing databases often contain an abundance of data, and non-parametric methods are in their element when this is the case. Therefore, it is not surprising that, in this setting, we can outperform standard parametric methods, while it may be surprising to some that we even outperform powerful semiparametric and Neural Network methods when judged by their *out-of-sample* predictive ability.

## 9. CONCLUSION

We apply recently developed non-parametric methods to the prediction of consumer purchase behavior. It is seen how the new methods can result in significantly improved *out-of-sample* purchase prediction relative to standard parametric, semiparametric, and Neural Network methods. A few words on the potential applicability of this approach are in order. It is evident that any profit maximization strategy on the part of a direct marketer hinges on the ability to identify those most likely to make a purchase. Bult and Wansbeek (1995) consider how such predictions could be instrumental for profit maximization, while Gönül and Shi (1998) consider how such predictions could be used in a utility/profit maximization framework for potential customers/firms. Though such an exercise lies beyond the scope of the current paper, the non-parametric approach used herein may prove highly valuable in these settings.

However, the non-parametric methods have a broader utility for the marketing community than even that presented herein and may be valuable in a wide range of settings not addressed in this paper. It is our sincere hope that these semiparametric and non-parametric methods spark the curiosity of all those interested in obtaining the highest possible return from their marketing dollar.

## NOTES

1. Bult and Wansbeek (1995), in a profit maximization framework, point out that in fact one might want to do the opposite thereby saving costs by avoiding repeated mailings to those who in fact are highly likely to make a purchase. Regardless of the objective, it is the ability to identify those most likely to make a purchase that has proven problematic in the past and is the focus of this paper.

2. Examples of categorical variables would include preferences (like, indifferent, dislike), purchase decisions (buy, don't buy), number of children and so on, while examples of continuous variables would include income, net wealth and the like.

3. This database contains customer buying history for about 100,000 customers of nationally known catalog and non-profit database marketing businesses.

4. Ichimura (1993) considers index functions of a general nature, but in practice the linear index is almost universally applied, hence we adopt this index specification for what follows.

5. Note that we subsume the multiplicative (inverse) bandwidth $1/h$ in the definition of the kernel function itself.

6. This is recorded in the database as 1 if the purchase was made in January through June, 2 if the purchase was made in July through December, and −1 if no purchase was made.

7. 12/71 was recorded as '0', 1/72 as '1' and so on.

8. A 'confusion matrix' is simply a tabulation of the actual outcomes versus those predicted by a model. The diagonal elements contain correctly predicted outcomes while the off-diagonal ones contain incorrectly predicted (confused) outcomes. A method that performs well relative to another could be detected by examining their respective confusion matrices; (i) the better performer would have a stronger diagonal (sum of the diagonal elements would be higher) and therefore a higher overall correct classification rate (CCR), and (ii) the better performer would in addition be expected to show more 'balance' in both the diagonal and off-diagonal elements.

## ACKNOWLEDGMENTS

I would like to thank but not implicate the editors of this volume and an anonymous referee whose sage advice and careful reading were extremely helpful. I would also like to thank Richard Kim and Jerry Goolsby for their helpful comments.

## REFERENCES

Ahmad, I. A., & Cerrito, P. B. (1994). Non-parametric estimation of joint discrete-continuous probability densities with applications. *Journal of Statistical Planning and Inference*, *41*, 349–364.

Aitchison, J., & Aitken, C. G. G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika*, *63*(3), 413–420.

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature*, *19*, 1483–1536.

Bult, J. R. (1993). Semiparametric vs. parametric classification models: An application to direct marketing. *Journal of Marketing Research, 30*, 380–390.

Bult, J. R., & Wansbeek, T. (1995). Optimal selection for direct mail. *Marketing Science, 14*(4), 378–394.

Bunch, D. S., & Batsell, R. R. (1989). A Monte Carlo comparison of estimators for the multinomial logit model. *Journal of Marketing Research, 26*, 56–68.

Chen, X., Racine, J., & Swanson, N. (2001). Semiparametric ARX neural network models with an application to forecasting information. *IEEE Transactions on Neural Networks, 12*(4), 674–684.

Gönül, F., & Shi, M. Z. (1998). Optimal mailing of catalogs: A new methodology using estimable structural dynamic programming models. *Management Science, 44*(9), 1249–1262.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics, 58*, 71–120.

Ichimura, H., & Lee, L. F. (1991), Semiparametric estimation of multiple index models. In: W. A. Barnett, J. Powell & G. Tauchen (Eds), *Non-parametric and Semiparametric Methods in Econometrics and Statistics* (pp. 3–50). Cambridge.

Ichimura, H., & Thompson, T. S. (1998). Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution. *Journal of Econometrics, 86*(2), 269–295.

Judge, G., Hill, R., Griffiths, W., Lütkepohl, H., & Lee, T. C. (1988). *Introduction to the Theory and Practice of Econometrics* (2nd ed.). John Wiley & Sons.

Kannan, P., & Wright, G. P. (1991). Modelling and testing structured markets: A nested logit approach. *Marketing Science, 10*, 58–82.

Leamer, E. E. (1978). *Specification Searches*. Wiley, New York.

Manski, C. F., & McFadden, D. (1986). *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge: MIT Press.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics, 3*, 205–228.

McFadden, D. (1984), Econometric analysis of qualitative response models. In: Z. Griliches & M. Intriligator (Eds), *Handbook of Econometrics* (pp. 1385–1457). North Holland.

McFadden, D., Puig, C., & Kerschner, D. (1977). Determinants of the long-run demand for electricity. Proceeedings of the American Statistical Association (Business and Economics Section), pp. 109–117.

Muus, L., van der Scheer, H., & Wansbeek, T. (2001). *A decision theoretic framework for profit maximization in direct marketing*. JAI Press.

Pagan, A., & Ullah, A. (1999). *Non-parametric Econometrics*. Cambridge: Cambridge University Press.

Racine, J., & Li, Q. (2001). *Non-parametric estimation of conditional distributions with both categorical and continuous data*. In submission.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Veall, M. R., & Zimmermann, K. F. (1992). Performance measures from prediction-realization tables. *Economics Letters, 39*, 129–134.

White, H. (1989). Learning in artificial neural networks: A statistical perspective. *Neural Computation, 1*, 425–464.

White, H., & Racine, J. (2001). Statistical inference, the bootstrap, and neural network modeling with application to foreign exchange rates. *IEEE Transactions on Neural Networks*, *12*(4), 657–674.

# APPENDIX: MODEL ESTIMATION SUMMARY

## *A.1. Probit Model Estimation Summary*

PROBIT ESTIMATION

| | |
|---|---|
| Number of observations = 4500.00 | R-squared = 0.126302 |
| Number of positive obs. = 393.000 | Kullback-Leibler R-sq = 0.148880 |
| Mean of dep. var. = 0.087333 | Log likelihood = −1134.93 |
| Sum of squared residuals = 313.821 | |
| Fraction of Correct Predictions = 0.913556 | |

| Parameter | Estimate | Standard Error | *t*-statistic | *P*-value |
|---|---|---|---|---|
| C | −0.640230 | 0.096342 | −6.64536 | [0.000] |
| X1 | 0.178449 | 0.019207 | 9.29101 | [0.000] |
| X2 | −0.315028 | 0.060194 | −5.23355 | [0.000] |
| X3 | 0.010375 | 0.053897 | 0.192492 | [0.847] |
| X4 | 0.011419 | 0.077655 | 0.147045 | [0.883] |
| X5 | −0.017793 | 0.144130E-02 | −12.3453 | [0.000] |

## *A.2. Logit Model Estimation Summary*

LOGIT ESTIMATION

| Choice | Frequency | Fraction | |
|---|---|---|---|
| 0 | 4107 | 0.9127 | (coefficients normalized to zero) |
| 1 | 393 | 0.0873 | |

| | |
|---|---|
| Number of observations = 4500.00 | $R$-squared = 0.135567 |
| Number of positive obs. = 393.000 | Kullback-Leibler $R$-sq = 0.155711 |
| Mean of dep. var. = 0.087333 | Log likelihood = −1125.82 |
| Sum of squared residuals = 310.353 | |
| Number of Choices = 9000 | |
| Fraction of Correct Predictions = 0.914444 | |

| Parameter | Estimate | Standard Error | t-statistic | P-value |
|---|---|---|---|---|
| C1 | –0.844080 | 0.1816090 | –4.64778 | [0.000] |
| X11 | 0.308135 | 0.0350990 | 8.77910 | [0.000] |
| X21 | –0.589361 | 0.1202130 | –4.90264 | [0.000] |
| X31 | 0.038611 | 0.105474 | 0.366068 | [0.714] |
| X41 | 0.021875 | 0.144698 | 0.151179 | [0.880] |
| X51 | –0.041433 | 0.345309E-02 | –11.9989 | [0.000] |

*A.3. Semiparametric Index Model Estimation Summary*

| | |
|---|---|
| Kernel function: | Second Order Epanechnikov Kernel |
| Number of observations: | 4500 |
| Smoothing parameter: | 0.9128 |
| Coefficient for regressor 1: | –0.0069 |
| Coefficient for regressor 2: | 4.0755 |
| Coefficient for regressor 3: | 0.0348 |
| Coefficient for regressor 4: | –0.0043 |
| Coefficient for regressor 5: | 2.0205 |
| MSE: | 0.0527 |
| MAE: | 0.1071 |

*A.4. Maximum Score Model Estimation Summary*

| | |
|---|---|
| Number of observations: | 4500 |
| Intercept | –0.0231 |
| Coefficient for regressor 1: | 0.1420 |
| Coefficient for regressor 2: | 0.9814 |
| Coefficient for regressor 3: | –0.0715 |
| Coefficient for regressor 4: | 0.1785 |
| Coefficient for regressor 5: | –0.3630 |

*A.5. Neural Network Model Estimation Summary*

| | |
|---|---|
| Number of inputs: | 5 |
| Number of training observations: | 4500 |
| Number of neurons (SIC-optimal): | 3 |
| Number of weights: | 22 |
| MSE: | 0.0527 |
| MAE: | 0.1055 |

*A.6. Non-parametric Model Estimation Summary*

| | |
|---|---|
| Kernel for ordered variables: | Ahmad & Cerrito Kernel |
| Number of observations: | 4500 |
| smoothing parameter 1: | 0.001 |
| smoothing parameter 2: | 0.547 |
| smoothing parameter 3: | 0.003 |
| smoothing parameter 4: | 0.529 |
| smoothing parameter 5: | 0.999 |
| smoothing parameter 6: | 0.012 |

# ESTIMATING MARKET-LEVEL MULTIPLICATIVE MODELS OF PROMOTION EFFECTS WITH LINEARLY AGGREGATED DATA: A PARAMETRIC APPROACH

Albert C. Bemmaor and Udo Wagner

## ABSTRACT

*Marketing researchers may be confronted with biases when estimating response coefficients of multiplicative promotion models based on linearly aggregated data. This paper demonstrates how to recover the parameters obtained with data which are aggregated in a compatible way with such models. It provides evidence that the geometric means of sales and of prices across stores can be predicted with accuracy from their arithmetic means and standard deviations. Employing these predictions in a market-level model results in parameter estimates which are consistent with those obtained with the actual geometric means and fairly close to coefficients derived at the individual store level.*

**Econometric Models in Marketing, Volume 16, pages 165–189.**

**ISBN: 0-7623-0857-5**

# 1. INTRODUCTION

Much recent research has dealt with the measurement of promotion effects on retail sales (see, e.g. Blattberg & Neslin, 1990). Typically, these studies are based on the analysis of store-level weekly sales data (see, e.g. Blattberg & Wisniewski, 1989; Montgomery, 1997). However, the client companies of marketing research firms such as A. C. Nielsen or IRI do not usually have access to store-level sales and promotion data but rather to cumulative data such as sales at the market level and mean prices across stores. These linearly aggregated data are incompatible with the estimation of multiplicative models of sales response because nonlinear relations postulated at the individual (store) level do not extend alike to more comprehensive levels (e.g. chain, market) if basically a summing-up procedure to create the new variables is employed. In the case of a multiplicative model, one would need the geometric means of the variables to obtain parameters which are in accordance with the store-level model.

Since such data are not available in general nor can be calibrated from sources at hand, market researchers frequently take on a more pragmatic point of view and postulate the same multiplicative model, e.g. the SCAN*PRO model (Wittink et al., 1988), at the *aggregate* level. The issue then arises to estimate the biases induced therefrom. Recently, Christen et al. (1997) analyzed these biases when the "true" model is a multiplicative model with constant slope coefficients across stores but with varying intercepts. They showed that the impact of promotion can be inflated substantially with market-level data, thereby falsely encouraging managers to run more promotions; they provided empirical evidence for these biases and developed a means to correct them based on simulations. The disadvantage of their correction method is the potential lack of generalizability.

Instead of carrying out an ex-post correction of the parameters, the study at hand suggests:

(i) *predicting* aggregated sales and price data which are compatible with the estimation of multiplicative store-level models; and
(ii) subsequently employing these variables to estimate the model's parameters at the market level.

This is accomplished on the basis of market-level data and additional descriptive information about the variability of sales and prices across cross-sections. Dealing with the aggregation problem in such a way, the contribution of this study is two-fold:

(1) It shows that the actual geometric means of variables can be predicted with accuracy from their arithmetic means and their corresponding standard deviations when they are lognormally distributed.
(2) It demonstrates that the parameter estimates of the response model obtained with the use of these predicted geometric means are consistent with those obtained when using actual geometric means.

Empirical evidence is provided by analyzing scanning data on six brands belonging to two product categories in two European countries.

The remainder of this article is organized as follows: The second section presents the issue and the proposed methodology, the third section reports the empirical analysis and the last one concludes the paper.

# 2. THE ISSUE

## *2.1. Relationships Between Multiplicative Models at Different Levels of Aggregation*

### *2.1.1. Model at the Store Level*

The issue of linear aggregation in a linear model has been discussed by Theil (1974). However, linear models have found limited applications in marketing; instead, marketing practitioners and academics have extensively used multiplicative models of sales response to price changes and promotional activities due to their enhanced face validity. The SCAN*PRO model (Wittink et al., 1988), for example, was developed by A. C. Nielsen to estimate price and promotion effects at the store level. In order to pinpoint the aggregation issue we make use of a simplified version of this model by concentrating on a single brand only, thereby neglecting potential cross-effects; further, we do not account for seasonality. Thus the model takes on the following form:

$$S_{kt} = \alpha_k \cdot P_{kt}^{\beta} \cdot \prod_{j=1}^{J} \gamma_j^{F_{jkt}} \qquad \forall k, t \tag{1}$$

$S_{kt}$: sales in store $k$ in week $t$;
$P_{kt}$: price relative to regular price in store $k$ in week $t$;
$F_{jkt}$: dummy variable which takes on a value of 1 when store $k$ runs merchandising support $j$ in week $t$ (e.g. display, feature) and 0 otherwise;
$\alpha_k$: parameter, varying across stores but constant over time, which may be interpreted as baseline sales in store $k$ ($\alpha_k > 0$);

$\beta, \gamma_j$: response parameters assumed constant across stores and time; $\beta$ may take on any value, but is expected to be negative; it is interpreted as a price elasticity.
$\gamma_j \geq 0$, but is expected to be greater than unity; it represents a sales increase due to promotional activity $j$.

For presentational convenience we will assume $J=1$ in the sequel and drop the index $j$ since the consideration of more than one promotional variable does not add to the complexity of the problem.

The SCAN*PRO model has been extensively applied (Bucklin & Gupta, 1999). It assumes that the price elasticity $\beta$ and the merchandising support multiplier $\gamma$ are constant across stores but that the intercepts $\alpha_k$ vary across stores. Store-level data are required for parameter calibration which might easily be carried out by means of a "pooled" regression analysis (over stores and time). For the subsequent discussion, we implicitly assume that (1) is the "true" model. We will compare this formulation with corresponding alternatives using aggregated data. Keep in mind, however, that when analyzing empirically observed markets the "true" model is unknown.

*2.1.2. Model at the Market Level with Arithmetic Means as Input Data*

From the manufacturer's point of view, the brand manager does not concentrate on a single store $k$ but rather on a more general level (e.g. chain or market). Furthermore, the data at hand usually do not provide detailed store information but have been aggregated by the provider by taking arithmetic means, i.e.:

$$\bar{S}_t = \frac{1}{K}\sum_{k=1}^{K} S_{kt}, \quad \bar{P}_t = \frac{1}{K}\sum_{k=1}^{K} P_{kt}, \quad \bar{F}_t = \frac{1}{K}\sum_{k=1}^{K} F_{kt} \qquad \forall t$$

where $K$ represents the total number of stores. (Instead of mean sales, cumulative sales may be reported as well; this would basically make no difference with respect to our derivations.) Such a procedure results in significant data reduction: Cross-sectional information is reduced to a single value of central tendency per period instead of looking at a variety of observations. Generalizing (1) to hold at market level and replacing the variables by their arithmetic means and the response parameters by $\alpha'$, $\beta'$, $\gamma'$, respectively, one ends up with:

$$\bar{S}_t = \alpha' \cdot \bar{P}_t^{\beta'} \cdot \gamma'^{\bar{F}_t} \qquad \forall t \qquad (2)$$

Looking at (2) from a marketing perspective, homogeneous parameters *and* homogeneous marketing activities across the different stores are postulated. Geweke (1985) called such a perspective the "representative agent approach", i.e. theory is developed at the micro unit level, model building and estimation are conducted with aggregate data.

Taking logarithms, (2) becomes:

$$\ln \bar{S}_t = \ln \alpha' + \beta' \cdot \ln \bar{P}_t + \bar{F}_t \cdot \ln \gamma' \qquad \forall t \tag{3}$$

Marketing researchers usually find (3) computationally very attractive since the response parameters may be estimated by an ordinary least-squares procedure. In more general terms, $\bar{S}_t$, $\bar{P}_t$ and $\bar{F}_t$ may be interpreted as an estimator of $E(S_t)$, $E(P_t)$, $E(F_t)$, respectively if we regard $S_t$, $P_t$ and $F_t$ as random variables varying across stores. Thus (3) can be written as:

$$\ln E(S_t) = \ln \alpha' + \beta' \cdot \ln E(P_t) + \bar{F}_t \cdot \ln \gamma' \qquad \forall t \tag{4}$$

As will be seen shortly, the aggregation problem arises only with respect to the variables which are subject to a logarithmic transformation; therefore we kept $\bar{F}_t$ in (4).

*2.1.3. Model at the Market Level with Geometric Means as Input Data*

Christen et al. (1997) show that the estimation of (2) is inconsistent with the estimation of (1), the reason being essentially due to *linear* aggregation. However, this problem can – at least in theory – be dealt with by using geometric means instead of arithmetic means in (2). In this case one calibrates the same model as in (1) but at a different level of aggregation, given the assumptions of constant response parameters across stores (for a formal proof and the relationship between the parameters ($\alpha_1^*, \ldots, \alpha_K - \alpha'$; $\beta - \beta'$; $\gamma - \gamma'$) see Leeflang et al., 2000, p. 274). Consequently we define:

$$g_{S_t} = \sqrt[K]{\prod_{k=1}^{K} S_{kt}}, \quad g_{P_t} = \sqrt[K]{\prod_{k=1}^{K} P_{kt}}, \qquad \forall t$$

and replace the arithmetic means in (2) by their corresponding geometric equivalents:

$$g_{S_t} = \alpha \cdot g_{P_t}^{\beta} \cdot \gamma^{\bar{F}_t} \qquad \forall t \tag{5}$$

Once again we take logarithms and arrive at:

$$\ln g_{S_t} = \ln \alpha + \beta \cdot \ln g_{P_t} + \bar{F}_t \cdot \ln \gamma \qquad \forall t \tag{6}$$

Arguing along the same lines as above, one finds

$$\ln g_{S_t} = \frac{1}{K} \sum_{k=1}^{K} \ln S_{kt}$$

and thus it may be seen as an estimator of $E(\ln S_t)$: the logarithm of the geometric mean of sales corresponds to the expected value of log sales. We define $E(\ln P_t)$ analogously and rewrite (6):

$$E(\ln S_t) = \ln \alpha + \beta \cdot E(\ln P_t) + \bar{F}_t \cdot \ln \gamma \qquad \forall t \tag{7}$$

*2.1.4. Comparison Between Models at the Market Level*

Starting from the same model at the store level, the preceding sub-sections presented the effects of different aggregation schemes. We now focus on the consequences resulting therefrom. In particular we compare (4) with (7) and put emphasis on the following:

- Relationship between $\ln E(X)$ and $E(\ln X)$
  In some sense we are confronted with a basic econometric problem, i.e. the relation between $\ln E(X)$ and $E(\ln X)$ (where $X$ denotes a random variable). It is well known (c.f. Judge et al., 1985, p. 147 f.) that these two expressions are not equal, in general.
- Errors in variables
  Since (7) is consistent with the "true" model (1), we therefore notice that (4) includes errors in both the dependent variable and one independent variable. This error can be approximated by means of a Taylor expansion. For example, in the case of $S_t$, we expand $E(\ln S_t)$ at $\bar{S}_t = E(S_t)$ and obtain:

$$E(\ln S_t) \approx \frac{1}{K} \sum_{k=1}^{K} \left[ \ln \bar{S}_t + \frac{S_{kt} - \bar{S}_t}{\bar{S}_t} - \frac{(S_{kt} - \bar{S}_t)^2}{2 \cdot \bar{S}_t^2} \right] = \ln E(S_t) - \frac{1}{2} \cdot CV_{S_t}^2 \tag{8}$$

  where $CV_{S_t}$ represents the coefficient of variation of sales (standard deviation/mean) in period $t$. Therefore, one would need a measure of the coefficient of variation of the variables across stores to obtain a more accurate prediction of the logarithm of the geometric mean.
- Identification problem
  We use (8) to illustrate another problem encountered when employing (4), i.e. arithmetic means, for estimation. Keeping in mind that (7) is consistent with the SCAN*PRO model (1), we replace $E(\ln X)$ by their approximations (8) and get:

$$\ln E(S_t) - \frac{1}{2} CV_{S_t}^2 \approx \ln \alpha + \beta \cdot \left(\ln E(P_t) - \frac{1}{2} CV_{P_t}^2\right) + \bar{F}_t \cdot \ln \gamma \qquad \forall t \tag{9}$$

  Hence, the comparison of (4) with (9) shows that the parameters of (9) are not identified when one has no information on the variation of the variables across stores; $CV_{S_t}$ and $CV_{P_t}$ change over time. This fact is very much in line with standard results of econometric theory with respect to errors-in-variables models (c.f. Judge et al., 1985, p. 735).

- Additional information required
  In order to predict $E(\ln S_t)$ and $E(\ln P_t)$ from their arithmetic means, one needs *additional information* such as the standard deviation. This will permit one to approximate $E(\ln S_t)$ and $E(\ln P_t)$ in the way outlined above. As an alternative procedure, we propose a parametric approach, i.e. to postulate an assumption on the distribution of the variables across stores in order to obtain an exact expression for $E(\ln S_t)$ and $E(\ln P_t)$.

*2.1.5. Limitations of the Considered Model*

As mentioned above, we decided to use a rather simple model since we want to concentrate on the aggregation issue. If, on the other hand, a researcher is more interested in trying to describe a certain market as realistically as possible, he/she would have to take care of cross-effects between competing brands or stores as well as lead and lag effects of promotional activities or the existence of threshold or saturation effects when running marketing communication programs (for a recent discussion of these impacts within a SCAN*PRO scenario see van Heerde, Leeflang & Wittink, 2000, 2001). Moreover, these phenomena may cause additional problems when aggregating data, e.g. cross-effects between promotional campaigns are probably more pronounced at the store than at the market level (i.e. promotional variables are expected to be correlated negatively at the store level but almost uncorrelated at the market level – if a brand is on special offer in a certain store in a certain week, one is unlikely to find simultaneously a promoted competitive brand in the same store and the same product category; this competitor is more likely to run a merchandising support at a different store at the same time).

We do not consider these effects or others due to omitted variables here but focus on the *bias which originates from linear aggregation of the data.* In particular, we compare arithmetic and geometric means and expose the differences using the parsimonious model (1). Our results hold for multiplicative models in general and, therefore, can also be applied to the comprehensive version of SCAN*PRO (Wittink et al., 1988) which does account for cross-effects. If lead or lagged variables are considered by means of some multiplicative functional formulation, our method might still be employed. The modeling of threshold or saturation effects usually require other kinds of nonlinear response functions (e.g. of exponential type) or a semiparametric analysis, which are both not compatible with our approach. In order to reduce the complexity of the general aggregation problem, we thus assumed that (1) is the "true" model (c.f. sub-section 2.1.1).

### 2.2. *Predicting Geometric Means*

#### 2.2.1. *Appropriate Distributions to Describe the Variability Across Stores*

As already noticed in sub-section 2.1.4. we propose to make use of statistical distributions in order to describe sales and prices varying across stores as well as to approximate $E(\ln S_t)$ and $E(\ln P_t)$. Concentrating on sales first, we note that they are not negative by definition. Moreover, empirical data usually exhibit skewed distributions reflecting a larger number of small outlets and a small number of large stores. This pattern extends to a wide range of cases in other areas known as the "80/20 rule" (i.e. for the example above, about 20% of stores account for around 80% of sales). Lawrence (1988) shows that for such a situation the *lognormal distribution* is the natural choice. He further provides an extensive overview of various applications of the lognormal distribution in economics and business, e.g. sales in an industry.

In principle, several other density functions might be capable to reflect this property as well, e.g. the Weibull or the gamma distribution. From a more pragmatic point of view, we will demonstrate in the next sub-section that one obtains a very simple closed form solution for $E(\ln S_t)$ if $S_t$ is lognormally distributed. We studied the relevant behavior of the other two distributions and derived analytical expressions which turned out to be computationally less attractive because they require the evaluation of Gamma and Digamma functions respectively; these functions cannot be easily processed by statistical *standard* software packages like SPSS. In line with relation (8) these approximations depend on $\ln E(S_t)$ and on $CV_{S_t}$, differing with respect to $CV_{S_t}$ only. Therefore, the choice of the appropriate distribution is a more relevant issue for more dispersed data (i.e. with increasing $CV_{S_t}$).

Lawrence (1988) also presents some examples from the literature dealing with the modeling of price perceptions by means of lognormal distributions. We admit, however, that the theoretical and empirical support for lognormality of the price variable is limited. Keep in mind that our purpose is to estimate the geometric mean of price indices rather than to describe accurately their distribution, which certainly is a less demanding task. Moreover, price indices fluctuate around one and do not vary substantially so that rather small coefficients of variation have to be expected.

Therefore, we regard the variables of interest, i.e. sales and prices, to be random across stores in the sequel and assume that $S_t$ and $P_t$ independently follow a lognormal distribution in each time period $t$. This is a flexible two-parameter ($\mu_{\ln X}$, $\sigma_{\ln X}$) distribution with the following mean and variance:

$$E(X) = \exp(\mu_{\ln X} + 0.5 \cdot \sigma^2_{\ln X}) \tag{10}$$

$$V(X) = \exp(2 \cdot (\mu_{\ln X} + \sigma^2_{\ln X})) - \exp(2 \cdot \mu_{\ln X} + \sigma^2_{\ln X}) \tag{11}$$

where $X$ denotes a positive random variable (i.e. sales or prices in our case).

*2.2.2. A Proposed Methodology*

Since:

(i) the logarithm of a lognormally distributed random variable varies according to a normal distribution with the same parameters (this is the reason for our notation with respect to the parameters of the lognormal distribution, i.e. $E(\ln X) = \mu_{\ln X}$) and

(ii) $\ln g_X = \ln \sqrt[K]{\prod_{k=1}^{K} X_k} = \frac{1}{K} \sum_{k=1}^{K} \ln X_k$

we employ the method of moments and obtain estimates for the parameters of the lognormal distribution:

$$\hat{\mu}_{\ln X} = 2 \cdot \ln\left(\frac{1}{K} \sum_{k=1}^{K} X_k\right) - 0.5 \cdot \ln\left(\frac{1}{K} \sum_{k=1}^{K} X_k^2\right) \tag{12}$$

$$\hat{\sigma}^2_{\ln X} = \ln\left(\frac{1}{K} \sum_{k=1}^{K} X_k^2\right) - 2 \cdot \ln\left(\frac{1}{K} \sum_{k=1}^{K} X_k\right) \tag{13}$$

If we use $S_t$ and $P_t$ respectively, instead of $X$ in (10), we find a convenient way to determine $\ln E(S_t)$ and $\ln E(P_t)$. Therefore, parameter calibration is performed via (12) and (13). This is the very point where the additional information required by our methodology comes into play (i.e.

$$\frac{1}{K} \sum_{k=1}^{K} X_k^2,$$

a descriptive measure about the variability of sales and prices across stores per period). Finally, we substitute (10) in (4) and arrive at:

$$\hat{\mu}_{\ln S_t} + 0.5 \cdot \hat{\sigma}^2_{\ln S_t} = \ln \alpha' + \beta' \cdot (\hat{\mu}_{\ln P_t} + 0.5 \cdot \hat{\sigma}^2_{\ln P_t}) + \bar{F} \cdot \ln \gamma' \qquad \forall t \tag{14}$$

When $S_t$ and $P_t$ follow a lognormal distribution, $\ln E(S_t)$ and $\ln E(P_t)$ overestimate the logarithms of the geometric mean of sales and prices, i.e. $\mu_{\ln S_t}$,

$\mu_{\ln P_t}$ by $0.5 \cdot \sigma^2_{\ln S_t}$, $0.5 \cdot \sigma^2_{\ln P_t}$, respectively. Moreover, the squared coefficient of variation of a lognormally distributed random variable equals $CV^2_X = \exp(\sigma^2_{\ln X}) - 1$ which can be approximated by $\sigma^2_{\ln X}$ (i.e. linear Taylor expansion at zero). These results are consistent with (8) and (9).

If the variances across stores $\hat{\sigma}^2_{\ln S_t}$ and $\hat{\sigma}^2_{\ln P_t}$ are constant over time, (14) becomes similar to (7) (except for the intercept). Thus, arithmetic means permit the identification of the parameters $\beta$ and $\gamma$ of (7). Also, (14) shows that when the prices remain the same across stores (homogeneous marketing activities) in a given week, but the variances of sales across stores differ over time, the parameters $\beta$ and $\gamma$ are again unidentified. This result might apply at the chain level: A chain manager may decide in favor of identical marketing mix activities in all of his outlets. Nevertheless, sales will still fluctuate according to store-specific circumstances.

Again, one might think of this issue in rather general econometric terms: the errors-in-variables model. In our case, the arithmetic means may be regarded as proxy variables for their true but generally unavailable geometric counterparts. In accordance with the literature (e.g. Judge et al., 1985, p. 705 ff.), we have demonstrated that this deficiency results in identification problems. In particular, we have explained that the identification of the store-level parameters from the market-level model depends on the variation of the variances of the variables over time. This identification problem disappears when:

(1) we know these variances, and
(2) we assume that the variables are lognormally distributed across stores.

## 3. EMPIRICAL APPLICATION

In this section, we will apply the methodology proposed above to empirical data and demonstrate that it is possible to determine parameters which are consistent with those derived with the use of actual geometric means in a multiplicative market-level model.

### 3.1. The Data

The data consist of two product categories in two European countries: dishwasher detergent in Austria and a chocolate product in France. Both products are frequently bought consumer goods with a nationwide distribution. There is intensive competition in both markets in terms of marketing efforts (i.e. pricing and promotional activities) and data are collected on a regular basis by means of scanning equipment. We use data provided by the A. C. Nielsen

Company at the store level. This enables us to compare the results of models' estimates at different levels of aggregation. This situation is exceptional for the marketing manager of a single brand since he usually has information at his disposal which refers only to the overall market, or sometimes to the chain level. As stated above, the SCAN*PRO model is frequently applied under such circumstances.

We study three national brands in each market (Brands 1, 2 and 3 in Austria; Brands A, B and C in France). For each brand, we use weekly sales data over 52 weeks and a price index which is calculated as the ratio between the price actually paid for the brand analyzed and the corresponding regular price. Thereby, one takes into account the effects of price cuts and the like. For ease of presentation, we will use the term 'price' in the sequel although 'relative price' would probably be more appropriate. Consistent with (1), we will not include cross-effects of marketing activities due to the varying distribution of brands across stores.

The consideration of promotional variables is case specific. For some brands we use 'display' or 'feature' variables; for others 'bonus pack', 'leaflet' or a combination of them. All of these are dummy variables defined such that, for example, the variable 'bonus pack only' is set to one if the brand under consideration was offered by providing some add-up during the relevant time period; if, simultaneously, the brand was on display or featured or the like, or was not promoted at all, the variable is set to zero. The selection of which promotional activity per brand to include in the SCAN*PRO model was based on the individual marketing policy and on the econometric analysis at the store level. Once selected, we kept all variables for subsequent computations and report on the estimated response coefficients.

Since the main purpose of our research aims to demonstrate potential aggregation effects when analyzing marketing data rather than trying to describe the market under study in detail, we concentrate our investigation on stores which report sales for the relevant brands every week. This results in a number of stores fluctuating between 18 and 34. The data set for the chocolate product has been previously analyzed by Bemmaor, Franses and Kippers (1999).

### *3.2. Preliminary Analysis*

Figures 1 and 2 depict the nature of the data at hand. They exhibit the two main marketing variables, sales and prices respectively, varying according to time and cross-section. In order to present the facts more clearly, we do not show the raw data but the distributions of sales and prices (approximated by lognormal

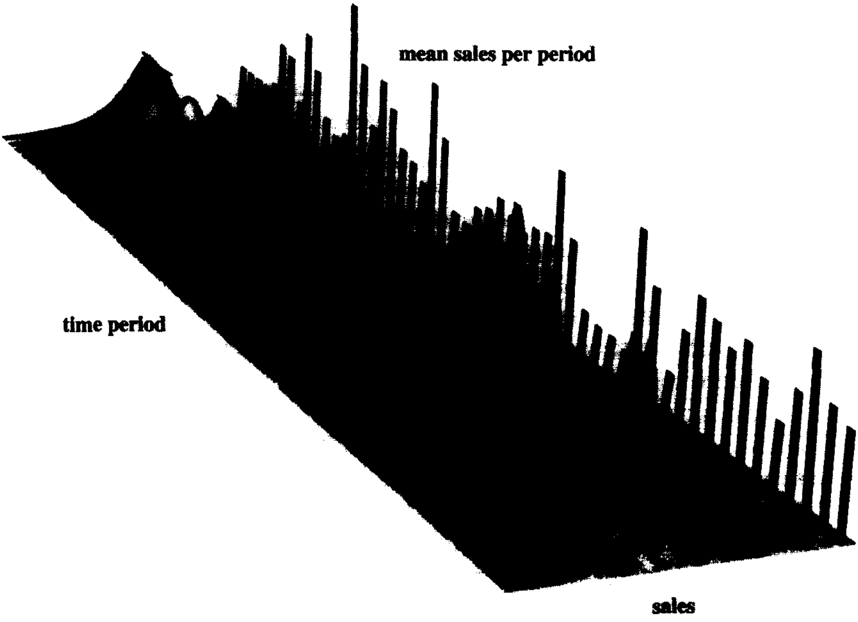


*Fig. 1.* Distribution of Sales Across Stores for Varying Time Periods (Brand 1, Austrian dishwasher detergent, 52 weeks).

density functions) for each of the 52 weeks for Brand 1 in the dishwasher detergent market. Simultaneously, we include mean sales and mean prices per period (averaged across cross-sections) using bar charts. For presentational reasons, different scales are used to display the density functions and the bar charts respectively for the two variables. Looking at averages instead of an ensemble of observations corresponds to the kind of data reduction encountered within the aggregation process from store to market level.

The sales distributions are highly skewed; their variances change from one period to the next. This pattern is in line with postulating lognormality. In fact, a Kolmogorov-Smirnov test performed on cross-sectional sales of Brand 1 per week could never reject this distributional assumption at the 5% level. Keep in mind that this is the more critical assumption (c.f. sub-section 2.2.1). The price distributions are much closer to symmetry because we are looking at indices, i.e. prices paid relative to regular prices (the latter have been calculated as the average prices paid in non-promoted stores for this data set). Therefore the lognormal distribution does not fit so well here with Kolmogorov-Smirnov tests rejecting this hypothesis in about 70% of the cases. As we used this

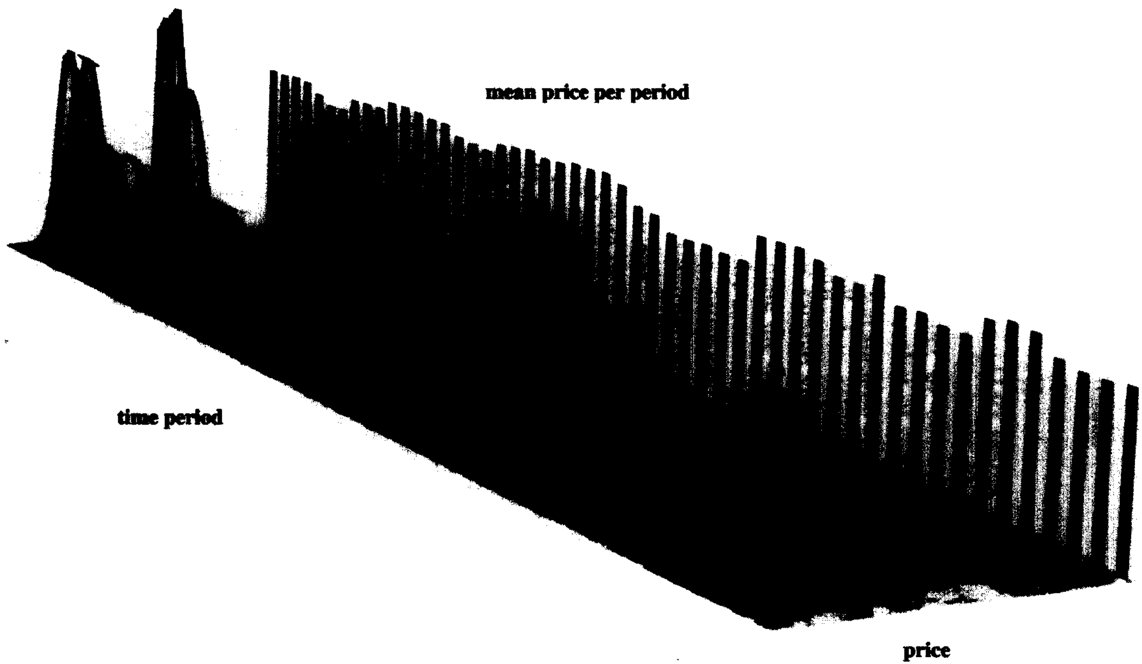


*Fig. 2.* Distribution of Prices Across Stores for Varying Time Periods (Brand 1, Austrian dishwasher detergent, 52 weeks).

*Table 1.* Correlation Coefficients Between Actual and Predicted Logarithms of Geometric Means.

| Product category/Brand | Sales | Prices |
|---|---|---|
| Dishwasher detergent | | |
| Brand 1 | 0.94 | 0.99 |
| Brand 2 | 0.97 | 1.0 |
| Brand 3 | 0.97 | 1.0 |
| Chocolate product | | |
| Brand A | 0.96 | 1.0 |
| Brand B | 0.99 | 1.0 |
| Brand C | 0.96 | 1.0 |

approximation mainly for presentational purposes in Figs 1 and 2, we are not substantially concerned about this mismatch. The distributional assumption is required to predict geometric means and we will analyze the accuracy of their forecasts later (c.f. Table 1).

As with sales data, the variances of prices differ in the course of time. Although the scaling effects outlined above should be considered, it is obvious that mean sales are characterized by more distinct fluctuations than mean prices. When reflecting the data as the realization of some stochastic process, the two figures illustrate another point. It is not possible to estimate ensemble variability by utilizing time series information, i.e. the stochastic process does not possess *ergodic* properties (e.g. Parzen, 1962, p. 72 f.). In fact, if we compare standard deviations over time to standard deviations across stores, the latter are typically much larger.

Postulating lognormal distributions for sales and prices, we used (12) to estimate the logarithm of their geometric means for every time period. In order to determine the goodness-of-fit between the actual and the predicted logarithms of the geometric means, we ran a correlation analysis over the 52 weeks first. The results are summarized in Table 1. Subsequently we checked for correspondence in magnitude by means of simple regression analyses; in all cases intercepts near zero and slope coefficients of about one were calculated. Clearly, the forecasts utilizing the lognormal distribution approximation are very accurate, both for sales and prices. The applicability in the case of sales was already claimed by Lawrence (1988); as expected, the fit for prices is even better, probably because their coefficients of variation are always much smaller (c.f. sub-section 2.2.1 and relation (8), Figs 1 and 2).

## *3.3. The Results*

### *3.3.1. Response Coefficients at the Store Level*

First, we estimate the linearization of (1) by taking logarithms over all stores by means of ordinary least-squares in order to achieve a benchmark for subsequent comparisons. The number of observations (c.f. Table 2) varies across brands due to the different number of stores which carry them over the 52-week period. The Durbin-Watson statistic diagnosed first-order autocorrelation, which is frequently the case when calibrating market response functions over a time domain. Since our sample sizes were rather large, we used the Durbin-Watson statistic to estimate the first-order autocorrelation coefficient $\hat{\rho}$ (Johnston, 1984, p. 315), thereby assuming $\rho$ to be constant across stores (this premise is analogous to the invariant response coefficients $\beta$ and $\gamma_j$). Subsequently, we performed the appropriate transformation of our variables and once again applied ordinary least-squares. For the results see Table 2, column 'Store-level model'.

We do not report the store-specific intercepts $\hat{\alpha}_k$ since they are of minor relevance here; in all cases they meet the condition of being positive and are highly significant most of the time. Although this model may not be the "true" one, we nevertheless feel that incorporating store-level information backs up the validity of the results. As can be seen, all parameters are highly significant and face valid, i.e. they possess the correct sign and the estimated influences of the promotional variables increase with intensified marketing efforts (e.g. Brand B's bonus pack only vs. bonus pack + leaflet, etc.). Please note that we actually report ln $\hat{\gamma}_j$ in Table 2, which means that a value of 0.18 (Brand B, bonus pack only) implies an estimated sales increase of about 20%.

### *3.3.2. Response Coefficients at the Market Level*

The next step consists of testing how well we can recover the parameters of the store-level model by means of three alternatives at the market level. For each brand in the two product categories, Table 2 displays the following types of results:

(1) the estimates obtained with the use of the actual geometric means (see sub-section 2.1.3. and espeoially Eq. (6); Table 2, column 'Actual geometric means');
(2) the estimates obtained with the use of the predicted geometric means

*Table 2.* Comparing the Estimates Across Aggregation Levels.*

| Product Category | Brand | Marketing mix variable | Store-level model | Market-level model: Actual geometric means | Market-level model: Predicted geometric means | Market-level model: Arithmetic means |
|---|---|---|---|---|---|---|
| Dishwasher Detergent | Brand A | Price | −2.97 (0.14) | −3.10 (0.57) | −3.24 (0.62) | −1.86 (0.98) |
| | | Display + feature | 0.37 (0.09) | 0.71$^{ns}$ (0.46) | 0.81$^{ns}$ (0.50) | 2.09 (0.91) |
| | | First-order autocorrelation | 0.42 | 0.82 (0.09) | 0.80 (0.09) | 0.44 (0.14) |
| | | Number of observations | 1560 | 52 | 52 | 52 |
| | Brand B | Price | −3.65 (0.15) | −2.93 (0.57) | −3.53 (0.66) | −4.73 (0.92) |
| | | Display + feature | 0.44 (0.10) | 0.89 (0.51) | 0.49$^{ns}$ (0.62) | 1.39$^{ns}$ (0.83) |
| | | First-order autocorrelation | 0.30 | 0.74 (0.10) | 0.67 (0.11) | 0.60 (0.12) |
| | | Number of observations | 1300 | 52 | 52 | 52 |
| | Brand C | Price | −1.64 (0.25) | −2.45 (1.12) | −2.71 (1.05) | −0.06$^{ns}$ (1.27) |
| | | Display only | 0.58 (0.12) | 1.03$^{ns}$ (0.65) | 1.17 (0.62) | 1.18$^{ns}$ (0.74) |
| | | Feature only | 0.73 (0.11) | 0.93$^{ns}$ (0.56) | 1.05 (0.52) | 1.32 (0.66) |
| | | Display + feature | 0.99 (0.10) | 0.88 (0.40) | 1.12 (0.38) | 1.69 (0.48) |
| | | First-order autocorrelation | 0.30 | 0.54 (0.14) | 0.56 (0.13) | 0.42 (0.14) |
| | | Number of observations | 936 | 52 | 52 | 52 |

* We report the estimates and the standard errors in parentheses corrected for first-order autocorrelation for all models. For the store-level model, we used the Durbin-Watson statistic of the model with the original variables to estimate the first-order autocorrelation coefficient (Johnston, 1984, p. 315); subsequently, we calibrated the coefficients by means of ordinary least-squares using the transformed variables. For the market-level models, we used nonlinear least-squares to estimate the full set of parameters (Johnston, 1984, p. 323).

$^{ns}$ not significant at the 5% level (one-tailed test, c.f. comments subsequent to Eq. (1))

***Table 2.*** (Continued).

| Product Category | Brand | Marketing mix variable | Store-level model | Market-level model: Actual geometric means | Market-level model: Predicted geometric means | Market-level model: Arithmetic means |
|---|---|---|---|---|---|---|
| Chocolate product | Brand A | Price | –2.53<br>(0.19) | –2.66[ns]<br>(1.78) | –3.27<br>(1.83) | –2.78[ns]<br>(2.57) |
| | | Bonus pack + display | 0.84<br>(0.07) | 1.30[ns]<br>(0.81) | 0.68[ns]<br>(0.83) | 1.60[ns]<br>(1.04) |
| | | Bonus pack + display + leaflet | 1.10<br>(0.11) | 1.50[ns]<br>(1.26) | 2.62<br>(1.29) | 5.40<br>(1.61) |
| | | First-order autocorrelation | 0.50 | 0.63<br>(0.11) | 0.63<br>(0.11) | 0.68<br>(0.11) |
| | | Number of observations | 1768 | 52 | 52 | 52 |
| | Brand B | Price | –1.84<br>(0.20) | –3.36[ns]<br>(3.00) | –3.14[ns]<br>(2.79) | –4.85[ns]<br>(4.05) |
| | | Bonus pack only | 0.18<br>(0.04) | 0.43[ns]<br>(0.42) | 0.38[ns]<br>(0.39) | 0.35[ns]<br>(0.52) |
| | | Bonus pack + leaflet | 0.51<br>(0.12) | 1.68[ns]<br>(1.33) | 1.11[ns]<br>(1.22) | 0.99[ns]<br>(1.60) |
| | | Bonus pack + display | 0.94<br>(0.08) | 1.54[ns]<br>(0.94) | 1.66<br>(0.86) | 2.93<br>(1.14) |
| | | Bonus pack + display + leaflet | 1.12<br>(0.15) | 0.91[ns]<br>(1.81) | 1.33[ns]<br>(1.67) | 2.95[ns]<br>(2.21) |
| | | First-order autocorrelation | 0.45 | 0.75<br>(0.10) | 0.75<br>(0.10) | 0.78<br>(0.10) |
| | | Number of observations | 1768 | 52 | 52 | 52 |
| | Brand C | Price | –2.18<br>(0.34) | –2.47[ns]<br>(1.89) | –2.00[ns]<br>(1.98) | –0.83[ns]<br>(1.82) |
| | | Bonus pack only | 0.27<br>(0.06) | 0.67[ns]<br>(0.44) | 0.60[ns]<br>(0.47) | 0.60[ns]<br>(0.40) |
| | | First-order autocorrelation | 0.42 | 0.61<br>(0.12) | 0.63<br>(0.14) | 0.63<br>(0.12) |
| | | Number of observations | 1404 | 52 | 52 | 52 |

(see sub-section 2.2.2. and once again Eq. (6) using (12) as approximation for the geometric means $g_{S_t}$ and $g_{P_t}$; Table 2, column 'Predicted geometric means');

(3) the estimates obtained with the use of the arithmetic means (see sub-section 2.1.2. and especially Eq. (3); Table 2, column 'Arithmetic means').

As above we have to account for first-order autocorrelation, which is achieved by using a nonlinear least-squares procedure (Johnston, 1984, p. 323). As expected, we find significant first-order autocorrelations in all cases. Once again we do not report the intercept $\alpha$ for a similar reason; in all cases they are positive and highly significantly different from zero. We first notice that for all the brands, the parameters of the store-level model (1) can be quite distinct (and frequently smaller in absolute terms) from those obtained from the market-level model estimated with the use of the actual geometric means (6). Although the market-level model (5) follows consistently from the store-level model (1), the equivalence between both equations applies if and only if (1) is the "true" model. Actual data shows that this might not be the case and calibrating (5) vs. (1) can lead to different estimates (for other potential reasons of this discrepancy see sub-section 2.1.5). On the whole, results based on (6) seem to be closer to the ones based on (1) than those employing (3).

Table 2 further reveals that the market-level estimates typically coincide with larger standard errors (shown in parentheses) than their store-level counterparts. This may in part be due to the loss of degrees of freedom but is mainly inherent in the aggregation process, i.e. larger sample sizes coincide with smaller variances. Consequently, we observe a large number of insignificant coefficients at the market level, especially for the chocolate product. Probably this effect is even more pronounced, since we allow for first-order autocorrelation: ordinary least-squares computations are expected to produce smaller standard errors (c.f. Johnston, 1984, p. 312). It is interesting to note that, e.g. Christen et al. (1997) neither report on first-order autocorrelations nor on significance levels at all, which is very much in line with common practice in marketing. A manager tends to evaluate response coefficients mainly on the basis of personal experience rather than on significance levels: he/she would not doubt, for example, that pricing activities do influence purchase behavior in (price-) competitive markets (Brands A, B, C); he/she would, however, look critically at the magnitude of the inferred elasticities. A statistician probably would be more conservative on this aspect.

Comparing the estimates obtained with the actual geometric means with those emerging from the predicted geometric means, we find that the calibrated coefficients are consistent in most cases. All significant parameters are face valid at a first glance but arithmetic means estimation results in substantially inflated coefficients in some cases (e.g. Brand A, bonus pack + display + leaflet).

### *3.3.3. Promotion Multipliers*

We want to expand further on this point and calculate the promotion multipliers $\hat{\gamma}_j$. As stated above, Table 2 actually reports ln $\hat{\gamma}_j$ and therefore we need to (back-) transform the response coefficients. Since we employed (3) and (6) for estimation purposes, we applied Kennedy's (1981) formula to compute the implied promotion effects; they are presented in Table 3 for each model. In some cases (e.g. Brand B, bonus pack + display + leaflet), we ended up with multipliers which were less than one. Since this would not make sense from a marketing point of view and because the estimates (ln $\hat{\gamma}_j$) were not significantly different from zero in all those instances, we set these multipliers equal to one.

Consistent with Christen et al. (1997), the response coefficients indicated by the model based on arithmetic means tend to be larger than the store-level multipliers. For this model, we sometimes obtain results which are far out from plausibility (e.g. Brand B, bonus pack + display); again, this is in line with the findings of Christen et al. (1997). Regarding the average impact of a promotional activity as inferred by the four alternative formulations, we observe that taking arithmetic means results in effects which are 95% larger for dishwasher detergent and 426% larger for the chocolate product when compared with the store-level scores. The effect becomes even more pronounced when computing the averages on significant response coefficients only. These percentages clearly underline the apparent deficiency resulting from employing arithmetic means within the SCAN*PRO model. If a marketing manager, who typically does not have access to store-level information as a benchmark for comparison, uses such figures, he clearly runs the risk of erroneously overstating the efficiency of promotional efforts.

The multipliers based on the predicted geometric means are fairly close to those obtained from the actual geometric means and both in turn roughly conform with the values achieved at the store level. Furthermore, the former do not exhibit a distinct pattern when compared with the latter ones.

### *3.3.4. Alternative Approaches to Correct for the Aggregation Bias*

Being aware of this fundamental deficiency of the SCAN*PRO model, Christen et al. (1997) introduced a debiasing procedure derived from simulated data. Having full knowledge of the underlying market at the store level, they generated a large number of different market conditions and promotional strategies, aggregated the data to market level, and systematically compared the

***Table 3.*** Comparing the Promotion Multipliers.*

| Product Category | Brand | Promotion variable | Store-level model | Market–level model: Actual geometric means | Market–level model: Predicted geometric means | Market–level model: Arithmetic means |
|---|---|---|---|---|---|---|
| Dishwasher Detergent | Brand 1 | Display + feature | 1.44 | *1.83**** | *1.98* | 5.34 |
| | Brand 2 | Display + feature | 1.55 | 2.14 | *1.35* | *2.84* |
| | Brand 3 | Display only | 1.77 | *2.27* | 2.66 | *2.47* |
| | | Feature only | 2.06 | *2.17* | 2.50 | 3.01 |
| | | Display + feature | 2.67 | 2.23 | 2.85 | 4.83 |
| | Average | | 1.90 | 2.12 | 2.27 | 3.70 |
| Chocolate product | Brand A | Bonus pack + display | 2.30 | *2.64* | *1.40* | *2.88* |
| | | Bonus pack + display + leaflet | 2.97 | *2.03* | 5.98 | 60.58 |
| | Brand B | Bonus pack only | 1.19 | *1.41* | *1.36* | *1.24* |
| | | Bonus pack + leaflet | 1.65 | *2.22* | *1.44* | *1.0*** |
| | | Bonus pack + display | 2.55 | *3.00* | 3.63 | 9.78 |
| | | Bonus pack + display + leaflet | 3.04 | *1.0*** | *1.0*** | *1.66* |
| | Brand C | Bonus pack only | 1.30 | *1.27* | 1.63 | *1.68* |
| | Average | | 2.14 | 1.94 | 2.35 | 11.26 |

* We computed the multiplier with the use of the expression given by Kennedy (1981), i.e. $\hat{\hat{\gamma}}_j = \exp(\ln \hat{\gamma}_j - 0.5 \cdot \hat{Var}(\ln \hat{\gamma}_j))$. This expression is based on the assumption that the residuals of the structural model are lognormally distributed.

** We set the multiplier equal to 1 when Kennedy's (1981) formula led to a number which was less than unity.

*** Promotion multipliers $\hat{\hat{\gamma}}_j$ whose corresponding estimates $\ln \hat{\gamma}_j$ are not significant are shown in *italics*.

estimated response coefficients with their true counterparts. They identified several determinants affecting the magnitude of the bias (e.g. average size of price cut, proportion of stores with promotions, $R^2$ of the regression at market level) and finally calibrated equations which can be used to approximate these biases for the SCAN*PRO response coefficients depending on the situation at hand. Unfortunately, we could not employ this procedure. As described above (c.f. Table 2) we had to deal with autocorrelated errors, which clearly influenced our parameter estimates. This type of environmental factor has not been considered by Christen et al. (1997), thus preventing the application of their method for our data.

An alternative debiasing procedure was proposed by Link (1995). Since bias occurs essentially when aggregating heterogeneous variables, subsets of stores for each time period are built that are close to homogeneous with respect to marketing activities. Contrary to the situation usually encountered in practice, this method implies, however, that data is available which has not yet been completely aggregated to market level. Christen et al. (1997) chose to group stores according to display and feature variables. This procedure is in line with the strategy Krishnamurthi, Raj and Selvam (1990) recommended to build groups of stores on the similarity of the explanatory variables. They proposed using an aggregate distance measure between cross-sections calculated from all marketing-mix variables and subsequently employing average linkage or Ward's minimum distance method for clustering.

Since the number of stores is rather limited for our data sets (between 18 and 34) and promotional activities have not been performed all the time, the first type of clustering is not feasible for us because it would result in empty groups. On the other hand, distance measures are expected to be dominated by the price variable in our case and, therefore, we decided to simply categorize according to the price indices, i.e. whether the brands have been offered below or above the regular price. We do not show the results in detail here because this procedure did not correct for the bias in the price elasticity in our case. As an example we report the estimates for Brand 1 (Price: –4.38; Display + feature: 1.99; First-order autocorrelation: 0.37). All coefficients are highly significant. When compared to the store level model (c.f. Table 2) we find a clear discrepancy.

### *3.4. Further Comments*

The improvement of the parameter estimates when employing the proposed methodology can be attributed to two facts (c.f. sub-section 2.2.2.):

(1) the availability of additional information (i.e. variances across stores); and

(2) the assumption that the variables follow a certain distribution (i.e. lognormal).

Can we still correct for the biases when standard deviations are *not known* ? In fact, this is the case analyzed by Christen et al. (1997). A way to overcome this problem might be maximum entropy estimation. It provides "... a criterion for setting up probability distributions on the basis of partial knowledge ..." and "... it is the least biased estimate possible on the given information" (Jaynes, 1957, p. 620). Assuming that the variables are lognormally distributed across stores and that we know their arithmetic means $\bar{X}$ only, we derive the maximum entropy estimators $\hat{\mu}^h_{\ln X}$, $\hat{\sigma}^h_{\ln X}$ of the parameters of the lognormal distribution and obtain (Wagner & Geyer, 1995):

$$\hat{\mu}^h_{\ln X} = \ln \bar{X} - 0.5 \tag{15}$$

$$\hat{\sigma}^h_{\ln X} = 1.0 \tag{16}$$

As shown in (15), the entropy estimator of the logarithm of the geometric mean $\hat{\mu}^h_{\ln X}$ is consistent (except for the constant 0.5) with the procedure proposed by Christen et al. (1997), which involves taking the logarithm of the means $\ln \bar{X}$ to estimate (4). However, empirical evidence shows that standard deviations vary over time (see Fig. 1), which is not in line with the constant $\hat{\sigma}^h_{\ln X}$ in (16). Therefore, knowledge of the means only is regarded as insufficient to identify both parameters of a lognormal distribution. Nevertheless, we thereby provided an additional interpretation of the estimation procedure recommended by Christen et al. (1997).

If managers find the proposed methodology useful to improve the reliability of response coefficients of the SCAN*PRO model estimated at the market level, they will probably ask their data providing marketing research firms to supply the necessary information additionally. This can be accomplished easily by, e.g. Nielsen or IRI since the required data are collected anyhow. Furthermore, standard deviations represent descriptive measures on an aggregate level and thus commonly observed reservations of data providing companies about giving away individual store-level figures does not apply in this case. It is interesting to note that Krishnamurthi, Raj and Selvam (1990) also addressing the aggregation problem but employing quite a different methodology conclude that a closer cooperation between data providers and their clients will be necessary because "... it will not be possible to assess the extent of aggregation bias if all one has is aggregate data and no information on how the data were aggregated."

# 4. CONCLUSION

This study illustrates that we can obtain good predictions of the geometric means of sales and prices from their arithmetic means and variances under the assumption that these variables are lognormally distributed across stores. Moreover, these forecasts can be calculated very easily when using statistical standard software packages. Consequently, we may employ these predictions in a market-level multiplicative sales response model. Tested on six brands in two product categories, this methodology leads to parameter estimates which are consistent with those obtained with the use of the actual geometric means. Compared to store-level parameters, the coefficients achieved with the use of the predicted geometric means do not exhibit a definite pattern and are comparable in most cases. Obviously, this result is preliminary due to the limited empirical evidence so far. Consistent with previous work, the estimates of the promotion multipliers based on the use of arithmetic means exhibit substantial overestimation when compared with store-level results. Finally, the paper exemplifies that knowledge of the means (or sums) of the variables only is insufficient to identify the parameters of the store-level model unless their variances across stores are constant over time.

The paper has predominantly aimed to address a very basic marketing modeling problem: the effects of aggregation over cross-sections on the estimates of response coefficients for a given (i.e. "true") model. We raised the issue from a more formal perspective and proposed a parametric methodology to partly overcome the problems involved. We are aware of the fact that there are still a lot of open questions and identify the following areas for further research:

- The method of moments to calibrate the parameters of the lognormal distribution should be replaced with a more efficient estimation principle (i.e. maximum likelihood).
- The proposed methodology could be extended to use alternative flexible distributions such as the Weibull, gamma or inverse gaussian.
- Alternative kinds of information might be employed in order to solve the identification problem differently, e.g. additional moments or other assumptions on the relationship between moments.
- Possibly, empirical regularities between the parameters of a lognormal or of alternative distributions for sales and prices might be detected. In such a case the requirement to refer to standard deviations might then not be necessary any more.

## ACKNOWLEDGMENTS

The authors are grateful to Cornelia Lamm and Markus Cserjan, formerly at A. C. Nielsen Austria, and to Gilles Gros, A. C. Nielsen France, for providing them with data. They further appreciate the comments given by Markus Christen when discussing a preliminary version of this paper and by Susan Cohen-Unger on improving the English style.

## REFERENCES

Bemmaor, A. C., Franses, P. H., & Kippers, J. (1999). Estimating the Impact of Displays and Other Merchandising Support on Retail Brand Sales: Partial Pooling with Examples. *Marketing Letters, 10*(1), 87–100.

Blattberg, R., & Neslin, S. A. (1990). *Sales Promotions: Concepts, Methods and Strategies.* Englewood Cliffs, New Jersey: Prentice-Hall.

Blattberg, R., & Wisniewski, K. J. (1989). Price-Induced Patterns of Competition. *Marketing Science, 8*(4), 291–309.

Bucklin, R. E., & Gupta, S. (1999). Commercial Use of UPC Scanner Data: Industry and Academic Perspectives. *Marketing Science, 18*(3), 247–273.

Christen, M., Gupta, S., Porter, J. C., Staelin, R., & Wittink, D. R. (1997). Using Market-Level Data to Understand Promotion Effects in a Nonlinear Model. *Journal of Marketing Research, 34* (August), 322–334.

Geweke, J. (1985). Macroeconomic Modeling and the Theory of the Representative Agent. *American Economic Review, 75* (May), 206–210.

Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *Physical Review, 106,* 620–630.

Johnston, J. (1984). *Econometric Methods* (3rd ed.). Auckland: McGraw-Hill Book Company.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., & Lee, T. C. (1985). *The Theory and Practice of Econometrics.* New York: John Wiley and Sons.

Kennedy, P. E. (1981). Estimation With Correctly Interpreted Dummy Variables in Semi-logarithmic Equations. *American Economic Review, 71* (September), 801.

Krishnamurthi, L., Raj, S. P., & Selvam, R. (1990). Statistical and Managerial Issues in Cross-Sectional Aggregation. Working paper, Northwestern University.

Lawrence, R. J. (1988). Applications in Economics and Business. In: E. L. Crow & K. Shimizu (Eds), *Lognormal Distributions: Theory and Applications* (pp. 229–266). Marcel Dekker, New-York.

Leeflang, P. S. H., Wittink, D. R., Wedel, M., & Naert, P. A. (2000). *Building Models for Marketing Decisions.* Boston: Kluwer Academic Publishers.

Link, R. (1995). Are Aggregate Scanner Data Models Biased? *Journal of Advertising Research, 35* (September/October), RC 8–12.

Montgomery, A. L. (1997). Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data. *Marketing Science, 16*(4), 315–337.

Parzen, E. (1962). *Stochastic Processes.* San Francisco: Holden-Day.

Theil, H. (1974). *Linear Aggregation of Economic Relations.* Amsterdam: North-Holland Publishing Company, Inc.

van Heerde, H. J., Leeflang, P. S. H., & Wittink, D. R. (2000). The Estimation of Pre- and Postpromotion Dips with Store-Level Scanner Data. *Journal of Marketing Research, 37* (August), 383–395.

van Heerde, H. J., Leeflang, P. S. H., & Wittink, D. R. (2001). Semiparametric Analysis to Estimate the Deal Effect Curve. *Journal of Marketing Research, 38* (May), 197–215.

Wagner, U., & Geyer, A. (1995). A Maximum Entropy Method for Inverting Laplace Transforms of Probability Density Functions. *Biometrika, 82*(4), 887–892.

Wittink, D. R., Addona, M. J., Hawkes, W. J., & Porter, J. C. (1988). SCAN*PRO: The Estimation, Validation and Use of Promotional Effects Based on Scanner Data. Working paper, Cornell University.

# MARKET STRUCTURE ACROSS STORES: AN APPLICATION OF A RANDOM COEFFICIENTS LOGIT MODEL WITH STORE LEVEL DATA

Pradeep Chintagunta, Jean-Pierre Dubé and Vishal Singh

## ABSTRACT

*Market structure analysis continues to be a topic of considerable interest to marketing researchers. One of the most common representations of the manner in which brands compete in a market is via market maps that show the relative locations of brands in multi-attribute space. In this paper, we use logit brand choice models to estimate a heterogeneous demand system capable of identifying such brand maps. Unlike the previous literature, we use only aggregate store-level data to obtain these maps. Additionally, by recognizing that there exists heterogeneity in consumer preferences both within a store's market area as well as across store market areas, the approach allows us to identify store-specific brand maps. The methodology also accounts for endogeneity in prices due to possible correlation between unobserved factors, such as shelf space and shelf location that affect brand sales, and prices. We provide an empirical application of our methodology to store level data from a retail chain for the laundry detergents product category. From a manager's perspective, our model enables micromarketing strategies in which promotions are targeted to those stores in which a brand has the most favorable, differentiated, position.*

**Econometric Models in Marketing, Volume 16, pages 191–221.**

**ISBN: 0-7623-0857-5**

# 1. INTRODUCTION

The analysis of market structure (Day, Shocker & Srivastava, 1979) continues to be an important area of research in marketing. The basic endeavor of this line of research is to identify, from consumer data, the extent to which different brands in a pre-defined product market compete with each other (Grover & Srinivasan, 1987; DeSarbo & Rao, 1986; Jain, Bass & Chen, 1990). The extent of inter-brand competition is characterized in one of two ways. Allenby (1989) for example, uses the matrix of price elasticities as the basis for identifying market structure. The idea here is that the greater the cross-price elasticities between two brands, the higher is the level of competition. An alternative measure of competition that has also been used by researchers is the extent to which preferences for two brands are correlated with one another in the marketplace, after controlling for the effects of marketing activities such as price. In this case, higher levels of correlation are associated with more intense levels of competition. The key advantage of this latter approach is that the representation of the market is not contaminated by short term (marketing) activities of firms and represents a more stable feature of the marketplace. We focus on the latter approach in developing our model of market structure in this paper.

While early work focused on identifying the pattern of inter-brand rivalry from stated preference data, the majority of the more recent literature has done so using revealed preference data. Researchers have also proposed hierarchical (Urban, Johnson & Hauser, 1984; Kannan & Wright, 1991; Ramaswamy & DeSarbo, 1990) as well as non-hierarchical approaches to understanding market structure (Elrod, 1988; Elrod & Keane, 1995; Chintagunta, 1998). In the hierarchical approach, consumers are assumed to be making a sequential set of decisions. By uncovering this sequence, researchers can obtain insights into the nature of perceived similarities across brands in a category. For example, consider the coffee category. Consumers can first decide to purchase either ground coffee or instant coffee. If they decide to choose ground coffee, they need to choose the kind of bean – Robusta or Arabica. Having made this decision, they have a choice of roasts. Finally, they can select a particular brand – Maxwell House, Folgers, etc. This indicates that two brands of ground coffee made with Arabica beans and French roast compete more closely with one another than do an Arabica bean brand and a Robusta bean brand. The non-hierarchical approach on the other hand allows for a general pattern of similarity across all brands in the product market considered. Our focus in this paper is on non-hierarchical approaches to market structure analysis that use revealed preference data.

A popular approach to understanding inter-brand similarity and rivalry using revealed preference data (usually scanner data) has been to construct market maps that pictorially depict the locations of brands in multi-attribute space (Elrod, 1988). The basic idea behind this approach is intuitively appealing. If consumers perceive two brands to be similar to one another, then these brands are likely to compete closely for that consumer's purchases. While this assumes that consumers are primarily inertial rather than variety seeking, most studies (with the possible exception of Kannan & Sanchez, 1994; Erdem, 1996) have found this to be the case. The appealing aspect of representing brands in multi-attribute space is that it relates quite closely with the notion of a perceptual map. Perceptual maps are regarded as fundamental building blocks for positioning analyses undertaken by marketers. Hence, approaches to depicting brands in multi-attribute space have become increasingly popular of late. We propose one such approach in this paper.

Researchers have proposed several approaches to depicting brands in multi-attribute space using consumer (revealed preference) data. Broadly, these approaches can be categorized based on the level of aggregation in the data used for the analysis. Specifically, previous studies have used either household level panel data or store level scanner data to uncover the underlying relationships among brands in the marketplace. This broad categorization nests several specific model formulations that lead to brand maps as an end result. For example, studies using store data have either exploited the structure of the cross-price elasticity matrix (as noted previously) in identifying the relative locations of brands in multi-attribute space (e.g. Allenby, 1989), or they have used the nature of price-quality tradeoffs made by consumers that are inherent in DEFENDER (Hauser & Shugan, 1983; Hauser & Wernerfelt, 1988) type models to identify brand locations along (or sometimes within) the price-quality frontier prevailing in the marketplace (Shugan, 1987; Waarts et al., 1991). The former approach allows researchers to account for the effects of other marketing activities (such as advertising and promotions) on brand sales, but is nevertheless focused on brand interactions along one specific marketing activity – price. The latter approach does not easily account for time-varying factors other than price as drivers of consumer choices. Further, as the model specification does not fall into the general class of discrete choice models (logit, probit, etc.), the approach has seen limited application in the marketing literature.

Household data on the other hand, have been used to locate brands in multi-attribute space by exploiting how heterogeneous consumers' preferences for brands in the marketplace are correlated across those brands. The basic idea here is that if intrinsic preferences for two brands are correlated across households, then these brands are going to be located close to one another in attribute space. The appealing feature of this approach is that the maps are obtained after controlling for the effects of marketing activities. At the same time, one can also study the effects of specific marketing activities on the inter-brand relationship in addition to the intrinsic brand similarities. As a consequence, models based on household data have abounded in the more recent marketing literature. The household level models that have been used to derive such maps have been choice models such as the multinomial logit model (Elrod, 1988; Chintagunta, 1994; Erdem, 1996) and the multinomial probit model (Elrod & Keane, 1995; Chintagunta & Honore, 1996) whose parameters are estimated using panel data from a group of households. In the estimation of these models, the covariance matrix of brand preferences that is estimated from the data is decomposed to provide the brand maps. This is accomplished by imposing a factor structure on the covariance matrix of preferences. The number of factors corresponds to the number of attributes and the parameters for the brands along those factors represent the brand locations along the attributes.

In this paper, we propose a methodology for obtaining brand maps from store level scanner data. Unlike extant approaches with such data however, we obtain these maps by decomposing the covariance matrix of brand preferences much like the methods using household level data. Such an approach requires us to be able to accommodate heterogeneity across consumers while using store level data as it is this heterogeneity in preferences across consumers that generates the covariance matrix which is decomposed as described for household data above. To accomplish this, we follow the recent literature in economics that treats store data as the aggregation of choices made by heterogeneous consumers in that market area (see Berry, Levinsohn & Pakes, 1995, hereafter referred to as BLP; Nevo, 2001). Brand choice decisions of individual consumers are modeled using a logit demand model – consistent with utility maximizing behavior and also with the models used in conjunction with household data to obtain market maps. Using this approach, we are able to recover the heterogeneity distribution across consumers in that store's market area. Hence, we are able to exploit the decomposition of the covariance matrix as has been done by researchers using household level scanner panel data. Accounting for heterogeneity with aggregate data therefore, allows us to derive market maps with store level data. Previous research using data at this level of aggregation (Shugan, 1987; Allenby, 1989) has also attempted to control for heterogeneity in the analysis. However, by imposing the structure that we do in this paper, we are able to obtain maps that correspond to those obtained from household data.

More interestingly, we demonstrate how we are able to obtain implications for market structure above and beyond what researchers have been able to accomplish with household level data. Specifically, we show that by making the distribution of heterogeneity in a particular store's market area a function of the demographic characteristics of consumers in that store area, we are able to recover market maps specific to each store in a retail chain. In other words, we are able to say whether or not two brands of soap – Ivory and Dove that are perceived to be similar by consumers in the market area of a store located in a northern suburb of Chicago, are perceived to be similar by consumers of a store (from the same chain) located in downtown Chicago. Hence, we are able to account for heterogeneity not only across consumers within a store area, but also across different store areas.

The key advantages of store data relative to household data are the following. First, store data are widely available to marketing managers and are used as a key resource for decision making. By contrast household data require a lot more computational resources to deal with and also tend to be more expensive for firms to acquire. Second, store level brand maps are useful for managerial decision making. Knowledge of differences in brand perceptions across stores has implications for managers at both manufacturer as well as retailer levels. Consider a manufacturer dealing with two retail chains in a market area – one with 10 stores and the other with 15. The manufacturer is concerned about a competitive threat from one particular rival brand. Prima facie, the manufacturer might be using the number of stores as a basis for allocating promotional money and so allocates more money to the chain with 15 stores. However, the manufacturer may be better able to allocate promotional moneys across the retailers if it knew that it was competing (perceptually) with the rival brand in only 2 of 15 stores from the larger retail chain whereas it was doing so in 8 of 10 stores from the smaller chain. Our methodology and model results would be a useful input to this allocation problem. With household panel data on the other hand, one does not typically have access to a large enough number of purchases in each store to be able to derive market maps for each store area. Similarly, there are important implications for the retailer as well. Knowledge of variations in inter-brand rivalries across stores will enable the retailer to better adjust shelf location, shelf space allocation etc. in order to maximize its category profits. An aggregate market structure analysis across all stores in the chain that does not allow for relative brand perceptions to vary across stores may not be very useful in this regard.

Our discussion above is not intended to imply that store data are to be preferred to household data under all circumstances. If one is interested, for example, in decomposing the impact of marketing activities into their effects on purchase incidence, brand choice, brand switching and purchase quantity decisions, this can best be accomplished using household level data. However, there might be certain situations in which one can get a richer set of managerially relevant information using store data.

Another important feature of our model and analysis is that we allow for factors other than observed marketing variables (i.e. prices and promotions) to impact the individual choices that are aggregated to the store level. At the store level, there are factors like shelf space, shelf location and store coupons that will affect the choices made by consumers in that store's market area. However, as researchers we do not observe these factors. We therefore account for them in a manner similar to how various observed factors such as price are included in logit brand choice models. Specifically, we include a brand specific and time varying unobserved "attribute" that influences brand utility and consequently, the consumer's choices. Further, as the unobserved attribute for a brand can be correlated with price, we account for this correlation by instrumenting for prices in the estimation. In this way we address the issue of price endogeneity in the estimation (see also Besanko, Gupta & Jain, 1998, hereafter referred to as BGJ). Additionally, our approach offers the same advantages as a mixed logit model (Kamakura & Russell, 1989) estimated with household data in that aggregate elasticities are free from the IIA restriction.

We provide an empirical illustration of our proposed approach using data from the liquid laundry detergent category. Our results indicate that brand locations vary substantially across stores. In particular, we find that certain brands may have more favorable conditions in some stores than others. In one store, the brand may be fairly differentiated from competitors, whereas in other stores the same brand is tightly packed in a cluster of brands. We use store-level characteristics to explain these differences.

The rest of this paper is organized as follows. In the next section, we describe the model formulation. This is followed by a section on the data. The penultimate section provides the results from the empirical applications. The final section concludes.

## 2. MODEL

### *2.1. Utility and Demand*

In this section, we describe the underlying consumer choice model generating the observed aggregate purchases in each store-week. We use the increasingly popular mixed logit specification (McFadden & Train, 2000). For a more general discussion of discrete choice models and their aggregation we refer the reader to BLP (1995). Formally, we assume that on a given shopping trip in week $t$ ($t = 1, \ldots, T$), $M_t$ consumers each select one of $J$ brands in the category or opt for the no-purchase alternative, whose utility is normalized to 0. In a store-week $t$, the brand $j$-specific factors are the attributes: $(x_{jt}, \xi_{jt})$. The vector $x$ includes brand-specific fixed-effects as well as an indicator for the incidence of a deal. This vector also includes the variable $p$ which denotes the brand's shelf-price. Finally, $\xi$ encompasses the effects of unobserved (to the econometrician) in-store product attributes, such as advertising, shelf-space and coupon availability that vary across store-weeks (BLP, 1995; BGJ, 1998).[1]

For a shopping trip during week $t$, the conditional utility consumer $h$ derives from purchasing product $j$ is given by:

$$u_{hjt} = \alpha_{hj} + x_{jt}\beta_h + \xi_{jt} + \varepsilon_{hjt},$$

$$h = 1, \ldots, H, j = 0, \ldots, J, t = 1, \ldots, T.$$

The coefficients $\beta_h$ capture consumer $h$'s tastes for attributes, $x$, which includes prices and marketing mix variables. The parameter $\alpha_{ht}$ captures household $h$'s idiosyncratic perception of brand $j$. The term $\varepsilon_{hjt}$ is an i.i.d. mean-zero stochastic term capturing consumer $h$'s idiosyncratic utility for alternative $j$ during week $t$. We assume that $\varepsilon_{hjt}$ has a type $I$ extreme value distribution. Since we do not observe the true distribution of consumer preferences, we assume tastes and brand perceptions are drawn from a multivariate normal distribution. For simplicity, we treat the taste parameters as i.i.d:

$$\beta_h = \bar{\beta} + \lambda' \nu_h, \ \nu_h \sim N(0, I)$$

where the vectors of means, $\bar{\beta}$, and the standard deviations, $\lambda$, are parameters to be estimated.

We do allow for a richer covariance structure for the vector of brand perceptions:

$$\alpha_h \sim N(\bar{\alpha}, \Sigma).$$

In theory, we could estimate the full $(J \times J)$ matrix $\Sigma$ directly. To recover our perceptual map, we could use a multi-dimensional scaling procedure. In practice, as the number of product alternatives grows, $\Sigma$ becomes increasingly difficult to identify. Instead, we use the factor structure:

$$\Sigma = L\omega\omega' L', \ \omega \sim N(0, I).$$

One interpretation for this structure is that $L$ is a $(J \times K)$ matrix of latent attributes for each of the $J$ brands, and $\omega$ is a $(K \times 1)$ vector of tastes for these attributes that is consumer-specific. The vector for each consumer is a draw from the standard multivariate normal distribution. The vector of mean brand perceptions, $\bar{\alpha}$, and the matrix of latent attributes, $L$, consist of parameters to be estimated. In addition to its parsimony, this approach allows us to estimate standard errors for the latent attributes. In the current context, we assume $K=2$. For identification purposes, we do the following: (1) The outside or "no purchase" option is located at the origin of the map (translational invariance); (2) One of the brands is located along the horizontal axis (rotational invariance); and (3) We set the variances of $\omega$ above to 1 in the estimation (scale invariance).

We simplify our notation by re-writing the consumer's indirect utility in terms of mean tastes and deviations from the mean:

$$u_{hjt} = \delta_{jt} + \mu_{hjt} + \varepsilon_{hjt}$$

where $\delta_{jt} = \bar{\alpha}_j + x_{jt}\bar{\beta} + \xi_{jt}$ is common to all consumers and $\mu_{hjt} = x_{jt}\lambda\nu_h + L_j\omega_h$ is consumer-specific and $L_j$ denotes the $j$th row of the $L$ matrix. After mixing the normally-distributed taste shocks with the extreme value disturbance, the probability $q_{jt}$ that a consumer chooses a particular product $j$ in week $t$ has the following form:

$$q_{jt} = \int_{-\infty}^{\infty} \frac{\exp(\delta_{jt} + \mu_{hjt})}{1 + \sum_{i=1}^{J} \exp(\delta_{it} + \mu_{hit})} \phi(\nu)\, \partial\nu, \tag{1}$$

$$h = 1, \ldots, H, j = 0, \ldots, J, t = 1, \ldots, T.$$

where $\phi(\cdot)$ is the pdf of a standard normal. From the store manager's perspective, (1) represents the share of consumers entering the store in week $t$ that purchase a unit of product $j$. Thus, the manager's expected demand for product $j$ in store-week $t$ is :

$$Q_{jt} = q_{jt}M_t. \tag{2}$$

We have two primary motivations for using this random coefficient's specification, as opposed to a simpler conditional logit (or homogeneous logit). The heterogeneity provides a means by which to recover the underlying brand maps. However, heterogeneity also provides a layer of flexibility in consumer responses which will be important in various other applications using the same demand model. The conditional logit's restrictive IIA property (the independence of irrelevant alternatives property) at the consumer level would manifest itself into our aggregate analysis in several ways. First, it can be shown that the assumption of homogeneous tastes leads to aggregate cross-elasticities that are driven by market shares (see BLP, 1995 for a thorough discussion). For instance, products with similar market shares are predicted to be close substitutes. In addition to the potentially unrealistic predicted substitution patterns, the cross-elasticities also restrict the implied retailer behavior in equilibrium. Multiproduct firms are restricted to set a uniform margin for each of the products in their line (Besanko, Dubé & Gupta, 2001). In an analysis of category management, this property would imply that all of the products in a category have the same mark-up over their wholesale prices. We solve this problem by allowing for consumer-specific deviations from mean tastes that are distributed normally. McFadden and Train (1998) show that the mixture of normals with the type I error, the *mixed logit*, is sufficiently flexible to approximate a broad set of parametric indirect utility functions, including the multinomial probit. Nonetheless, one could easily use alternative, non-Normal, distributions if desired.[2]

One of the complications of the mixed logit specification (1) is the lack of an analytic form for the multidimensional integral. While it is true that for a simple model with fewer than three random parameters one could solve the expression numerically (Hausman & Wise, 1978), most categories consist of more than three alternatives. Instead, we use direct Monte Carlo simulation, as in Nevo (2001).

### 2.2. *Local Interactions*

Other than marketing mix variables, we have not yet discussed store-specific covariates that allow expected demand to vary across stores. In practice, we do not expect each store in a chain to face the same distribution of consumers. Stores in different neighborhoods typically face different demographic distributions of consumers. Moreover, the presence of local competitors could alter a store's levels of demand in various categories. We expect differences in both the distribution of consumer types and the presence of local competitors to alter the the derived demand for goods facing each store. Figures 1 and 2 demonstrate differences in the strategic role of brands across stores. The share of sales for the category leader, Tide, and the brand Wisk show considerable heterogeneity across the various store areas. While Tide's shares vary from 0.08 to 0.24, Wisk's vary within the tighter interval of 0.1 to 0.2. Since wholesale prices are the same across stores, these share patterns must reflect differences in the derived demand for a product in a given store. Note however, that there do apprear to be significant outliers for both brands.

To account for differences in demand across stores, we introduce variables specific to each store area into the model formulation. Specifically, these

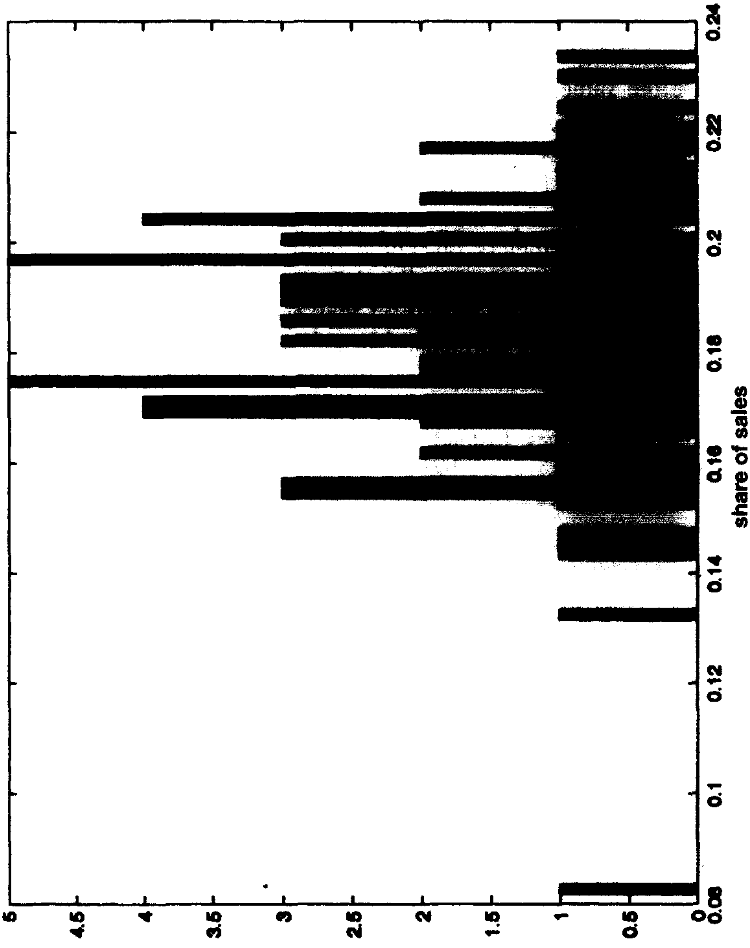


*Fig. 1.* Distribution of Share of Sales of 64 oz Tide Across Stores.

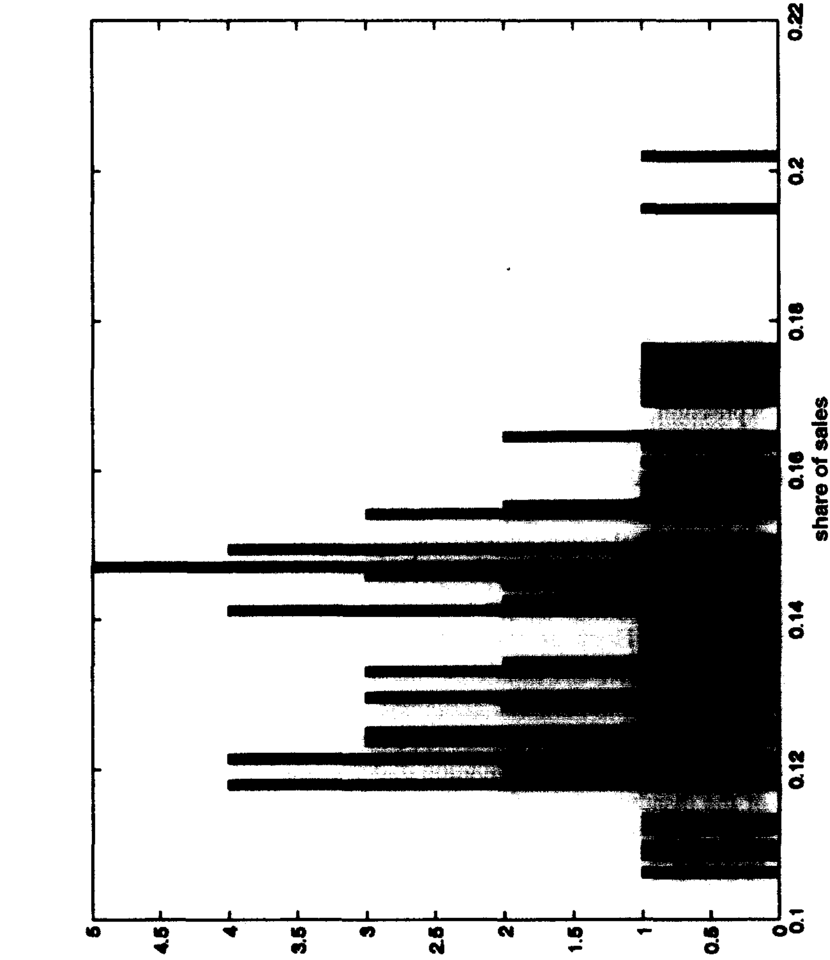


*Fig. 2.* Distribution of Share of Sales of 64 oz Wisk Across Stores.

variables are intoduced in two different ways. First, we include a store-specific category intercept that is a function of these store characteristics. Formally, this intercept will shift the size of the category share (relative to the no purchase share) across stores. Operationally, this involves shifting each of the brand intercepts by an equal amount. Second, we interact store characteristics with the latent attributes for each brand j, $L_j$. This allows consumers' inter-brand preference correlations to vary across stores. It is this variation in preference correlations across stores that gives us store-specific perceptual maps (recall our discussion in the introduction about preference correlations being used as a measure of market structure). We use a detailed set of variables that proxy for both differences in the mean demographic profiles and levels of competition facing each store. By including these variables in the formulation, the indirect utility expression that a consumer $h$ in store $s$ has for brand $j$ in week $t$ is given by the following:

$$\begin{aligned} u_{hsjt} &= \alpha_{hsj} + D_s\gamma + x_{jt}\beta_h + \xi_{sjt} + \varepsilon_{hsjt}, \\ \alpha_{hsj} &= L_j\omega_h + L_j\omega_h D_s\theta_j. \end{aligned} \tag{3}$$

In the above expression, $D_s$ is the row vector of characterisitcs for store area $s$ and $k$ is the dimension of the latent attribute vector. The term $D_s\gamma$ which has a common effect across all brands except the outside good shifts category demand up or down. The term $D_s\theta_j$ on the other hand interacts with $L_j$ to generate store-specific preference correlations. Using this approach, we are able to estimate all the demand parameters including the effects of store characteristics jointly in a single stage estimation process. This is in contrast with other studies (e.g. Hoch et al., 1995) that first estimate the demand elasticities and then regress these elasticities on store characteristics.

An alternative approach to the one proposed above is that used by Nevo (2001). He captures differences in consumer profiles across city-markets by sampling from the empirical joint distribution of demographics collected by the Census. One of the disadvantages of our disaggregate data is that comparable joint-demographic distributions are not available at the individual store's market-level (e.g. zip code or city block). Only marginal distributions are reported by the census. So we use mean demographics instead. To the best of our knowledge, no empirical study has explicitly modeled store competition in determining aggregate demand for a retailer.[3] Most applications of store-level data treat retailers as local monopolists (e.g. Slade, 1995; BGJ, 1998). Our telephone interviews with local store managers suggest that stores do condition on their competitors' action. Since we do not observe competitors' prices in our data, we cannot model competition explicitly. Instead, we assume that local market power is captured, on average, by proximity to competitors.

## 3. ESTIMATION

We now outline the estimation procedure for the mixed logit. Since one of our objectives in this analysis is the determination of the level of aggregation at which stores determine prices, we estimate demand alone. Unlike BLP (1995), we do not require additional supply-side moments for identification as we observe the exogenous wholesale prices and use these as instruments. This approach also ensures that our demand-side estimates are not subject to specification error from incorrectly assuming static category management by retailers. Since our estimation methodology is quite similar to that used by BLP (1995), we only provide an outline. We refer the more interested reader to BLP (1995) for a more technical description and to Nevo (2001) for a more thorough discussion of the implementation of the methodology.

A primary concern in empirical papers using similar discrete choice models is the potential for estimation bias due to correlation between prices and the unobserved product attribute, $\xi$. Using weekly store-level data, our primary concern lies in unmeasured store-specific covariates that influence demand and also shift prices. Several papers have documented evidence of an estimation bias in models that do not control for this problem using weekly supermarket data (BGJ, 1998; Chintagunta, 2001; Villas-Boas & Winer, 1999). For instance, we do not observe shelf-space; however, increasing shelf-space allocation typically incurs costs that raise prices, such as allocation fees and opportunity costs. At the same time, it is well known that shelf-space influences consumer brand choices (Dreze, Hoch & Purk, 1994). While characterizing the precise nature of measurement error in our data is beyond the scope of the paper, we use standard instrumental variable techniques to avoid estimation biases.

In order to facilitate the direct instrumentation of prices, we use the inversion procedure proposed by Berry (1994). We begin by partitioning the observed product characteristics as $X_{jt} = [x_{jt}, p_{jt}]$, where by assumption $E(x\xi_{jt} | x_{jt}) = 0$ and $E(p\xi_{jt} | p_{jt}) \neq 0$. Following Berry (1994), we invert (1) to recover the vector $\delta_t(\Theta)$ of mean utilities as a function of the parameter vector $\Theta$, and we set up the estimation procedure in terms of $\delta_t$. Since the inverse of (1) does not have a simple analytical form, we resort to numerical inversion. In particular, we use the contraction-mapping of BLP (1995). The approach requires, for each $t$, picking some initial guess of the mean utility vector $\delta_t$ and iterating (4) until the following $J$ expressions converge:

$$\delta_{jt}^{n+1} = \delta_{jt}^{n} + \ln(q_{jt}) - \ln[\bar{q}_{jt}(X_t, \delta_t^n; \Theta)], j = 1, \ldots, J \tag{4}$$

where the superscript $n$ refers to an iteration, and $q_{jt}$ is the observed market share for brand $j$ in period $t$.

The advantage of using $\delta_{jt}$ for estimation is that the prediction error, $\delta_{jt} - X_{jt}\beta$, is simply the unobserved product characteristic, $\xi_{jt}$. The fact that $\xi_{jt}$ enters (4) linearly facilitates instrumentation. Moreover, with some intuition for the source of the unobserved attribute, we are able to impose reasonable covariance restrictions to set up our method of moments procedure.

We now set up a generalized method of moments (GMM) procedure to estimate the system of mean utilities. Let $\xi_t$ be the $(J \times 1)$ matrix of unobserved attributes for each of the products in store-week $t$. Similarly, we define our instruments, $Z_t$, an $I$-dimensional vector including the exogenous product characteristics as well as other potential covariates that may be correlated with $p_{jt}$, but not with $\xi_{jt}$ (We describe these variables in the data section). Our key identifying condition is the conditional mean-independence assumption $E(\xi_t \otimes Z_t | Z_t) = 0$ and $E(\xi_t \xi_t' | Z_t) = \Psi$ a finite $(J \times J)$ matrix. We are now able to construct our moment conditions:

$$h_t(\Theta) = \xi_t \otimes Z_t,$$

where at the true parameter values, $\Theta_0$, $E(h_t(\Theta_0)) = 0$. For estimation, we compute the corresponding sample analogue of these moment conditions:

$$h_T(\Theta) = \frac{1}{T} \sum_{t=1}^{T} \xi_t \otimes Z_t. \tag{5}$$

Our goal is to find values of $\Theta$ close enough to $\Theta_0$ to set the sample moments as close as possible to zero. We estimate $\Theta$ by minimizing the following quadratic expression:

$$G(\Theta) = (h_{JT}(\Theta))' \mathbf{W} (h_{JT}(\Theta)).$$

The matrix $\mathbf{W}$ is a $(JT \times JT)$ weight matrix. Hansen (1982) shows that the most efficient choice of $\mathbf{W}$ is a consistent estimate of the inverse of the variance of the moment conditions:

$$\mathbf{W} = E\{(h_T(\Theta))(h_T(\Theta))'\}$$
$$= E\{\xi_t \xi_t' \otimes Z_t Z_t'\}.$$

We obtain such an estimate by first estimating with homoscedastic errors to compute $W$.

While we assume $\{\xi_t\}_t$ is i.i.d. for estimation purposes, misspecifying its dependence structure will only affect the efficiency, not the consistency of our estimates.[4] We should also point out that in simulating the market shares (1), we effectively simulate the moments used for estimation, (5). McFadden (1989) and Pakes and Pollard (1989) both show that the method of simulated moments (MSM) still produces consistent estimates. However, the efficiency of these estimates is reduced due to simulation error. Only with sufficiently many simulation draws can one reach asymptotic efficiency with MSM. We use 30 draws and assume this number is sufficient to eliminate any noticeable simulation noise. Alternatively, one could implement variance-reducing simulation methods as in BLP (1995).

## 4. DATA

We use data from Dominick's Finer Foods (DFF), which is the second largest supermarket chain in the Chicago metropolitan area. DFF operates close to 100 stores in the Chicago area. The data consist of weekly sales, prices, promotions, and profit margins at the individual UPC-level for the liquid laundry detergent category. We focus on 52 weeks of data in the year 1992. Our data are for 83 weeks for which data on all 52 weeks are available. We present descriptive statistics for those products included in the analysis in Table 1. These data consist of means across store-weeks. We also report the standard deviation of prices across store weeks.[5] Note that the variable WP denotes wholesale prices – the prices paid by Dominicks to its suppliers. There are five brands included in the analysis. Three of these brands (Wisk, Surf & All) are marketed by Lever Brothers. The other two, Tide and Cheer are marketed by Procter & Gamble. Tide is the largest brand in the category. The table also indicates that brands are sold in different sizes – most typically 64 ounces and 128 ounces. In the analysis, instead of having a separate intercept for each brand-size combination, we include 5 brand dummies (one each for Surf, Wisk, All, Cheer & Tide with the outside good serving as the base) and one size dummy to account for

***Table 1.*** Descriptive Statistics (Laundry Detergent).

| Product | Size | Unit Share | Price/unit | Std Price/unit | WP/unit | Prom |
|---|---|---|---|---|---|---|
| Surf | 64 | 6.2% | 4.09 | 0.31 | 3.01 | 0.27 |
| Wisk | 128 | 7.0% | 8.10 | 0.89 | 6.62 | 0.13 |
| Wisk | 64 | 14.1% | 4.14 | 0.45 | 3.53 | 0.17 |
| All | 64 | 12.8% | 3.11 | 0.18 | 2.41 | 0.24 |
| All | 128 | 10.9% | 5.72 | 0.51 | 4.37 | 0.13 |
| Cheer | 64 | 6.3% | 4.20 | 0.23 | 3.62 | 0.16 |
| Cheer | 128 | 5.3% | 8.20 | 0.50 | 6.83 | 0.25 |
| Tide | 128 | 18.9% | 8.27 | 0.74 | 7.03 | 0.41 |
| Tide | 64 | 18.4% | 4.39 | 0.38 | 3.79 | 0.24 |

size differences (1 = 64 ounces, 0 otherwise). Interestingly, we find that while Tide is the largest share brand (units-wise), it is also the highest priced brand and is promoted most often from among the set of brands analyzed. The promotion variable is an indicator for whether the given product had an in-aisle display or newspaper feature that week. In the appendix, we provide a precise description of how we construct the relevant brands for analysis.

We supplement our store data with an extensive set of descriptive variables, from Spectra (see Hoch et al., 1995), characterizing the underlying consumer base and local competition associated with each store. ZIP code level demographic data was obtained from the 1990 census. The following four criteria were used in selecting the demographic variables – prior research, significance in the homogenous models, multicollinearity, and managerial relevance. For example, while we had data on the median income in the ZIP code, we chose the variable HVAL150 (percent of homes with a value over $150,000) as a proxy for income because income was highly correlated with other included variables. Of the five demographic variables that were used, only SHOPINDX (ability to shop – percent of population with car and single family house) and ETHNIC (percent of population that are Black or Hispanic) have a correlation of over 0.5. AGE60 (percent of population over age 60) also represents the retired variable (correlation of 0.88). The final demographic variable included HHLARGE is the percentage of households with five or more members. The two competitive variables used in the study are distance from the nearest Jewel (the largest supermarket in the area) and minimum of the distance from the nearest Cubfoods and Omni (the two main EDLP operations). Our initial models had also included variables on competitor volume but these had limited explanatory power and were dropped in the final models.

Recall that our estimation accounts for price endogeneity by instrumenting for prices. The set of instruments we use are as follows. We use the brand specific intercepts, the store characteristics, the promotional variables (that are assumed to be exogenous) and the wholesale prices of the various brands. Assuming promotions to be exogenous may appear to be counter-intuitive. However, our conversations with the store managers revealed that promotional decisions are typically made in advance of the weekly pricing decisions, thereby providing some support for our assumption that promotions are exogenous. Wholesale prices are likely to be correlated with retail prices. However, they are unlikely to be correlated with store specific factors such as store coupons, etc. As these latter factors are what we believe affect the $\xi_{jt}$, using wholesale prices as instruments appears to be a reasonable assumption.

Summary statistics for the demographic and competitive variables are provided in Table 2. We find considerable variation in the demographic and

***Table 2.*** Demographic and Competitive Variables.

| Variable | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|
| AGE60 | 17% | 6% | 6% | 31% |
| ETHNIC | 15% | 19% | 2% | 99% |
| HHLARGE | 12% | 3% | 1% | 22% |
| HVAL150 | 34% | 24% | 0.40% | 92% |
| SHOPINDX | 74% | 24% | 0% | 99% |
| JEWELDIST | 1.29 (mi) | 0.86 | 0.06 | 3.96 |
| EDLPDIST | 5.03 (mi) | 3.48 | 0.13 | 17.85 |

competitive characteristics across stores. For example, DFF stores cater to market areas with Black and Hispanic representation ranging from 2 to 99% of the population. In terms of consumer wealth, the proportion of consumers in DFF markets with houses valued over $150,000 ranges from below 1% to 92%. In terms of competition, some stores are located right next to both rival supermarkets and warehouse stores. Others locate over 4 miles from the nearest Jewel and 18 miles from the nearest EDLP store. We expect these differences to generate noticeable variation in the nature of demand across stores. Moreover, we expect these differences to generate variation in the perceived distance between brands across stores' markets.

## 5. RESULTS

We now present our estimates for the demand parameters along with the resulting perceptual maps. Recall that our model formulation consists of the following sets of parameters. (a) The 5 brand specific intercepts and one size dummy. (b) The mean price effect and promotion effect (c) The heterogeneity in price sensitivity parameter ($\lambda$). (d) The effects of store characteristics on the category purchase shares ($\gamma$ in equation (3)). (e) The mean values of the latent brand attributes, $L_j$ for brand $j$. (f) The interaction effects between the latent attributes and the store characteristics ($\theta$ in equation (3)) that allow for store specific inter-brand preference correlations. We discuss these sets of parameter estimates in turn. We estimate two model specifications. In the first, we assume that the variance in tastes, and hence the perceptual maps, are constant across stores. In the second, we allow for the variation in tastes to vary across stores. Model 1 therefore, only includes effects (a) through (e). Model 2 accounts for effects (a) through (f).

We summarize the estimation results for the set of parameters (a) through (d) for both model specifications in Table 3. The rank ordering of the mean brand

***Table 3.*** Taste Parameters (Laundry Detergent).

| | Model I | | Model II | |
|---|---|---|---|---|
| Variable | Param | se | Param | se |
| Surf | −0.948192 | 0.2107 | −0.299 | 0.223 |
| Wisk | −0.231744 | 0.117 | 0.352 | 0.250 |
| All | −1.824576 | 0.0802 | −1.541 | 0.298 |
| Cheer | −0.411774 | 0.1717 | 0.030 | 0.218 |
| Tide | 0.964308 | 0.0934 | 1.516 | 0.221 |
| Size (64 oz) | 0.7356 | 0.0079 | 0.752 | 0.008 |
| Price | −12.144 | 0.0891 | −13.925 | 0.583 |
| Price std | 1.7224 | 0.1664 | 2.205 | 0.395 |
| Prom | 0.1153 | 0.0095 | 0.085 | 0.011 |
| AGE60 | 1.0993 | 0.0826 | 1.070 | 1.093 |
| ETHNIC | −0.4839 | 0.0422 | −0.680 | 0.169 |
| HHLARGE | 0.771 | 0.2148 | 0.173 | 1.245 |
| HVAL150 | 0.5596 | 0.0241 | 0.140 | 0.378 |
| SHOPINDX | 0.1605 | 0.0317 | −0.156 | 0.206 |
| JEWDIST | 0.0327 | 0.0054 | 0.044 | 0.016 |
| DISTEDLP | 0.0204 | 0.0015 | 0.025 | 0.004 |

intercepts appear to be fairly robust to the specification of variance. The coefficients for prices and promotions have the usual sign. We find that Tide is the highest-valued brand, on average, consistent with our earlier observation that it has the highest unit share while being the highest priced brand. We also find that overall category demand levels are lower in markets with a higher proportion of ethnic households. Further, levels of demand appear to be higher for stores farther away from competing chains' prime stores regardless of whether these are EDLP or high/low pricing stores (Jewel). Note that these results do not tell us whether proximity to a competitor will make the store more or less profitable. To assess the impact of proximity on market power, we would need to compute marginal effects subject to a model of category management.[6]

Next, we turn to the mean values of the latent brand attributes across stores. We assume that there are 2 latent attributes so the values corresponding to each brand can be interpreted as locations on a two-dimensional map. In Table 4, we present the set of parameters corresponding to these latent attributes (i.e. the set of parameters (e) above). We report only the results from Model 2 here. Note that Tide has been constrained to lie along attribute or dimension 1 to ensure rotational invariance of the derived map. First, we find that a brand's location

***Table 4.*** Common Component across stores for Latent Factors (Laundry Detergent).

| Brand | Dim 1 | se | Dim 2 | se |
|---|---|---|---|---|
| Surf | −0.153 | 0.322 | −0.774 | 0.272 |
| Wisk | 0.239 | 0.281 | −0.645 | 0.286 |
| All | −1.030 | 0.250 | −0.064 | 0.180 |
| Cheer | −1.066 | 0.206 | −0.311 | 0.166 |
| Tide | −0.806 | 0.143 | – | – |

does differ along the two dimensions. This implies that a one-factor or a single latent attribute would not have sufficed for these data to capture the nature of inter-brand correlations. On average, we find that All, Cheer and Tide are perceived to be similar along attribute 1, with Surf and Wisk being perceived similarly along this attribute (they are not statistically distinguishable from zero). Along dimension 2, we find Tide and All perceived as being similar. Surf and Wisk once again are close together with Cheer taking an intermediate location in between the two sets of brands. The results from this table therefore indicate that on average, consumers differ in the way they perceive the brands along the two attributes.

Having discussed the mean attribute locations, we now turn to the interactions between the latent attributes and the demographic variables (parameters identified as set (f) above). Again, these results are from Model 2 (model 1 does not incorporate these effects). Recall that the interactions allow us to obtain store-specific perceptual maps. Note that statistical significance in the interaction effects provide evidence of varying brand perceptions across store areas. In Table 5 we provide the parameter estimates and the standard errors of the brand-specific interaction effects. We find that the manner in which store characterisitics influence the latent attributes varies considerably across brands. The most significant interactions appear to be for the Tide brand. Here we find that three store characterisitics play a role in influencing latent attributes across stores. These variables are the proportion of large households in the store area, the proportion of houses with values exceeding $150,000 as well as the ability of consumers in the store area to shop. We also find some statistically significant effects for Cheer for the proportion of high-value households and the proportion of ethnic households and for All, the proportion of large households. Interestingly, consistent with our previous finding we do not find statistically significant effects for the large Lever brothers brands, Wisk and Surf.

Next, to fix our ideas about across store heterogeneity in perceptions, we plot the perceptual maps for 4 stores (identified as stores 62, 74, 89 and 103). The plots are provided in Figs 3, 4, 5 and 6. These stores were chosen because they varied along specific store characteristics. Store 62 has the highest proportion of households with residences valued over $150,000. Store 74 has the highest proportion of ethnic households. Store 89 is the opposite of store 62, it has the lowest proportion of families with homes valued over $150,000. Finally, store 103 has the lowest proportion of ethnic households. Our choice of these two variables is based on the significant interaction effects obtained in Table 4. In other words, we are interested in demonstrating that store differences in perceptions can be attributed to store characteristics.

***Table 5.*** Store-Specific Component for Latent Factors (Laundry Detergent).

| Brand | | Param | se | *t*-stat |
|---|---|---|---|---|
| Surf | AGE60 | 1.603 | 1.704 | 0.941 |
| | ETHNIC | −2.872 | 2.138 | −1.343 |
| | HHLARGE | −0.416 | 1.511 | −0.275 |
| | HVAL150 | 0.913 | 1.121 | 0.814 |
| | SHOPINDX | −0.422 | 2.026 | −0.208 |
| Wisk | AGE60 | 1.395 | 1.006 | 1.388 |
| | ETHNIC | 0.876 | 0.656 | 1.336 |
| | HHLARGE | 0.172 | 0.666 | 0.258 |
| | HVAL150 | 0.404 | 0.906 | 0.445 |
| | SHOPINDX | −0.263 | 0.996 | −0.265 |
| All | AGE60 | −0.142 | 2.841 | −0.050 |
| | ETHNIC | 2.663 | 4.035 | 0.660 |
| | HHLARGE | −5.492 | 2.399 | −2.290 |
| | HVAL150 | 4.296 | 2.859 | 1.503 |
| | SHOPINDX | −0.062 | 5.283 | −0.012 |
| Cheer | AGE60 | 0.155 | 0.533 | 0.291 |
| | ETHNIC | −1.161 | 0.591 | −1.967 |
| | HHLARGE | 0.364 | 0.425 | 0.855 |
| | HVAL150 | −0.713 | 0.350 | −2.038 |
| | SHOPINDX | −0.682 | 0.507 | −1.344 |
| Tide | AGE60 | −0.370 | 1.080 | −0.343 |
| | ETHNIC | 0.387 | 0.843 | 0.459 |
| | HHLARGE | −2.120 | 0.540 | −3.926 |
| | HVAL150 | −0.984 | 0.548 | −1.795 |
| | SHOPINDX | −3.198 | 0.741 | −4.313 |

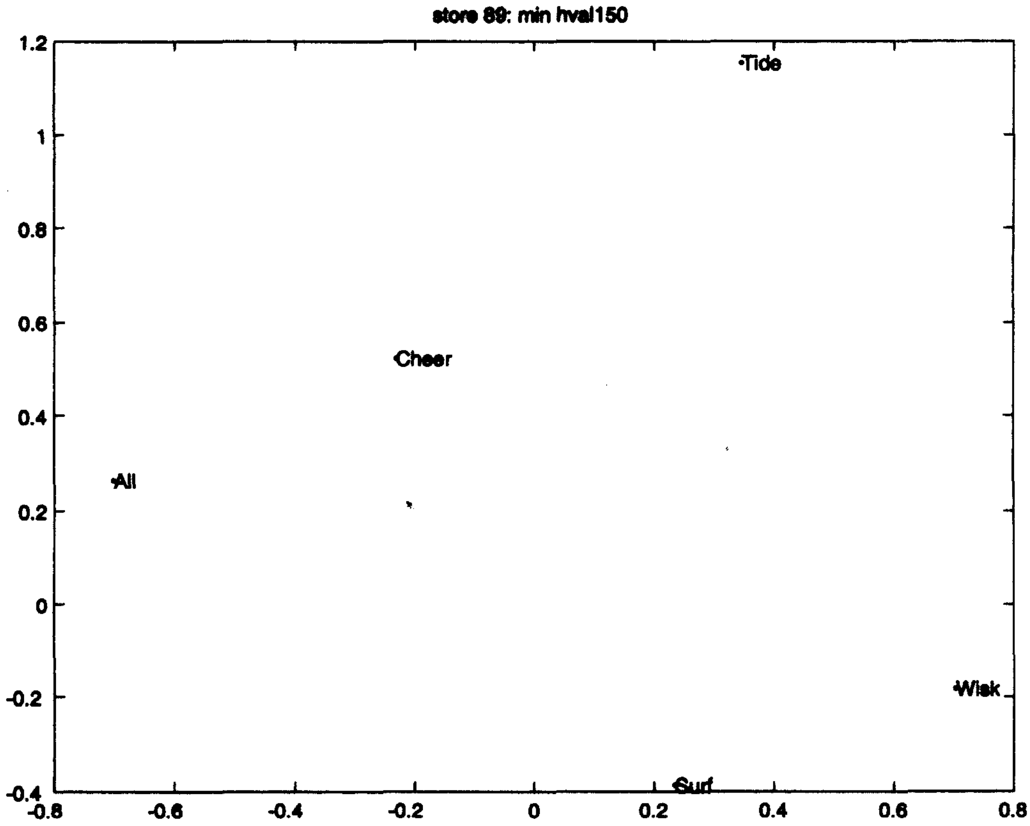


*Fig. 3.* Store 89: Lowest Proportion of Houses Valued Over $150,000.

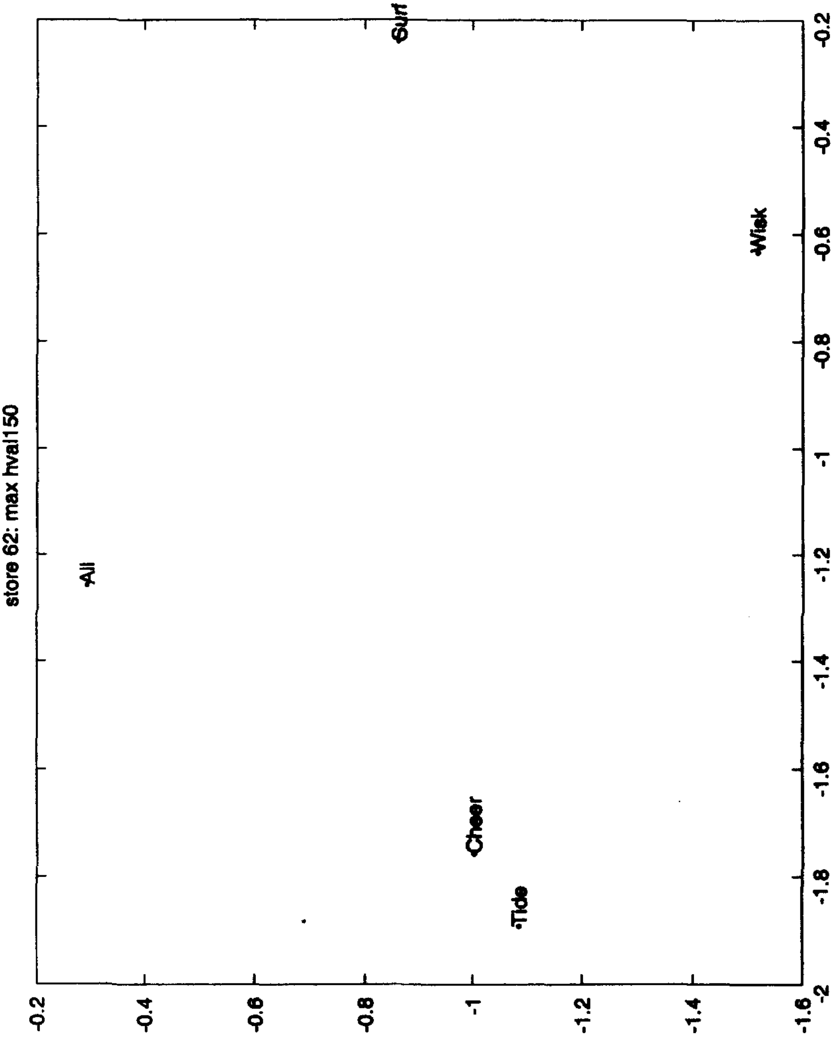


*Fig. 4.* Store 62: Highest Proportion of Houses Valued Over $150,000.

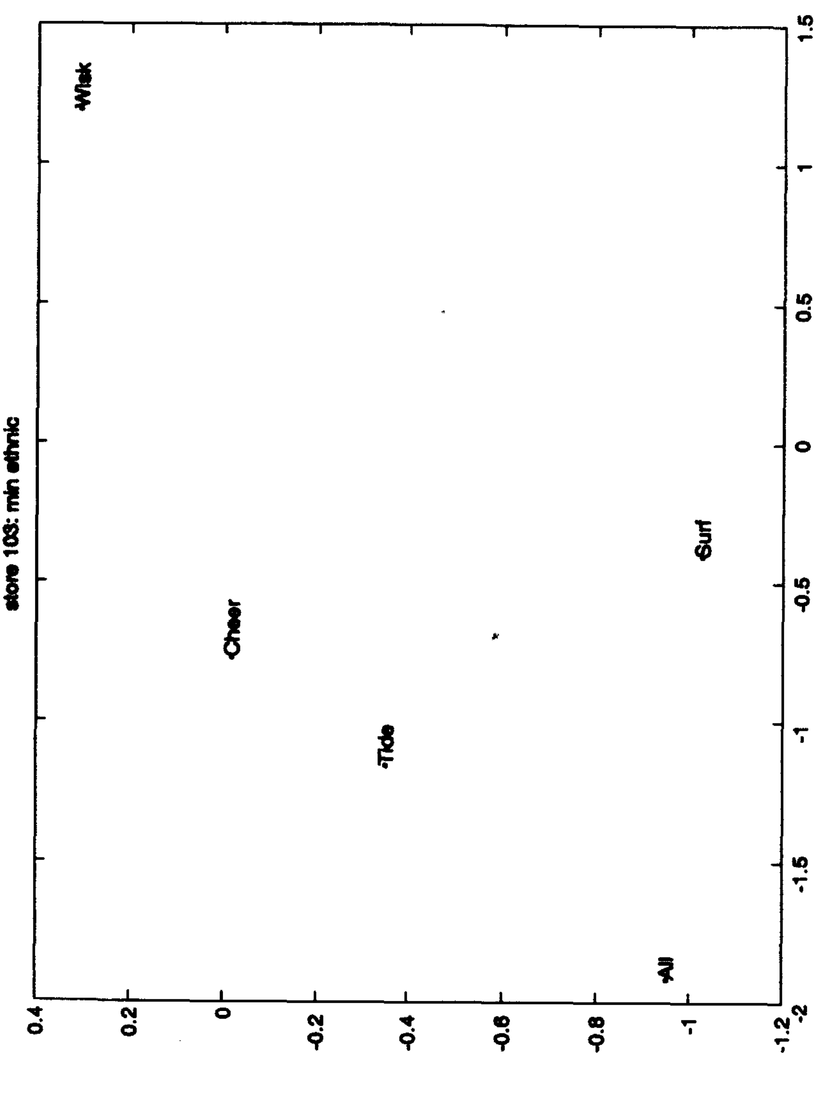


*Fig. 5.* Store 103: Lowest Proportion of Ethnic Households.

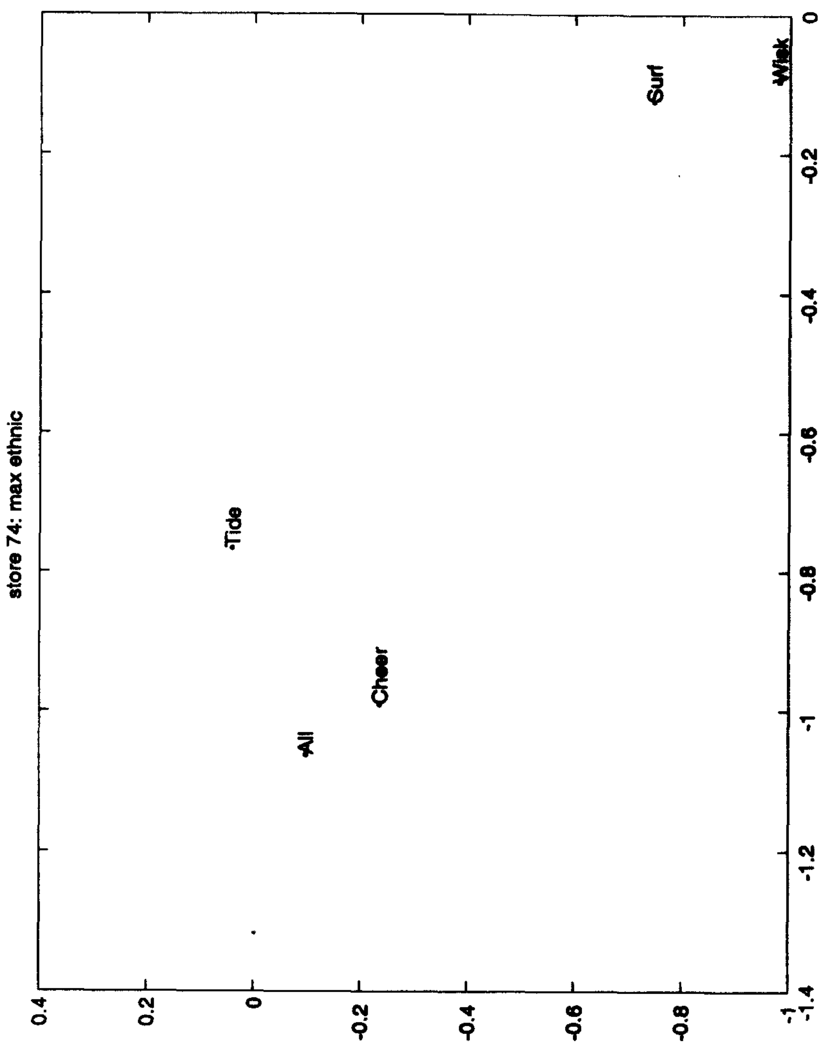


*Fig. 6.* Store 74: Highest Proportion of Ethnic Households.

Looking at Fig. 4 which provides the map for store 62, we find the following. First, the Procter & Gamble brands Tide and Cheer are perceived as being quite similar to one another, whereas the positioning of the Lever brothers brands is a lot more diffuse. Specifically, each of All, Surf and Wisk appear to have unique perceptions in the marketplace. Further, they seem to be perceived differently from the P&G brands. This store could be problematic for P&G as switching is most likely to occur within its own portfolio of brands.

We find that a similar pattern emerges when we look across two stores that differ in their proportions of ethnic households. Store 74, whose map is depicted in Fig. 6 has the highest propotion of ethnic households. Store 103 has the lowest proportion (Fig. 5). We find that for store 74, there are two distinct groupings of brands – one consisting of Tide, Cheer and All and the other with Surf and Wisk. In some sense for such neighborhoods, All seems to be working as a good fighter brand for Lever Brothers.When the proportion of ethnic households is small as in store 103 (Fig. 5), we find that preferences are a lot more diffuse in nature with all brands having fairly unique locations.

To summarize our findings from the various maps, statistically significant effects of interactions between the latent attributes and store characteristics do appear to translate into differences in perceptions of the brands along these attributes.

Next, we try to assess whether brand locations along the attributes are statistically significantly different across stores. In other words, is the location along an attribute, after including the effects of store characteristics, different across stores? To address this issue, we focus on the Tide brand. According to Table 5, this brand shows the most interactions with store characteristics. In Fig. 7, we plot the locations of the Tide brand in the 4 stores considered above – 62, 74, 89 and 103. Around each location, we provide the 95confidence region (i.e. two-dimensional confidence interval). The figure shows that while there is some overlap in these regions for pairs of stores, there are other store pairs for which there is no overlap. This implies that the locations of the brand is statistically significantly different at least across some store pairs.

## 6. CONCLUSION

We use standard choice models to estimate demand curves for a given category. By decomposing the covariance in consumer tastes for brands, we are able to recover perceptual maps. Unlike the existing literature, we use aggregate store-level data, treating weekly market shares as an aggregation of individual choices. The advantage of using aggregate data is that we are able to recover store-specific maps that allow for perceived brand locations to vary across local

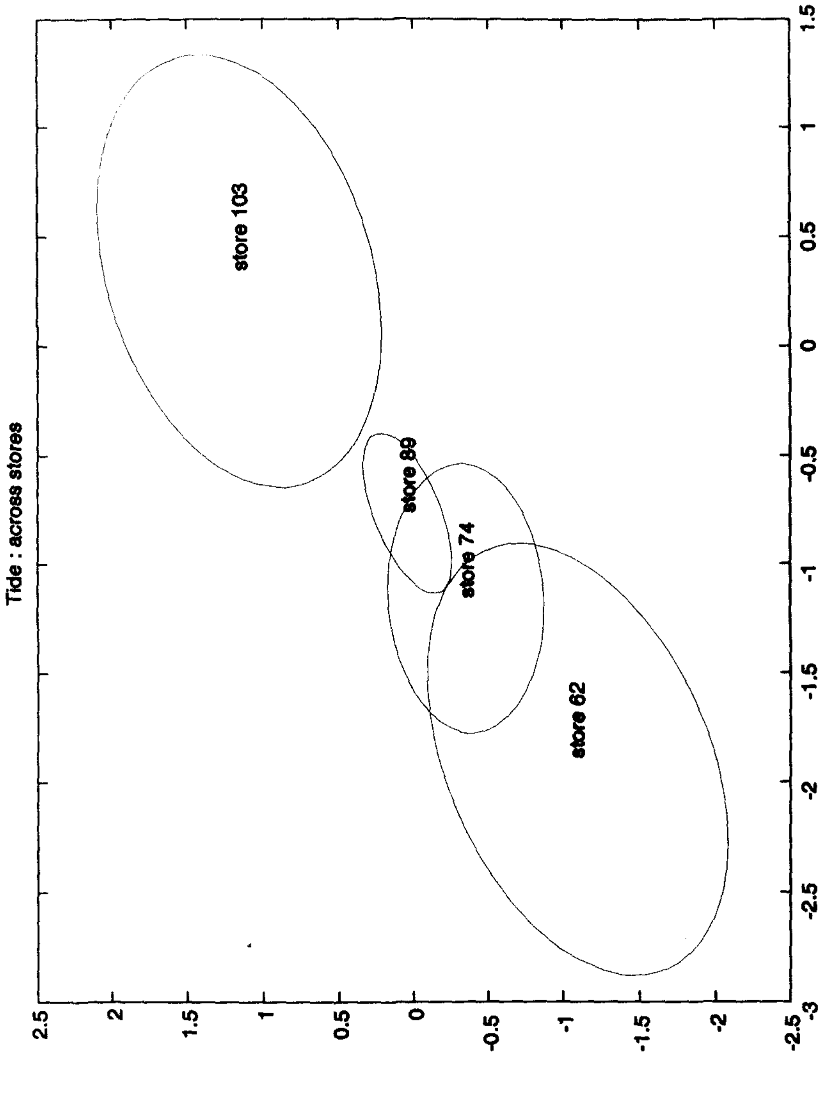


*Fig. 7.* Confidence Regions for Tide in Latent Attribute Space.

markets. Given the increasing emphasis on category management (Zenor, 1994) and micro-marketing (Hoch et al., 1995; Montgomery, 1997) store-specific perceptual maps enable retailers and manufacturers to target promotions to those markets in which they have the strongest positioning. Alternatively, one might target promotions to those markets in which the brand has the least perceived differentiation in a brand-building effort.

Applying the model to weekly sales of laundry detergents, we recover the demand system. From the parameters of the estimated demand system, we are able to plot the maps across stores. Our main finding is that the maps do vary substantially across stores. These differences could allow managers to consider targeting specific stores with promotions and pricing.

In the current work, we focus entirely on aggregate data. Despite our efforts to allow for taste distributions to vary across stores, we are limited to the use of mean Zip-code level data. Recently, stores are increasingly collecting their own microdata sets using loyalty cards, tracking individual purchases within specific stores. Future work might consider combining these micro data with the aggregate data into an integrated estimation framework (BLP, 1998 and Petrin, 2000). The combination of both data sources could allow for more sophisticated treatments of the store-specific brand locations and, in turn, the brand maps.

## NOTES

1. Since we estimate a full set of product fixed-effects, we do not need to worry about unmeasured physical product attributes, as in BLP 1995. We are concerned with unobserved weekly in-store product-specific effects.
2. For instance, Besanko, Dubé and Gupta (2001) use a finite-mixture model using comparable aggregate weekly store data.
3. One exception is Pesendorfer (2001), who models the timing of sales in a category as the outcome of inter-store competition.
4. We programmed the code for this estimation routine in MATLAB version 5.3.
5. Although not reported, we find that cross-store variation explains, on average, 15% of a brand's total price variation. For 64 oz Cheer, we find that cross-store variation explains over 45% of the total price variation. Thus, we expect our data to be capable of identifying significant cross-store effects.
6. We are currently working on this issue regarding the impact of proximity to competitors to store profitability.

## ACKNOWLEDGMENTS

The authors are grateful to Philip Hans Franses, Alan Montgomery and an anonymous reviewer for comments. The first two authors would also like to acknowledge the Kilts Center for Marketing at the University of Chicago for providing research funds. The second author acknowledges research support from the Beatrice Foods Faculty Fund at the Graduate School of Business. All correspondence may be addressed to the authors at the University of Chicago, Graduate School of Business, 1101 East 58th Street, Chicago, IL 60637 or via e-mail at jdube@gsb.uchicago.edu or pradeep.chintagunta@gsb.uchicago.edu.

## REFERENCES

Allenby, G. M. (1989). A Unified Approach to Identifying, Estimating and Testing Demand Structures with Aggregate Scanner Data. *Marketing Science*, *8*, 265–281.

Berry, S. (1994). Estimating Discrete-Choice Models of Product Differentiation. *Rand Journal of Economics*, *25*, 242–262.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, *63*, 841–890.

Berry, S., Levinsohn, J., & Pakes, A. (1998). Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. Yale University, Working Paper.

Besanko, D., Gupta, S., & Jain, D. (1998). Logit Demand Estimation Under Competitive Pricing Behavior: An Equilibrium Framework. *Management Science*, (November).

Besanko, D., Dubé, J.-P., & Gupta, S. (2001). Heterogeneity and Target Marketing Using Aggregate Retail Data: A Structural Approach. Working Paper, Kellogg Graduate School of Management.

Brownstone, D., & Train, K. (1998). Forecasting New Product Penetration with Flexible Substitution Patterns. *Journal of Econometrics*, *89*, 109–129.

Chintagunta, P. (1994). Heterogeneous Logit Model Implications for Brand Positioning. *Journal of Marketing Research*, *31*, 304–311.

Chintagunta, P. (1998). Inertia and Variety Seeking in a Model of Brand Purchase Timing. *Marketing Science*, *17*, 253–272.

Chintagunta, P. (2001). A Heterogeneous Aggregate Logit Demand Model. University of Chicago, Working Paper.

Chintagunta, P., & Honore, B. (1996). Investigating the effects of marketing variables and unobserved heterogeneity in a multinomial probit model. *International Journal of Research in Marketing*, *13*, 1–15.

Day, G. S., Shocker, A. D., & Srivastava, R. K. (1979). Customer-Oriented Approaches to Identifying Product-Markets. *Journal of Marketing*, *43*, 8–20.

Desarbo, W., & Rao, V. R. (1986). A Constrained Unfolding Methodology for Product Positioning. *Marketing Science*, *5*, 1–19.

Dreze, X., Hoch, S. J., & Purk, M. E. (1994). Shelf Management and Space Elasticity. *Journal of Retailing*, *70*, 301–326.

Elrod, T. (1988). Choice Map: Inferring a product market map from panel data. *Marketing Science*, *7*, 21–40.

Elrod, T., & Keane, M. (1995). A factor-analytic probit model for representing the market structure in panel data. *Journal of Marketing Research*, *32*, 1–16.

Erdem, T. (1996). A Dynamic Analysis of Market Structure Based on panel data. *Marketing Science*, *16*, 359–378.

Grover, R., & Srinivasan, V. (1987). A simultaneous approach to market segmentation and market structuring. *Journal of Marketing Research*, *24*, 139–153.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, *50*, 1092–1054.

Hauser, J. R., & Shugan, S. M. (1983). Defensive Marketing Strategies. *Marketing Science*, *2*, 319–360.

Hauser, J. R., & Wernerfelt, B. (1988). Competitive Price and Positioning Strategies; Existence and Uniqueness of Price Equilibria in Defender. *Marketing Science*, *7*, 76–94.

Hausman, J. A., & Wise, D. A. (1978). A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences. *Econometrica*, *46*, 403–426.

Hoch, S. J., Kim, B.-D., Montgomery, A. L., & Rossi, P. E. (1995). Determinants of store-level price elasticity. *Journal of Marketing Research*, *32*, 17–30.

Jain, D., Bass, F. M., & Chen, Y. M. (1990). Estimation Of Latent Class Models With Heterogeneous Choice. *Journal of Marketing Research*, *27*, 94–102.

Kadiyali, V., Chintagunta, P. K., & Vilcassim, N. J. (2000). Manufacturer-retailer channel interactions and implications for channel power: An empirical investigation of pricing in a local market. *Marketing Science*, *19*, 127–147.

Kannan, P. K., & Sanchez, S. M. (1994). Competitive market structure: A subset selection analysis. *Management Science*, *40*, 1484–1500.

Kannan, P. K., & Wright, G. P. (1991). On "Testing Competitive Market Structures". *Marketing Science*, *10*, 338–348.

Kamakura, W., & Russell, G. (1989). Probabilistic Choice Model for market segmentation and elasticity structure. *Journal of Marketing Research*, *26*, 379–390.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration. *Econometrica*, *57*, 995–1026.

McFadden, D., & Train, K. (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Economerics*, *15*, 447–470.

Montgomery, A. L. (1997). Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing Science*, *16*, 315–338.

Nevo, A. (2001). Measuring Market Power in the Ready-To-Eat Cereal Industry. *Econometrica*, *69*, 307–340.

Pakes, A., & Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, *57*, 1027–1057.

Pesendorfer, M. (2001). Retail Sales: A Study of Pricing Behavior in Supermarkets. *Journal of Business* (forthcoming).

Petrin, A. (1999). Quantifying the Benefits of New Products: The Case of the Minivan. University of Chicago. Working Paper.

Ramaswamy, V., & Desarbo, W. S. (1990). SCULPTRE: A New Methodology for Deriving and Analyzing Hiera. *Journal of Marketing Research*, *27*, 418–428.

Slade, M. E. (1995). Product Market Rivalry with Multiple Strategic Weapons: an analysis of price and advertising competition. *Journal of Economics and Management Strategy*, *4*, 445–476.

Shugan, S. M. (1987). Estimating Brand Positioning Maps Using Supermarket Scanning Data. *Journal of Marketing Research*, *24*, 1–18.

Urban, G. L., Johnson, P. L., & Hauser, J. R. (1984). Testing Competitive Market Structures. *Marketing Science*, *3*, 83–113.

Villas-Boas, M., & Winer, R. (1999). Endogeneity in Brand Choice Models. *Management Science*, *45*(10), 1324–1338.

Waarts, E., Carree, M., & Wierenga, B. (1991). Full-Information Maximum Likelihood Estimation of Brand Positioning Maps Using Supermarket Scanning Data. *Journal of Marketing Research*, *28*, 483–490.

Zenor, M. J. (1994). The profit benefits of category management. *Journal of Marketing Research*, *31*, 202–214.

# APPENDIX

*The Laundry Detergent Category*

The laundry detergent category is dominated by brands from two manufacturers (P&G and Unilever) that account for 84% of category sales. Tide is the market leader with a market share of 32%, followed by Wisk with a share of 16%. The store brand has a limited presence in this category and thus we omit it from our analysis. Although detergents are sold in a number of sizes, 64 and 128 oz account for over 80% category volume.

Despite the dominance by two manufacturers, the detergent category is highly competitive with over 100 UPCs and 15 brands. Such large number products dictate some aggregation across products. Further, selecting brands for empirical analysis requires a balance between category representation and aggregation bias. Our approach was to run a correlation of prices across stores and weeks, and bundling those UPCs within a brand-size that had a price correlation of over 0.8. In other words, we only aggregate across UPCs whose prices co-move highly enough to believe they are priced jointly. To be more concrete, we present the UPCs, their sizes, category shares, prices, and the price correlation for Wisk in Table 6. The items in bold appear in our empirical analysis. Overall, our empirical analysis uses data on 19 UPCs and 9 brand-size combinations. The included products account for 61% of the category sales.

***Table 6.*** Aggregation Across UPCs.

| UPC | Product | Size (oz) | Price | Cat. Share | Price Correlation | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1111187202 | WISK POWDER | 128 | 8.45 | 0.3% | 1.00 | | | | | |
| 1111187404 | NP WISK LIQ DET | 128 | 8.10 | 4.4% | 0.79 | 1.00 | | | | |
| 1111187405 | WISK LIQUID | 96 | 6.35 | 1.5% | 0.57 | 0.17 | 1.00 | | | |
| 1111187406 | NP WISK LIQ DET | 64 | 4.13 | 5.4% | 0.13 | 0.01 | 0.24 | 1.00 | | |
| 1111187410 | WISK UNSCENTED | 64 | 4.16 | 2.7% | 0.12 | –0.02 | 0.19 | 0.98 | 1.00 | |
| 1111187412 | WISK LIQUID | 32 | 2.54 | 1.3% | 0.61 | 0.29 | 0.67 | 0.31 | 0.27 | 1.00 |

# ECONOMETRIC ANALYSIS OF THE MARKET SHARE ATTRACTION MODEL

Dennis Fok, Philip Hans Franses and Richard Paap

## ABSTRACT

*Market share attraction models are useful tools for analyzing competitive structures. The models can be used to infer cross-effects of marketing-mix variables, but also the own effects can be adequately estimated while conditioning on competitive reactions. Important features of attraction models are that they incorporate that market shares sum to unity and that the market shares of individual brands are in between 0 and 1. Next to analyzing competitive structures, attraction models are also often considered for forecasting market shares.*

*The econometric analysis of the market share attraction model has not received much attention. Topics as specification, diagnostics, estimation and forecasting have not been thoroughly discussed in the academic marketing literature. In this chapter we go through a range of these topics, and, along the lines, we indicate that there are ample opportunities to improve upon present-day practice. We also discuss an alternative approach to the log-centering method of linearizing the attraction model. This approach leads to easier inference and interpretation of the model.*

**Econometric Models in Marketing, Volume 16, pages 223–256.**

**ISBN: 0-7623-0857-5**

# 1. INTRODUCTION

The implementation of econometric models has become increasingly fashionable in marketing research. The main reason for this is that nowadays marketing research can involve the analysis of large amounts of data on revealed preferences, such as sales, market shares, brand choices and interpurchase times, and stated preferences such as opinions, attitudes and purchase intentions. Many firms collect data on these performance measures for their current and their prospective customers, and they usually try to relate these measures with individual-specific characteristics and marketing-mix efforts. See Leeflang et al. (2000) and Franses and Paap (2001a) for recent surveys on quantitative models for revealed preference data. The main reason for considering econometric models is that in many cases the number of data points and the number of variables is rather large, and hence simply performing a range of bivariate analyses seems impractical.

The econometric analysis of a certain model for the above mentioned measures usually involves a range of steps. The first step amounts to specifying a model given the available data, the relevant explanatory variables, and the marketing problem at hand. Once the model has been specified, one needs to estimate the parameters and their associated confidence regions. Third, one usually considers the empirical validity of the model by performing diagnostic tests on its adequacy, where one typically focuses on the properties of the unexplained part of the model. Given the potential availability of two or more adequate rival models, one seeks to compare these models either on within-sample fit or on out-of-sample forecasting performance. Finally, one can use the ultimately obtained model for forecasting or for policy analysis. It should be noted that the focus in econometric textbooks tends to be on parameter estimation, but it is by no means the single most important issue. Indeed, in practice it is often difficult to specify the model and to compare it with alternatives.

In this chapter we will consider the econometric analysis of a popular model in marketing research, which is the market share attraction model. This model is typically considered for data on market shares, where the data have been collected at a weekly or monthly interval. Market share attraction models are seen as useful tools for analyzing competitive structures, see Cooper and Nakanishi (1988) and Cooper (1993), among various others. The models can be used to infer cross-effects of marketing-mix variables, but one can also learn about the effects of own efforts while conditioning on competitive reactions.

Important features of attraction models are that they rightfully assume that market shares sum to unity and that the market shares of individual brands are in between 0 and 1. This complicates the econometric analysis, as we will see below. Typically, an attraction model can be written as a system of equations concerning all market shares, and the parameters can be estimated using standard methods, see for example Cooper (1993) and Bronnenberg et al. (2000).

Interestingly, a casual glance at the relevant marketing literature on market share attraction models indicates that there seem to have been little attention on how to specify the attraction model, how to estimate its parameters, how to analyze its virtues in the sense that the models capture the salient data characteristics, and about how to use the models for forecasting. In sum, it seems that an (empirical) econometric view in these models is lacking. Therefore, in this chapter we aim to contribute to this view by addressing these issues concerning attraction models when they are to be used for describing and forecasting market shares. The first issue concerns the specification of the models. A literature check immediately indicates that many studies simply assume one version of an attraction model to be relevant and start from there. In this chapter we first start with a fairly general and comprehensive attraction model, and we show how various often applied models fit into this general framework. We also indicate how one can arrive from the general model at the more specific models, thereby immediately suggesting a general-to-simple testing strategy. Second, we discuss the estimation of the model parameters. We show that a commonly advocated method is unnecessarily complicated and that a much simpler method yields equivalent estimates. Along these lines, we also propose a few diagnostic measures, which to our knowledge have rarely been used, but which really should come in handy. Finally, we address the issue of generating forecasts for market shares. As the market share attraction model ultimately gets analyzed as a system of equations for (natural) log transformed shares, generating unbiased forecasts is far from trivial. We discuss a simulation-based method which yields unbiased forecasts.

The outline of this chapter is as follows. In Section 2, we first discuss the basics of the attraction model by reviewing various specifications of the model. We discuss the interpretation of the model in Section 3, and we discuss parameter estimation of the model in Section 4. We discuss diagnostic measures in Section 5. We touch upon the topic of model selection in Section 6. Forecasting in the attraction model is discussed in Section 7. In Section 8, we illustrate some of the techniques using scanner data. We conclude in Section 9 with suggestions for further research.

## 2. REPRESENTATION

In this section we start off with discussing a general market share attraction model and we deal with various of its nested versions which currently appear in the academic marketing literature. We first start with the so-called fully extended attraction model in Section 2.1. This model has a flexible structure as it includes many variables. Naturally this increases the empirical uncertainty about the relevant parameters. Therefore, in practice one may want to consider restricted versions of this general model. In Section 2.2, we discuss some of the restricted versions, where we particularly focus on those models which are often applied in practice.

### *2.1. A General Market Share Attraction Model*

Let $A_{i,t}$ be the attraction of brand $i$ at time $t$, $t = 1, \ldots, T$, given by

$$A_{i,t} = \exp(\mu_i + \varepsilon_{i,t}) \prod_{j=1}^{I} \prod_{k=1}^{K} x_{k,j,t}^{\beta_{k,j,i}} \qquad \text{for } i = 1, \ldots, I, \tag{1}$$

where $x_{k,j,t}$ denotes the $k$-th explanatory variable (such as price level, distribution, advertising spending) for brand $j$ at time $t$ and where $\beta_{k,j,i}$ is the corresponding coefficient for brand $i$. The parameter $\mu_i$ is a brand-specific constant. Let the error term $(\varepsilon_{1,t}, \ldots, \varepsilon_{I,t})'$ be normally distributed with zero mean and $\Sigma$ as a possibly non-diagonal covariance matrix, see Cooper and Nakanishi (1988). As we want the attraction to be non-negative, $x_{k,j,t}$ has to be non-negative, and hence rates of changes are usually not allowed. The variable $x_{k,j,t}$ may be a 0/1 dummy variable to indicate promotional activities for brand $j$ at time $t$. Note that for this dummy variable, one should transform $x_{k,j,t}$ to $\exp(x_{k,j,t})$ to avoid that $A_{i,t}$ becomes zero in case of no promotional activity.

The attraction specification in (1) is known as the Multiplicative Competitive Interaction [MCI] specification. A more general version of the attraction model uses a transformation of the explanatory variables; that is, it uses $f(x_{k,j,t})$ instead of $x_{k,j,t}$. When $f(\cdot)$ is taken to be the exponential function one obtains a specification known as the Multinomial Logit [MNL] specification. The difference between the MCI and the MNL specification is the assumed pattern of the elasticity of marketing instruments. The MCI specification assumes that the elasticity declines with increasing values of the explanatory variable, while the MNL specification assumes the elasticity increases up to a specific level and then decreases. The ultimate choice of a specification therefore depends on the marketing instruments used. The MNL specification seems to be appropriate for advertising spending, while the MCI specification would better fit pricing, see Cooper (1993) or Cooper and Nakanishi (1988) for elaborate discussions on the choice of $f(\cdot)$. In order not to complicate matters, we only consider the MCI specification, but note that all results can be extended to the MNL specification.

The market shares for the $I$ brands follow from the, what is called, Market Share Theorem, see Bell et al. (1975). This theorem states that the market share of brand $i$ is equal to its attraction relative to the sum of all attractions, that is,

$$M_{i,t} = \frac{A_{i,t}}{\sum_{j=1}^{I} A_{j,t}} \qquad \text{for } i = 1, \ldots, I. \tag{2}$$

The model in (1) with (2) is usually called the market share attraction model. Notice that the definition of the market share of brand $i$ at time $t$ given in (2) implies that the attraction of the product category is the sum of the attractions of all brands and that $A_{i,t} = A_{l,t}$ results in $M_{i,t} = M_{l,t}$.

The interesting aspect of the attraction model is that the $A_{i,t}$ in (1) is unobserved. As we will see below, this implies that neither $\mu_i$ nor $\beta_{k,j,i}$ is identified. Another consequence is that the market researcher should make a decision on the specification of $A_{i,t}$ prior to empirical analysis. As we will indicate, there are many possible specifications. For example, to describe potential dependencies in market shares over time, which describe purchase reinforcement effects, one may include lagged attractions $A_{i,t}$ in (1). For example, one may consider

$$A_{i,t} = \exp(\mu_i + \varepsilon_{i,t}) A_{i,t-1}^{\gamma_i} \prod_{j=1}^{I} \prod_{k=1}^{K} x_{k,j,t}^{\beta_{k,j,i}}. \tag{3}$$

However, due to the fact that we do not observe $A_{i,t}$, it turns out only possible to estimate the parameters in this model if the lag parameter $\gamma_i$ is assumed to be the same across brands, see Chen et al. (1994). As this may be viewed as too restrictive, an alternative strategy to account for dynamics is to include lagged values of the observed variables $M_{j,t}$ and $x_{k,j,t}$ in (1). The most general autoregressive structure follows from the inclusion of lagged market shares and lagged explanatory variables of all brands. In that case, the attraction specification with a $P$-th order autoregressive structure becomes

$$A_{i,t} = \exp(\mu_i + \varepsilon_{i,t}) \prod_{j=1}^{I} \left( \prod_{k=1}^{K} x_{k,j,t}^{\beta_{k,j,i}} \prod_{p=1}^{P} \left( M_{j,t-p}^{\alpha_{p,j,i}} \prod_{k=1}^{K} x_{k,j,t-p}^{\beta_{p,k,j,i}} \right) \right), \tag{4}$$

where the $\alpha_{p,j,i}$ parameters represent the effect of lagged market shares on attraction and where the $\beta_{p,k,j,i}$ parameters represent the effect of lagged explanatory variables. To illustrate, this model allows that the market share for brand 1 at $t-1$ has an effect on that of brand 2 at $t$, and also that there is a relationship between brand 2's market share and the price of brand 1 at $t-1$. The lagged endogenous variables capture dynamics in purchase behavior that cannot be attributed to specific marketing instruments. For example, consider state dependence in behavior. If brand $i$ is purchased at time $t$ by consumers who act state dependent, there will be a higher probability that they will purchase brand $i$ again at time $t+1$. Whether the brand was chosen at time $t$ because it was promoted or just by chance does not influence the dynamics in the behavior. On the other hand, part of the dynamics in the behavior can be attributed to specific marketing instruments. As an example, consider price promotions. A well-known feature of promotions is the post-promotional dip, see Van Heerde et al. (2000). In the period after a promotion it is often observed that sales or market shares decrease temporarily, as due to the promotion there has been stock piling by the consumers. To capture such dynamic patterns we include lagged exogenous variables in our attraction specification.

The flexibility of this general specification is reflected by the potentially large number of parameters. For example with $I=4$ brands, $K=3$ explanatory variables and $P=2$ lags, there are over 150 parameters to estimate (although they are not all identified, see below). It is however not necessary that the order $P$ for the lagged market shares and lagged explanatory variables is the same. To obtain a different lag order for the explanatory variables, one can restrict the corresponding $\beta_{p,k,j,i}$ parameters to be zero.

The model that consists of Eqs (4) and (2) is sometimes called the fully extended multiplicative competitive interaction [FE-MCI] model, see Cooper (1993). To enable parameter estimation, one can linearize this model in two steps. First, one can take one brand as the benchmark brand. Choosing brand $I$ as the base brand leads to

$$\frac{M_{i,t}}{M_{I,t}}=\frac{\exp(\mu_i+\varepsilon_{i,t})\prod_{j=1}^{I}\left(\prod_{k=1}^{K}x_{k,j,t}^{\beta_{k,j,i}}\prod_{p=1}^{P}\left(M_{j,t-p}^{\alpha_{p,j,i}}\prod_{k=1}^{K}x_{k,j,t-p}^{\beta_{p,k,j,i}}\right)\right)}{\exp(\mu_I+\varepsilon_{I,t})\prod_{j=1}^{I}\left(\prod_{k=1}^{K}x_{k,j,t}^{\beta_{k,j,I}}\prod_{p=1}^{P}\left(M_{j,t-p}^{\alpha_{p,j,I}}\prod_{k=1}^{K}x_{k,j,t-p}^{\beta_{p,k,j,I}}\right)\right)}. \tag{5}$$

In Section 4.2, we will discuss another approach to linearizing the model, but we will show that both transformations lead to the same parameter estimates, while the estimation procedure based on (5) is much simpler. Next, one can take the natural logarithm (denoted by log) of both sides of (5). Together, this results in the $(I-1)$-dimensional set of equations given by

$$\log M_{i,t} - \log M_{I,t} = (\mu_i - \mu_I) + \sum_{j=1}^{I} \sum_{k=1}^{K} (\beta_{k,j,i} - \beta_{k,j,I}) \log x_{k,j,t}$$
$$+ \sum_{j=1}^{I} \sum_{p=1}^{P} \left( (\alpha_{p,j,i} - \alpha_{p,j,I}) \log M_{j,t-p} + \sum_{k=1}^{K} (\beta_{p,k,j,i} - \beta_{p,k,j,I}) \log x_{k,j,t-p} \right) + \eta_{i,t}, \tag{6}$$

for $i = 1, \ldots, I-1$. Note that not all $\mu_i$ parameters $(i = 1, \ldots, I)$ are identified. Also for each $k$ and $p$, one of the $\beta_{k,j,i}$ and $\beta_{p,k,j,i}$ parameters is not identified. In fact, only the parameters $\tilde{\mu}_i = \mu_i - \mu_I$, $\tilde{\beta}_{k,j,i} = \beta_{k,j,i} - \beta_{k,j,I}$ $\tilde{\beta}_{p,k,j,i} = \beta_{p,k,j,i} - \beta_{p,k,j,I}$ are identified. This is however sufficient to completely identify elasticities, see Section 3 below and Cooper and Nakanishi (1988, p. 145). Finally, one can only estimate the parameters $\tilde{\alpha}_{p,j,i} = \alpha_{p,j,i} - \alpha_{p,j,I}$.

The error variables in (6) are $\eta_{i,t} = \varepsilon_{i,t} - \varepsilon_{I,t}$, $i = 1, \ldots, I-1$. Hence, given the earlier assumptions on $\varepsilon_{i,t}$, $(\eta_{1,t}, \ldots, \eta_{I-1,t})'$ is normally distributed with mean zero and $((I-1) \times (I-1))$ covariance matrix $\tilde{\Sigma} = L\Sigma L'$, where $L = (\mathbf{I}_{I-1} \vdots \mathbf{i}_{I-1})$ with $\mathbf{I}_{I-1}$ an $(I-1)$-dimensional identity matrix and where $\mathbf{i}_{I-1}$ is an $(I-1)$-dimensional unity vector. Note that therefore only $\frac{1}{2} I(I-1)$ parameters of the covariance matrix $\Sigma$ can be identified.

In sum, the general attraction model can be written as a $(I-1)$-dimensional $P$-th order vector autoregression with exogenous variables [sometimes abbreviated as VARX(P)], given by

$$\log M_{i,t} - \log M_{I,t} = \tilde{\mu}_i + \sum_{j=1}^{I} \sum_{k=1}^{K} \tilde{\beta}_{k,j,i} \log x_{k,j,t}$$
$$+ \sum_{j=1}^{I} \sum_{p=1}^{P} \left( \tilde{\alpha}_{p,j,i} \log M_{j,t-p} + \sum_{k=1}^{K} \tilde{\beta}_{p,k,j,i} \log x_{k,j,t-p} \right) + \eta_{i,t}, \tag{7}$$

$i = 1, \ldots, I-1$, where the covariance matrix of the error variables $(\eta_{1,t}, \ldots, \eta_{I-1,t})'$ is $\tilde{\Sigma}$. Note that the model is only valid for the observations starting at time $t = P+1$. For inference, it is common practice to condition on the first $P$ initial values of the log market shares and the explanatory variables as is also done in vector autoregressions, see Lütkepohl (1993). For further reference, we will consider (7) as the general attraction specification. We will take it as a starting point in our within-sample model selection strategy, which follows the general-to-specific principle, see Section 6 below.

It is not possible to write the log market shares in (7) as a function of current and lagged explanatory variables and disturbances only. It is even not possible to solve (7) for $\log M_{i,t} - \log M_{I,t}$. This is mainly due to the complex dynamic structure. This means that it is difficult to derive restrictions for stationarity of the log market shares themselves. In practice, this may not be a serious problem. Indeed, Srinivasan and Bass (2000) and Franses et al. (2001) consider testing for unit roots in market shares in a different model and their results suggest that generally market shares appear to be stationary. In Section 2.2, we show that if the dynamic specification is somehow restricted, it does become possible to solve (7) for log relative market shares.

## 2.2. *Various Restricted Models*

As can be understood from (7), the general attraction model contains many parameters and in practice this will absorb many degrees of freedom. Therefore, one usually assumes a simplified version of this general model. Obviously, the general model can be simplified in various directions, and, interestingly, the academic marketing literature indicates that in many cases one simply assumes some form without much further discussion. Selecting an appropriate model may be a non-trivial exercise, as there are many possible simpler models. One can for example impose restrictions on the $\beta$ coefficients, on the covariance structure $\Sigma$, and on the autoregressive parameters $\alpha$. In this section we will discuss a few of these potentially empirically relevant restrictions on the attraction specification in (4).

*Restricted Covariance Matrix [RCM]*

If the covariance matrix of the error variables $\varepsilon_{i,t}$ in (4) is a diagonal matrix, where each $\varepsilon_{i,t}$ has its own variance $\sigma_i^2$, that is, $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_I^2)$, then the covariance matrix for the $(I-1)$-dimensional vector $\eta_{i,t}$ in (7) becomes

$$\text{diag}(\sigma_1^2, \ldots, \sigma_{I-1}^2) + \sigma_I^2 \mathbf{i}_{I-1}\mathbf{i}'_{I-1}, \quad (8)$$

where $\mathbf{i}_{I-1}$ denotes a $(I-1)$-dimensional unity vector. In Section 6 we discuss how one can examine the validity of (8). If this restriction holds, the errors in the attraction specifications are independent, implying that the unexplained components of the attraction equations are uncorrelated.

*Restricted Competition [RC]*

One can also assume that the attraction of brand $i$ only depends on its own explanatory variables. This amounts to the assumption that marketing effects of competitive brands do not have an attraction effect, see for example Kumar

(1994) among others. For (4), this corresponds to the restriction $\beta_{k,j,i}=0$ (and $\beta_{p,k,j,i}=0$) for $j \neq i$. More precisely, this RC restriction implies that (4) reduces to

$$A_{i,t}=\exp(\mu_i+\varepsilon_{i,t})\prod_{k=1}^{K} x_{k,i,t}^{\beta_{k,i}}\prod_{j=1}^{I}\prod_{p=1}^{P}\left(M_{j,t-p}^{\alpha_{p,j,i}}\prod_{k=1}^{K} x_{k,i,t-p}^{\beta_{p,k,i}}\right) \qquad \text{for } i=1,\ldots,I, \tag{9}$$

where we write $\beta_{k,i}$ for $\beta_{k,i,i}$ and $\beta_{p,k,i}$ for $\beta_{p,k,i,i}$. Consequently, the linearized multiple equation model in (7) becomes

$$\log M_{i,t}-\log M_{I,t}=\tilde{\mu}_i+\sum_{k=1}^{K}\beta_{k,i}\log x_{k,i,t}-\sum_{k=1}^{K}\beta_{k,I}\log x_{k,I,t}$$
$$+\sum_{j=1}^{I}\sum_{p=1}^{P}\left(\tilde{\alpha}_{p,j,i}\log M_{j,t-p}+\sum_{k=1}^{K}\beta_{p,k,i}\log x_{k,i,t-p}-\sum_{k=1}^{K}\beta_{p,k,I}\log x_{k,I,t-p}\right)+\eta_{i,t} \tag{10}$$

for $i=1,\ldots,I-1$. Notice that this means that the coefficients $\beta_{k,I}$ are equal across the $(I-1)$ equations and that these restrictions should be taken into account when estimating the parameters. The RC assumption in (9) imposes $K(P+1)I(I-2)$ restrictions on the parameters in the general model in (7), which amounts to a substantial increase in the degrees of freedom. In Section 6 we will discuss how this restriction can be tested.

*Restricted Effects [RE]*

An even further simplified model arises if we assume, additional to RC, that the $\beta$ parameters are the same for each brand, that is, $\beta_{k,i}=\beta_k$ (and $\beta_{p,k,i}=\beta_{p,k}$), see Danaher (1994) for an implementation of this combined restrictive model. This model assumes that marketing efforts for brand $i$ only have an effect on the market share of brand $i$, and also that these effects are the same across brands. In other words, price effects, for example, are the same for all brands. It should be noted here that these similarities do not hold for *elasticities*, as will become apparent in Section 3. One may coin this model as an attraction model with restricted effects. Based on (4), the attraction for brand $i$ at time $t$ then further simplifies to

$$A_{i,t}=\exp(\mu_i+\varepsilon_{i,t})\prod_{k=1}^{K} x_{k,i,t}^{\beta_k}\prod_{j=1}^{I}\prod_{p=1}^{P}\left(M_{j,t-p}^{\alpha_{p,j,i}}\prod_{k=1}^{K} x_{k,i,t-p}^{\beta_{p,k}}\right) \qquad \text{for } i=1,\ldots,I, \tag{11}$$

and the linearized multiple equation model (7) simplifies to

$$\log M_{i,t} - \log M_{I,t} = \tilde{\mu}_i + \sum_{k=1}^{K} \beta_k(\log x_{k,i,t} - \log x_{k,I,t}) + \sum_{j=1}^{I} \left( \sum_{p=1}^{P} \tilde{\alpha}_{p,j,i} \log M_{j,t-p} + \sum_{k=1}^{K} \beta_{p,k}(\log x_{k,i,t-p} - \log x_{k,I,t-p}) \right) + \eta_{i,t} \quad (12)$$

for $i=1, \ldots, I-1$. This RE assumption imposes an additional $K(P+1)(I-1)$ parameter restrictions on the $\beta$ coefficients of (7). Of course, it may occur that the restrictions only hold for a few and not for all $\beta_{k,j,i}$ parameters, that is, for only a few marketing variables. In that case, less parameter restrictions should be imposed.

*Restricted and Common Dynamics [RD, CD]*

Finally, one may want to impose restrictions on the autoregressive structure in (4), implying that the purchase reinforcement effects are the same across brands. For example, the restriction that the attraction of brand $i$ at time $t$ only depends on its own lagged market shares $M_{i,t}$ corresponds with the restriction $\alpha_{p,j,i}=0$ for $j \neq i$ in (4). The corresponding multivariate model, representing an attraction model with Restricted Dynamics [RD], then becomes

$$\log M_{i,t} - \log M_{I,t} = \tilde{\mu}_i + \sum_{j=1}^{I} \sum_{k=1}^{K} \tilde{\beta}_{k,j,i} \log x_{k,j,t} + \sum_{j=1}^{I} \sum_{p=1}^{P} \left( \alpha_{p,i} \log M_{j,t-p} - \alpha_{p,I} \log M_{I,t-p} + \sum_{k=1}^{K} \tilde{\beta}_{p,k,j,i} \log x_{k,j,t-p} \right) + \eta_{i,t}, \quad (13)$$

for $i=1, \ldots, I-1$, where we again save on notation by using $\alpha_{p,i}$ instead of $\alpha_{p,i,i}$. Note that now the $\alpha_{p,I}$ parameters are the same across the $(I-1)$ equations and hence that these restrictions should be imposed when estimating the model parameters. To illustrate, Chen et al. (1994) additionally impose that $P=1$ and $\alpha_{1,i}=\gamma$, which yields the estimable version of the attraction model in (3) which assumes that the purchase reinforcement effects are the same across brands. For further reference, we will call this last restriction the Common Dynamics [CD] restriction.

To illustrate the common dynamics model we consider a simple attraction model with $P=1$ and restricted effects, that is,

$$A_{i,t} = \exp(\mu_i + \varepsilon_{i,t}) x_{i,t}^{\beta_0} x_{i,t-1}^{\beta_1} M_{i,t-1}^{\alpha} \qquad \text{for } i=1, \ldots, I, \quad (14)$$

with $(\varepsilon_{1,t}, \ldots, \varepsilon_{I,t})' \sim \text{NID}(\mathbf{0}, \Sigma)$. This attraction specification corresponds to the following set of $I-1$ linear equations

$$\log M_{i,t} - \log M_{I,t} = \tilde{\mu}_i + \alpha(\log M_{i,t-1} - \log M_{I,t-1}) + \beta_0(\log x_{i,t} - \log x_{I,t}) + \beta_1(\log x_{i,t-1} - \log x_{I,t-1}) + \eta_{i,t}. \tag{15}$$

This equation is basically an Autoregressive Distributed Lag model for the variable $(\log M_{i,t} - \log M_{I,t})$. To determine the dynamic effects of lagged $x_{i,t}$ and $x_{I,t}$ on the market shares we solve (15) for $(\log M_{i,t} - \log M_{I,t})$

$$\log M_{i,t} - \log M_{I,t} = \alpha^t(\log M_{i,0} - \log M_{I,0}) + \sum_{\tau=0}^{t-1} \alpha^\tau(\tilde{\mu}_i + \beta_0(\log x_{i,t-\tau} - \log x_{I,t-\tau}) + \beta_1(\log x_{i,t-\tau-1} - \log x_{I,t-\tau-1}) + \eta_{i,t-\tau}) \tag{16}$$

for $i = 1, \ldots, I-1$. It is easy to see that the relative log market shares are stationary if $|\alpha| < 1$ as under this restriction the influence of the market shares at time 0 vanishes for $t \to \infty$. Under stationarity, the effect of lagged explanatory variables on current log market shares decreases exponentially.

The above discussion shows that various attraction models, which are considered in the relevant literature and in practice for modeling and forecasting market shares, are nested within the general attraction model in (4). The fact that these models are nested automatically suggests that an empirical model selection strategy can be based on a general-to-simple strategy, see Franses and Paap (2001b).

## 3. INTERPRETATION

As the market shares get modeled through the attraction specification, and as this implies a reduced form of the model where parameters represent the impact of marketing efforts on the logarithm of relative market shares, the parameter estimates themselves are not easy to interpret. To facilitate an easier interpretation, one usually resorts to elasticities. In fact, it turns out that the reduced-form parameters are sufficient to identify these (cross-)elasticities.

For model (4), the instantaneous elasticity of the $k$-th marketing instrument of brand $j$ on the market share of brand $i$ is given by

$$\frac{\partial M_{i,t}}{\partial x_{k,j,t}} \frac{x_{k,j,t}}{M_{i,t}} = \beta_{k,i,j} - \sum_{r=1}^{I} M_{r,t}\beta_{k,r,j}, \tag{17}$$

see Cooper (1993). To show that these elasticities are identified, one can rewrite them such that they only depend on the reduced-form parameters, that is,

$$\frac{\partial M_{i,t}}{\partial x_{k,j,t}}\frac{x_{k,j,t}}{M_{i,t}} = (\beta_{k,j,i} - \beta_{k,j,I})(1 - M_{i,t}) - \sum_{r=1 \wedge r \neq i}^{I-1} M_{r,t}(\beta_{k,j,r} - \beta_{k,j,I}), \tag{18}$$

see (6). Under Restricted Competition, these elasticities simplify to

$$\frac{\partial M_{i,t}}{\partial x_{k,j,t}}\frac{x_{k,j,t}}{M_{i,t}} = (\delta_{i=j} - M_{j,t})\beta_{k,j}, \tag{19}$$

where $\delta_{i=j}$ is the Kronecker $\delta$ which has a value of 1 if $i$ equals $j$ and 0 otherwise. Under Restricted Effects, we simply have

$$\frac{\partial M_{i,t}}{\partial x_{k,j,t}}\frac{x_{k,j,t}}{M_{i,t}} = (\delta_{i=j} - M_{j,t})\beta_{k}. \tag{20}$$

It is easy to see that the elasticities converge to zero if a market share goes to 1. From a marketing perspective, this seems rather plausible. If a brand controls almost the total market, its marketing efforts will have little if any effect on its market share. Secondly, in case the market share is an increasing function of instrument $X$, then if $X$ goes to infinity the elasticity will go to 0. These two properties may seem straightforward, but among the best known market share models, the attraction model is the only model satisfying these properties, see also Cooper (1993). Whether the above two properties hold in a practical attraction model depends on the specific transformation of variables used, although the MCI and the MNL specification both lead to elasticities satisfying these properties.

## 4. PARAMETER ESTIMATION

In this section we discuss two methods for parameter estimation, and we show that they are equivalent. The first method is rather easy, whereas the second (which seems to be commonly applied) is more difficult.

### *4.1. Using a Base Brand*

To estimate the parameters in attraction models, we consider the $(I-1)$-dimensional set of linear equations which results from log-linearizing the attraction model given in (7). In general, these equations can be written in the following form

$$\begin{aligned}
y_{1,t} &= w'_{1,t}b_1 &+ z'_{1,t}a &+ \eta_{1,t}\\
y_{2,t} &= w'_{2,t}b_2 &+ z'_{2,t}a &+ \eta_{2,t}\\
\vdots &= \vdots &+ \vdots &+ \vdots\\
y_{I-1,t} &= w'_{I-1,t}b_{I-1} &+ z'_{I-1,t}a &+ \eta_{I-1,t},
\end{aligned} \tag{21}$$

where $y_{i,t} = \log M_{i,t} - \log M_{I,t}$, $\eta_t = (\eta_{1,t}, \ldots, \eta_{I-1,t})' \sim \text{NID}(\mathbf{0}, \tilde{\Sigma})$, and where $w_{i,t}$ are $k_i$-dimensional vectors of explanatory variables with regression coefficient vector $b_i$, which is different in each equation, and where $z_{i,t}$ are $n$-dimensional vectors of explanatory variables with regression coefficient vector $a$ which is the same across the equations, $i = 1, \ldots, I-1$. Each (restricted) version of the general attraction model discussed in Section 2.2 can be written in this format, see Franses and Paap (2001b).

To discuss parameter estimation, it is convenient to write (21) in matrix notation. We define $y_i = (y_{i,1}, \ldots, y_{i,T})'$, $W_i = (w_{i,1}, \ldots, w_{i,T})'$, $Z_i = (z_{i,1}, \ldots, z_{i,T})'$ and $\eta_i = (\eta_{i,1}, \ldots, \eta_{i,T})'$ for $i = 1, \ldots, I-1$. In matrix notation, (21) then becomes

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{I-1} \end{pmatrix} = \begin{pmatrix} W_1 & \mathbf{0} & \ldots & \mathbf{0} & Z_1 \\ \mathbf{0} & W_2 & \ldots & \mathbf{0} & Z_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & W_{I-1} & Z_{I-1} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_{I-1} \\ a \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_{I-1} \end{pmatrix} \tag{22}$$

or

$$y = X\gamma + \eta \tag{23}$$

with $\eta \sim \text{N}(\mathbf{0}, (\tilde{\Sigma} \otimes \mathbf{I}_T))$, where $\otimes$ denotes the familiar Kronecker product.

One method for parameter estimation of (23) is ordinary least squares [OLS]. Generally, however, this leads to consistent but inefficient estimates, where the inefficiency is due to the (possibly neglected) covariance structure of the disturbances. Only if the explanatory variables in each equation are the same, or in the unlikely case that $\tilde{\Sigma}$ is a diagonal matrix, and provided that there are no restrictions on the regression parameters ($w_{i,t} = \mathbf{0}$ for all $i$, $t$), OLS provides efficient estimates, see Judge et al. (1985, Chapter 12), among others. Therefore, one should better use generalized least squares [GLS] methods to estimate the model parameters. As the covariance matrix of the disturbances is usually unknown, one has to opt for a feasible GLS procedure, where we use the OLS estimator of the covariance matrix of the disturbances. This procedure is known as Zellner's (1962) seemingly unrelated regression [SUR] estimation method. Under the assumption of normality, an iterative SUR estimation method will lead to the maximum likelihood [ML] estimator of the model parameters, see Zellner (1962).

To estimate the parameters in attraction models, and to facilitate comparing various models, we favor ML estimation. The log of the likelihood function of (23) is given by

$$\ell(\gamma, \tilde{\Sigma}) = -\frac{T(I-1)}{2}\log(2\pi) + \frac{T}{2}\log|\tilde{\Sigma}^{-1}| - \frac{1}{2}(y - X\gamma)'(\tilde{\Sigma}^{-1}\otimes \mathbf{I}_T)(y - X\gamma). \tag{24}$$

The parameter values which maximize this log likelihood function are consistent and efficient estimates of the model parameters.

For the FE-MCI model without any parameter restrictions in (7), the ML estimator corresponds with the OLS estimator, as the explanatory variables are the same across equations. In that case,

$$\hat{\gamma}_{OLS} = (X'X)^{-1}X'y \tag{25}$$

such that $\hat{\gamma}_{OLS} = (\hat{b}_{OLS,1}, \ldots, \hat{b}_{OLS,I-1}, \hat{a}_{OLS})'$, see (22), and

$$\hat{\tilde{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T} \hat{\eta}_t\hat{\eta}_t', \tag{26}$$

where $\hat{\eta}_t$ consists of stacked $\hat{\eta}_{i,t} = y_{i,t} - w_{i,t}'\hat{b}_{OLS,i} - z_{i,t}'\hat{a}_{OLS}$.

For the attraction models with restrictions on the regression parameters, that is, for the RC model in (10), the RE model in (12), and the RD model in (13), one can opt for the iterative SUR estimator which converges to the ML estimator. Starting with the OLS-based estimator for $\tilde{\Sigma}$ in (26), one constructs the feasible GLS estimator

$$\hat{\gamma}_{SUR} = (X'(\hat{\tilde{\Sigma}}^{-1}\otimes \mathbf{I}_T)X)^{-1}X'(\hat{\tilde{\Sigma}}^{-1}\otimes \mathbf{I}_T)y, \tag{27}$$

that is the SUR estimator, see Zellner (1962). Next, we replace the estimate of the covariance matrix $\tilde{\Sigma}$ by the new estimate of $\tilde{\Sigma}$, that is (26), where $\hat{\eta}_t$ now consists of stacked $\hat{\eta}_{i,t} = y_{i,t} - w_{i,t}'\hat{b}_{SUR,i} - z_{i,t}'\hat{a}_{SUR}$, to obtain a new SUR estimate of $\gamma$. This routine is repeated until the estimates for $\gamma$ and $\tilde{\Sigma}$ have converged. Under the assumption of normally distributed disturbances, the final estimates are the ML estimates of the model, that is, they maximize the log likelihood function (24).

A little more involved are the restrictions on the $\tilde{\Sigma}$ matrix. To estimate the attraction model under the restriction (8), one can either directly maximize the log likelihood function (24) with $\tilde{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_{I-1}^2) + \sigma_I^2\mathbf{i}_{I-1}\mathbf{i}_{I-1}'$ using a numerical optimization algorithm like Newton-Raphson or one can again use an iterative SUR procedure. In the latter approach, the new estimate of $\tilde{\Sigma}$ is obtained by maximizing

$$\ell(\tilde{\Sigma}) = -\frac{T(I-1)}{2}\log(2\pi) + \frac{T}{2}\log|\tilde{\Sigma}^{-1}| - \frac{1}{2}\hat{\eta}'(\tilde{\Sigma}^{-1}\otimes \mathbf{I}_T)\hat{\eta}, \qquad (28)$$

where $\hat{\eta}$ are the residuals from the previous SUR regression. Again, we need a numerical optimization routine to maximize (28). Especially in cases where there are many brands, the optimization of (28) can become cumbersome. It can however be shown, see Appendix A, that the optimization can be reduced to numerically maximizing a concentrated likelihood over just $\sigma_I^2$ where one uses

$$\hat{\sigma}_i^2 = \frac{\hat{\eta}_i'\hat{\eta}_i}{T} - \hat{\sigma}_I^2 \qquad \text{for } i = 1, \ldots, I-1, \qquad (29)$$

where $\hat{\eta}_i = (\hat{\eta}_{i,1}, \ldots, \hat{\eta}_{i,T})'$. Given an estimate of $\sigma_I^2$, this relationship can be used to obtain estimates of $\sigma_1^2, \ldots, \sigma_{I-1}^2$.

Finally, in all the above cases the standard errors for the estimated regression parameters $\gamma$ are to be estimated by

$$\hat{V}(\hat{\gamma}) = (X'(\hat{\tilde{\Sigma}}^{-1} \otimes \mathbf{I}_T)X)^{-1}, \qquad (30)$$

where one should include the appropriate ML estimator for $\tilde{\Sigma}$. When taking the square roots of the diagonal elements of this matrix, one obtains the appropriate standard errors.

### *4.2. An Alternative Estimation Method*

The above estimation routine is based on the reduced-form model, which is obtained from reducing the system of equations using the base-brand approach. An alternative method is the, what is called, log-centering method advocated by Cooper and Nakanishi (1988). We will now show that this method is equivalent to the above method, although a bit more complicated.

The log-centering approach is based on the following transformation. After taking the natural logs for the $I$ model equations, the log of the geometric mean market share over the brands is subtracted from all equations. The reduced-form model is now specified relative to the geometric mean. So instead of reducing the system of equations by using a base brand, this methodology reduces the system by the "geometric average brand". Note that the reduced-form model in this case still contains $I$ equations.

To demonstrate the equivalence of parameters obtained through the log-centering technique of Cooper and Nakanishi (1988) and those using the base-brand approach, we show that there exists an exact relationship between these sets of parameters. The parameters for the base-brand specification can uniquely be determined from the parameters for the log-centering specification and vice versa. Given the 1-to-1 relationship the likelihoods are the same, that is, the discussed feasible GLS estimator yields the same maximum value of the likelihood as we can use the invariance principle of maximum likelihood, see for example Greene (1993, p. 115). All that needs to be shown is the 1-to-1 relationship between the parameters in the two specifications.

Consider a general attraction specification, that is

$$A_{i,t} = \exp(\mu_i + \varepsilon_{i,t}) \prod_{j=1}^{I} \prod_{k=1}^{K} z_{k,j,t}^{\beta_{k,j,i}}, \tag{31}$$

where $z_{k,j,t}$ may contain any kind of explanatory variable, such as lagged market shares, promotion and price. The market shares are again defined by

$$M_{i,t} = \frac{A_{i,t}}{\sum_{j=1}^{I} A_{j,t}}. \tag{32}$$

Written in a vector notation the model for the natural logarithm of attraction becomes

$$\log A_t := \begin{pmatrix} \log A_{1,t} \\ \vdots \\ \log A_{I,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_I \end{pmatrix} + \sum_{k=1}^{K} \begin{pmatrix} \beta_{k,1,1} & \beta_{k,2,1} & \cdots & \beta_{k,I,1} \\ \beta_{k,1,2} & \beta_{k,2,2} & \cdots & \beta_{k,I,2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k,1,I} & \beta_{k,2,I} & \cdots & \beta_{k,I,I} \end{pmatrix} \begin{pmatrix} \log z_{k,1,t} \\ \vdots \\ \log z_{k,I,t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{I,t} \end{pmatrix}$$

$$= \mu + \sum_{k=1}^{K} B_k \log z_{k,t} + \varepsilon_t. \tag{33}$$

The definition of market share in (32) implies that $\log M_{i,t} = \log A_{i,t} - \log \sum_{j=1}^{I} A_{j,t}$. In a vector notation this gives

$$\log M_t := \begin{pmatrix} \log M_{1,t} \\ \vdots \\ \log M_{I,t} \end{pmatrix} = \log A_t - \mathbf{i}_I \log \sum_{j=1}^{I} A_{j,t}, \tag{34}$$

where $\mathbf{i}_I$ denotes a $(I \times 1)$ unity vector.

As the model in (34) cannot be estimated directly due to the nonlinear dependence of $\log(\sum_{j=1}^{I} A_{j,t})$ on the model parameters, a reduced-form model should be considered. The log-centering method now subtracts the average of the log market shares from the equations to give a reduced-form specification. The dependent variable in this system of equations is now

$$\begin{pmatrix} \log M_{1,t} \\ \vdots \\ \log M_{I,t} \end{pmatrix} - \begin{pmatrix} \frac{1}{I}\sum_{j=1}^{I} \log M_{j,t} \\ \vdots \\ \frac{1}{I}\sum_{j=1}^{I} \log M_{j,t} \end{pmatrix} = \begin{pmatrix} 1-\frac{1}{I} & -\frac{1}{I} & \dots & -\frac{1}{I} \\ -\frac{1}{I} & 1-\frac{1}{I} & \dots & -\frac{1}{I} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{I} & -\frac{1}{I} & \dots & 1-\frac{1}{I} \end{pmatrix} \log M_t$$

$$= H_{lc} \log M_t, \tag{35}$$

where $H_{lc}$, with rank $I-1$, denotes the transformation matrix corresponding to the log-centering approach. The reduced-form model then becomes

$$H_{lc} \log M_t = H_{lc} \log A_t - H_{lc}\mathbf{i}_I \log \sum_{j=1}^{I} A_{j,t}, \tag{36}$$

which equals

$$H_{lc} \log M_t = H_{lc}\mu + \sum_{k=1}^{K} H_{lc}B_k \log z_{kt} + H_{lc}\varepsilon_t \tag{37}$$

as $H_{lc}\mathbf{i}_I = 0_{I\times I}$. Due to the reduced rank of $H_{lc}$, the system in (37) contains $I$ equations, but it only has $I-1$ independent equations.

Alternatively, the base-brand approach in Section 4.1 gives as the dependent variables in the reduced-form model

$$\begin{pmatrix} \log M_{1,t} \\ \vdots \\ \log M_{I-1,t} \end{pmatrix} - \begin{pmatrix} \log M_{I,t} \\ \vdots \\ \log M_{I,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix} \log M_t$$

$$= H_{bb} \log M_t,$$

with $H_{bb}$ as the relevant transformation matrix. As $H_{bb}\mathbf{i}_I = \mathbf{0}_{I-1\times I}$, the reduced-form model becomes

$$H_{bb} \log M_t = H_{bb} \log A_t = H_{bb}\mu + \sum_{k=1}^{K} H_{bb}B_k \log z_{kt} + H_{bb}\varepsilon_t, \tag{39}$$

which is to be compared with (37). This system contains only $I-1$ equations.

The 1-to-1 relation between the parameters in the two approaches follows from the fact that the equation $CH_{lc} = H_{bb}$ yields a unique solution $C$, given by

$$C = \begin{pmatrix} 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -1 \\ 1 & 1 & \dots & 1 & 1 \end{pmatrix}. \tag{40}$$

Hence, the matrix $C$ relates the "log-centered" parameters to the "base-brand" parameters. The inverse transformation from the base-brand specification to the log-centered specification follows from applying the Moore-Penrose inverse of $C$, denoted by $C^+$, that is,

$$C^+ = \begin{pmatrix} 1-\frac{1}{I} & -\frac{1}{I} & \dots & -\frac{1}{I} \\ -\frac{1}{I} & 1-\frac{1}{I} & \dots & -\frac{1}{I} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{I} & -\frac{1}{I} & \dots & 1-\frac{1}{I} \\ -\frac{1}{I} & -\frac{1}{I} & \dots & -\frac{1}{I} \end{pmatrix}. \tag{41}$$

Note that the matrix $C^+$ satisfies $H_{lc} = C^+ H_{bb}$.

The above shows that the transformations yield equivalent parameters. For example, assume that the log-centered form of the model is estimated, giving estimates of $H_{lc}\mu$, $H_{lc}B_k$ and $H_{lc}\Sigma H'_{lc}$. By multiplying the estimated system of equations by $C$ we get $CH_{lc}\mu$, $CH_{lc}B_k$ and $CH_{lc}\Sigma H'_{lc}C'$ as model coefficients. Using the invariance principle of maximum likelihood and the relation $CH_{lc} = H_{bb}$, these coefficients are the maximum likelihood estimates of $H_{bb}\mu$, $H_{bb}B_k$ and $H_{bb}\Sigma H'_{bb}$. These coefficients are exactly the same as the coefficients used in the base-brand specification, see (39). Using the inverse of $C$, the procedure can be used the other way around. We can also obtain estimates of the coefficients in a log-centered specification from the estimates in a base-brand specification by multiplying them with $C^+$.

In our opinion, the main reason to prefer taking a base brand to reduce the model is that the statistical analysis of the resulting model is more straightforward as compared to the log-centering technique. Recall that the log-centered reduced-form model contains $I$ equations whereas the base brand reduced-form model only has $I-1$ equations. One of the equations in the log-centered specification is however redundant. This redundancy leads to some difficulties in the estimation and interpretation, as estimation usually requires the (inverse) covariance matrix of the residuals. In the log-centering case the residuals are linearly dependent, and the covariance matrix is therefore non-invertible. Further, direct interpretation of the coefficients obtained from the base-brand approach is easier as each coefficient only concerns two brands, while a coefficient in the log-centering approach always involves all brands.

Another advantage of using the base-brand approach concerns markets where the number of brands changes over time. In this case the "geometric average" brand may consist of a different number of brands across weeks. The variability in the market share of this average brand will fluctuate with the numbers of brands available. Using this average brand as a base brand, as proposed in the log-centering approach, will therefore introduce complicated forms of heteroscedasticity. If a brand is available during the entire sample period, the base-brand approach can be straightforwardly applied without introducing heteroscedasticity. If such a brand is not available, a different base brand can be considered for different weeks. This will also introduce some heteroscedasticity, but of a more manageable form than would be the case for the log-centering approach.

# 5. DIAGNOSTICS

In this section we present some basic diagnostics for the market share attraction model. First of all we present a test on the normality assumption in the attraction specification. Next, we discuss tests for outliers and tests for structural breaks.

## *5.1. Normality*

An important assumption made in the development of the attraction model is the normality of the unexplained attractions. Much of the inference is based on this assumption. For example, significance tests of parameters are based on the normality assumption. Therefore, as in every model, it is important to test the distributional assumption.

One can test the normality of each of $\hat{\eta}_1, \ldots, \hat{\eta}_{I-1}$ separately using the familiar normality test by Bowman and Shenton (1975) which is based on the skewness, denoted by $\sqrt{b_1}$, and the kurtosis, denoted by $b_2$, of the residuals for every brand. However, Doornik and Hansen (1994) argue that this test is unsuitable except in very large samples. Instead, they propose to use the sum of squared transformed skewness and kurtosis measures, where the transformation involved is as in D'Agostino (1970). The resultant test statistic equals

$$E_p = z_1(b_1)^2 + z_2(b_1, b_2)^2, \tag{42}$$

where $z_1(\cdot)$ and $z_2(\cdot,\cdot)$ are the relevant transformation functions. Under the hypothesis of normally distributed $\eta_i$, $i=1,\ldots,I-1$, the test statistic is asymptotically $\chi^2(2)$ distributed. Note that the normality of $\eta_i$ depends on the normality of both $\varepsilon_i$ and $\varepsilon_I$. It is however not possible to test the normality of the individual $\varepsilon_i$. Therefore, it is easier to use a joint test on the normality of all disturbances. Doornik and Hansen (1994) show that a joint test statistic for multivariate normality can easily be obtained by summing the individual test statistics. The resulting statistic has a $\chi^2(2(I-1))$ distribution under the null hypothesis of joint normality.

### 5.2. *Outliers*

As Franses et al. (1999) suggest, scanner data may contain several aberrant observations. Therefore, it is important to check for such observations as they may have a large influence on the parameter estimates.

Testing for outliers in market shares is not straightforward. A sudden event in the market share of one brand is by definition accompanied by an opposite effect in the remainder of the market. Outliers in market shares can therefore not be attributed to a single brand. It is then easier to test for an outlier in attractions. To test for this in the attraction of brand $j$ at time $T_b$, we simply include $\exp(D_t)$ in the attraction specification of brand $j$. The dummy variable $D_t$ is defined as

$$D_t = \begin{cases} 1 & \text{if } t=T_b \\ 0 & \text{elsewhere.} \end{cases} \tag{43}$$

Note that due to the multiplicative specification of attraction we need the exponential transformation to ensure that the new variable does not affect the attraction if $t \neq T_b$. For the specification of the reduced-form model it matters whether the brand with the aberrant observation is the base brand or not. In case $j<I$, so that brand $j$ is not the base brand, we just add the variable $D_t$ to the reduced-form equation for $\log M_{j,t} - \log M_{I,t}$. In case the brand with the aberrant observation happens to be the base brand the variable $-D_t$ is added to the equations for $\log M_{i,t} - \log M_{I,t}$, $i=1,\ldots,I-1$, where the corresponding coefficients are restricted to be equal across the equations.

Whether the observation at $T_b$ actually corresponds with an outlier in the attraction of brand $j$ can now easily be tested by testing the significance of $D_t$ in the reduced-form model. In case the observation does turn out to be an outlier, one can opt to remove the observations at $T_b$ from the data set to prevent the outlier from influencing the estimation results. One can also choose to include the above introduced variable into the model and base the interpretation of the model on the resulting specification. In fact, the inclusion of $D_t$ "removes" the influence of the market share at time $T_b$ of brand $j$.

### 5.3. Structural Breaks

Brand introductions or brand/line extensions can change the entire market structure. Less radical changes like brand repositioning can also change (part of) the market structure. These changes can cause only average market share to change, which corresponds to changes in brand intercepts in the attraction model, but it can also change the competitive structure on specific instruments.

Testing for a structural break is much like testing for outliers. To test for a structural break in the attraction of brand $j$ starting from time $T_b$, one can just add the variable $\exp(D_t^*)$ to the attraction specification of brand $j$, with

$$D_t^* = \begin{cases} 1 & \text{if } t \geq T_b \\ 0 & \text{elsewhere.} \end{cases} \tag{44}$$

Using the same reasoning as above, the reduced-form specifications can be obtained. The significance of $D_t^*$ in the reduced-form model indicates whether there has been a break at time $T_b$.

The above methodology only considers a break in the level of the attraction. The structural break can also be in the effect of one of the marketing instruments. For example, due to a repositioning of brand $j$, the price elasticity of this brand may change. To test for this, one can add the variable $\exp[D_t^* \log(P_{j,t})]$ to the attraction specification of brand $j$, and correspondingly to the reduced-form equations.

## 6. MODEL SELECTION

Attraction models are often considered for forecasting market shares. It is usually assumed that, by imposing in-sample specification restrictions, the out-of-sample forecasting accuracy will improve. Exemplary studies are Brodie and Bonfrer (1994), Danaher (1994), Naert and Weverbergh (1981), Leeflang and Reuyl (1984), Kumar (1994) and Chen et al. (1994), among others. A summary of the relevant studies is given in Brodie et al. (2001). A common characteristic of these studies, an exception being Chen et al. (1994), is that they tend to compare one or two specific forms of the attraction model with various more naive models. In this section we consider the question of obtaining the best (or a good) choice for the specification from the wide range of possible attraction specifications.

There are of course many possible approaches to obtain a suitable attraction specification. One could consider a set of popular specifications and select the optimal model using an information criterion, like the BIC (Schwarz, 1978), or use statistical tests to determine the "best" model. In a Bayesian setting one could even derive posterior probabilities for the proposed models. One may select the model with the highest posterior probability or one can combine several models. For example, to construct forecasts, one can use the posterior probabilities to weight forecasts generated by the different models. Another strategy is to start with a general model and try to simplify it using statistical tests. In this chapter we opt for this general-to-simple model selection strategy, following Hendry (1995). In Franses and Paap (2001b) it is shown that this strategy tends to work well in empirical applications.

The starting point of the model selection strategy is the most extended attraction model, that is, model (7) without any restrictions. Of course, in practice the size of the model is governed by data availability and sample size. The first step of a model selection strategy concerns fixing the proper lag order $P$ of the model. It is well known that an inappropriate value of $P$ leads to inconsistent and inefficient estimates. To perform valid inference on the restrictions on the explanatory variables and covariance matrix it is therefore necessary to first determine the appropriate lag order. Furthermore, imposing incorrect restrictions on the explanatory variables and covariance matrix may lead to selecting an incorrect lag order. Lag order selection may be based on the BIC criterion. Another strategy may be a sequential procedure, where one starts with a large value of $P$ and tests for the significance of the $\tilde{\beta}_{P,k,j,i}$ and $\tilde{\alpha}_{P,j,i}$ parameters and imposes these restrictions when they turn out to be valid. These tests usually concern many parameter restrictions and may therefore have little power. Instead, one may therefore base the lag order determination on Lagrange Multiplier [LM] tests for serial correlation in the residuals, see Lütkepohl (1993) and Johansen (1995, p. 22). The advantage of these tests is that they concern less parameter restrictions and hence have more power. We would recommend to start with a model of order 1 and increase the order with 1 until the LM tests do not indicate the presence of any serial correlation.

Once $P$ is fixed, we propose to test the validity of the various restrictions on (7) as proposed in Section 2.2. We test for the validity of restriction (8) on the covariance matrix $\tilde{\Sigma}$ [RCM] in model (7). Additionally, we test in model (7) for restricted dynamics [RD], common dynamics [CD], and, for each explanatory variable $k$, for restricted competition [RC], for restricted effects [RE] (12) and even for the absence of this variable. Finally, we propose to test for the significance of the lagged explanatory variables in the general model.

Next, we recommend to perform an overall test for all restrictions which were not rejected in the individual tests. If this joint test is not rejected, all restrictions are imposed, and this results in a final model that can be used for forecasting. However, if the joint test indicates rejection, one may want to decide to relax some restrictions, where the $p$-values of the individual tests can be used to decide which of these restrictions have to be relaxed. Note that apart from the lag order selection stage we perform the individual tests in the general model and that we do not directly impose the restrictions if not rejected. Hence, the model selection approach in this stage does not depend on the sequence of the tests. Furthermore, as we use a general-to-specific strategy, we do not a priori exclude model specifications.

To apply our general-to-simple model selection strategy, we have to test for restrictions on the covariance matrix $\tilde{\Sigma}$ and on the other model parameters (collected in $\gamma$) in (7). To test these parameter restrictions, we opt for Likelihood Ratio [LR] tests, see for example Judge et al. (1985, p. 475). Denoting the ML estimates of the parameters under the null hypothesis by $(\hat{\gamma}_0, \hat{\tilde{\Sigma}}_0)$ and the ML estimates under the alternative hypothesis by $(\hat{\gamma}_a, \hat{\tilde{\Sigma}}_a)$, then

$$\mathrm{LR} = -2(\ell(\hat{\gamma}_0, \hat{\tilde{\Sigma}}_0) - \ell(\hat{\gamma}_a, \hat{\tilde{\Sigma}}_a)) \underset{asy}{\sim} \chi^2(\nu), \tag{45}$$

where $\ell(\cdot)$ denotes the log-likelihood function as defined in Section 4 and where $\nu$ is the number of parameter restrictions.

## 7. FORECASTING

There has been considerable research on forecasting market shares using the market share attraction model. Most studies discuss the effect of the estimation technique used in combination with the parametric model specification on the forecasts, see for example Leeflang and Reuyl (1984), Brodie and de Kluyver (1984) and Ghosh et al. (1984), among others. More recent interest has been on the optimal model specification under different conditions, see, for example, Kumar (1994) and Brodie and Bonfrer (1994). The available literature, however, is not specific as to how forecasts of market shares should be generated. In this section we show that forecasting market shares turns out not to be a trivial exercise and that in order to obtain unbiased forecasts one has to use simulation methods.

Furthermore, in empirical applications it should be recognized that parameter values are obtained through estimation. The true parameter values are usually unknown, and parameter values are at best obtained through unbiased estimators of the true values. In a linear model this parameter uncertainty can be ignored when constructing unbiased forecasts. However, in nonlinear models this may not be true, see for example Hsu and Wilcox (2000).

### 7.1. Forecasting Market Shares

To provide some intuition why forecasting in a market share attraction model is not a trivial exercise, consider the following. The attraction model ensures logical consistency, that is, market shares lie between 0 and 1 and they sum to 1. These restrictions imply that the model parameters can be estimated from a multivariate reduced-form model with $I-1$ equations. The dependent variable in each of the $I-1$ equations is the natural logarithm of a relative market share. More formally, it is $\log m_{i,t} \equiv \log \frac{M_{i,t}}{M_{I,t}}$, for $i=1, 2, \ldots, I-1$. The base brand $I$ can be chosen arbitrarily.

Of course, one is usually interested in predicting $M_{i,t}$ and not in the logs of the relative market shares. It is then important to recognize that, first of all, $\exp(\text{E}[\log m_{i,t}])$ is not equal to $\text{E}[m_{i,t}]$ and that, secondly, $\text{E}[M_{i,t}/M_{I,t}]$ is not equal to $\text{E}[M_{i,t}]/\text{E}[M_{I,t}]$, where E denotes the expectation operator. Therefore, unbiased market share forecasts cannot be obtained by routinized data transformations, see also Fok and Franses (2001b) for similar statements.

To forecast the market share of brand $i$ at time $t$, one needs to consider the relative market shares

$$m_{j,t} = M_{j,t}/M_{I,t} \qquad \text{for } j=1, 2, \ldots, I,$$

as $m_{1,t}, \ldots, m_{I-1,t}$ form the dependent variables (after log transformation) in the reduced-form model (7). As $M_{I,t} = 1 - \sum_{j=1}^{I-1} M_{j,t}$, we have that

$$M_{I,t} = \frac{1}{1 + \sum_{j=1}^{I-1} m_{j,t}}$$

$$M_{i,t} = M_{I,t} m_{i,t} = \frac{m_{i,t}}{1 + \sum_{j=1}^{I-1} m_{j,t}} \qquad \text{for } i=1, 2, \ldots, I-1. \tag{47}$$

Note that $m_{I,t} = M_{I,t}/M_{I,t} = 1$ and hence (47) can be summarized as

$$M_{i,t} = \frac{m_{i,t}}{\sum_{j=1}^{I} m_{j,t}} \qquad \text{for } i=1, 2, \ldots, I. \tag{48}$$

As the relative market shares $m_{i,t}$, $i = 1, \ldots, I-1$ are log-normally distributed by assumption, see (7), the probability distribution of the market shares involves the inverse of the sum of log-normally distributed variables. The exact distribution function of the market shares is therefore complicated. Moreover, correct forecasts should be based on the expected value of the market shares, and unfortunately, for this expectation there is no simple algebraic expression. Appropriate forecasts therefore cannot be obtained from the expectations directly.

If we ignore parameter uncertainty for the moment, we need to calculate the expectations of the market shares given in (48). This cannot be done analytically. However, we can calculate the expectations using simulations. The relevant procedure works as follows. We use model (7) to simulate relative market shares for various disturbances $\eta$ randomly drawn from a multivariate normal distribution with mean 0 and covariance matrix $\tilde{\Sigma}$. In each run, we compute the market shares where parameter values and the realization of the disturbance process are assumed to be given. The market shares averaged over a number of replications now provide their unbiased forecasts. Notice that we only need the parameters of the reduced-form model in the simulations.

To be more precise about this simulation method, consider the following. The one-step ahead forecasts of the market shares are simulated as follows, first draw $\eta_t^{(l)}$ from $N(0, \tilde{\Sigma})$, then compute

$$m_{i,t}^{(l)} = \exp(\tilde{\mu}_i + \eta_{i,t}^{(l)}) \prod_{j=1}^{I} \left( \prod_{k=1}^{K} x_{k,j,t}^{\tilde{\beta}_{k,j,i}} \prod_{p=1}^{P} \left( M_{j,t-p}^{\tilde{\alpha}_{p,j,i}} \prod_{k=1}^{K} x_{k,j,t-p}^{\tilde{\beta}_{p,k,j,i}} \right) \right), \quad i = 1, \ldots, I-1, \tag{49}$$

with $m_{I,t}^{(l)} = 1$ and finally compute

$$M_{i,t}^{(l)} = \frac{m_{i,t}^{(l)}}{\sum_{j=1}^{I} m_{j,t}^{(l)}} \qquad \text{for } i = 1, \ldots, I, \tag{50}$$

where $l = 1, \ldots, L$ denotes the simulation iteration and where the FE-MCI specification is used, see (4). Every vector $(M_{1,t}^{(l)}, \ldots, M_{I,t}^{(l)})'$ generated this way amounts to a draw from the joint distribution of the market shares at time $t$. Using the average over a sufficiently large number of draws we calculate the expected value of the market shares. By the weak law of large numbers we have

$$\operatorname*{plim}_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} M_{i,t}^{(l)} = \mathrm{E}[M_{i,t}]. \tag{51}$$

For finite $L$ the mean value of the generated market shares is an unbiased estimator of the market share. The estimate may differ from the expected market share, but this difference is only due to simulation error and this error will rapidly converge to zero if $L$ gets large. Of course, the value of $L$ can be set at a very large value, depending on available computing power.

The lagged market shares in (7) are of course only available for one-step ahead forecasting and not for multiple-step ahead forecasting. Hence, one has to account for the uncertainty in the lagged market share forecasts. One can now simply use simulated values for lagged market shares, thereby automatically taking into account the uncertainty in these lagged variables. Note that we do assume that the marketing efforts of all market players are known. It is possible to also model these efforts and use the estimated model to obtain forecasts that also account for that uncertainty. The models describing the marketing efforts can be used to simulate future values of the levels of the marketing instruments. To take into account the uncertainty of future marketing efforts for forecasting market shares, we use the simulated efforts instead of forecasted efforts to obtain draws from the joint distribution of market shares in (49) and (50).

### 7.2. *Parameter Uncertainty*

The model parameters, including those in $\tilde{\Sigma}$, usually have to be estimated from data. This implies that the parameter estimators are random variables. If estimated parameters are used for forecasting in combination with a nonlinear model, we should also take into account the uncertainty of these estimates. To take account of the stochastic nature of the estimator, we explicitly take the expectation of the market shares over the unknown parameters.

Unfortunately, the relevant distribution of the parameters is not known. To overcome this difficulty, we propose to use parametric bootstrapping to draw parameters from their distribution. Summarizing all parameters in $\theta$, we sample $\theta$ using the following scheme:

- Use the estimated parameters $\hat{\theta}$, the realizations of the exogenous variables and the first $P$ observed realizations as starting values to generate artificial realizations of the market shares.

- Re-estimate the model based on this artificial data.

The thus obtained parameters $\hat{\theta}^{(l)}$, $l = 1, \ldots, L$, where $L$ denotes the number of draws, can be seen as draws from the small sample distribution of $\hat{\theta}$. Based on these draws we calculate $\text{E}(M_{i,t})$ as $\frac{1}{L}\sum_{l=1}^{L} \text{E}[M_{i,t} \mid \hat{\theta}^{(l)}]$.

In the market share attraction model the forecasting scheme becomes more complicated as the market shares do not depend linearly on the disturbances. From (7), (46) and (48) we have $M_{i,t} = g_i(X_t, \ldots, X_{t-P}, M_{t-1}, \ldots, M_{t-P}, \eta_t, \theta)$, where $X_t$ contains all exogenous variables at time $t$, $M_t = (M_{1,t}, \ldots, M_{I,t})'$, and $g_i(\cdot)$ is a nonlinear function. As in this case $M_{i,t}$ also nonlinearly depends on the model disturbances $\text{E}[M_{i,t} \mid \hat{\theta}] \neq g_i(X_t, \ldots, X_{t-P}, M_{t-1}, \ldots, M_{t-P}, \mathbf{0}, \hat{\theta})$. To obtain unbiased forecasts, we therefore have to take the expectation of $g_i(\cdot)$ with respect to $\eta_t$ **and** $\theta$, that is

$$\text{E}(M_{i,t}) = \int_{\theta} \int_{\eta_t} g_i(X_t, \ldots, X_{t-P}, M_{t-1}, \ldots, M_{t-P}, \eta_t, \theta)\phi(\eta_t \mid \theta) f(\theta) d\eta_t d\theta, \quad (52)$$

where $\phi(\eta_t \mid \theta)$ denotes the distribution function of the (normally) distributed disturbances given the parameters and $f(\theta)$ denotes the distribution of the parameters. Again we choose to calculate the complex integral using simulation. The parameter vectors are simulated using the bootstrap methodology described above. For every bootstrap realization of $\hat{\theta}^{(l)}$ we calculate $\text{E}[M_{i,t} \mid \hat{\theta}^{(l)}]$, $i = 1, \ldots, I$ using the simulation technique in Section 7.1. The average of the forecasts over all generated parameter vectors constitutes unbiased forecasts of the market shares under uncertain parameters. It is not necessary to use many simulation rounds conditional on the parameters. Theoretically it suffices to use one round for every $\hat{\theta}^{(l)}$.

In a classical setting we have to rely on bootstrapping techniques to account for parameter uncertainty. A Bayesian analysis of market share models has the advantage that it provides a more natural approach to account for parameter uncertainty. To obtain the posterior distribution of the parameters of the market share attraction model, one can rely on Markov chain Monte Carlo [MCMC] methods, see Casella and George (1992) for a simple introduction and Paap (2002) for a recent survey. As byproduct of this sampler we can obtain forecasts which account for parameter uncertainty.

## 8. AN ILLUSTRATION

To illustrate some of the methods put forward in this chapter, we consider a data set containing market shares, prices and two 0/1 dummy variables (feature and display). The market at hand concerns four brands of peanut butter. Three of these brands are large national brands, the fourth one aggregates smaller brands and private labels. The data set is part of the so-called ERIM data base of the University of Chicago. The data we consider are collected from July 1986 until December 1988 by ACNielsen in Sioux Falls, South Dakota.

We first apply our model selection strategy to this market to obtain a suitable model specification. Next, we use some statistical tests to assess the validity of the model. For model selection we use 111 of the 124 available observations. The remaining 13 observations, corresponding to a quarter of a year, are used as out-of-sample data to demonstrate our forecasting strategy. Along the way we show that this strategy performs better than constructing market share forecasts from forecasts of log relative market shares. We also compare the forecasting performance of the selected model to various attraction specifications proposed in the literature.

For model selection, we first select the appropriate lag order to capture the dynamics in this market. An LM test for serial correlation in the residuals of an attraction model, where all variables are also included with one lag, indicates that it is not necessary to increase the lag order ($p$-value 0.2291). For this market, we can therefore fix $P = 1$. Next we test whether we can impose restrictions on the parameters in (7). First, we test the validity of the restricted covariance matrix restriction. An LR test on this restriction indicates that it cannot be rejected ($p$-value 0.0905). Additionally, we use the same type of test to assess the validity of restricted forms of competition on every marketing instrument separately, again in (7). For price and display, the $p$-values are not distinguishable from 0, for feature this value is 0.2393. Further testing shows that we cannot restrict the competition on feature as restricted elasticities ($p$-value 0.0016). The LR test on restricted dynamics indicates that this restriction can be imposed ($p$-value 0.0844), however common dynamics is rejected ($p$-value 0.0148).

Summarizing, in this market we can impose restricted dynamics and the restricted competition assumption on the use of display. Tests concerning the inclusion of lagged prices, feature and display show that the coefficients for these variables cannot be restricted to zero. Finally, we end up with a model with lag order one, a restricted covariance matrix, and restricted competition on feature and restricted dynamics. A joint test of all restrictions does not get rejected. Therefore we continue with this attraction specification.

In the resulting model we test for normality of the residuals. The test for multivariate normality does not indicate significant deviations from the normal distribution. Finally, we test for a structural break in the mean attractions. As for our data set, we do not have information on a relevant point in time of a possible break, we test for a break halfway in our sample, that is a break at week 75. Such a break could correspond to a change in the market structure. The $p$-value of the corresponding LR test is 0.5164 indicating that there seems to be no break. The market appears to have been quite stable during the observational period.

We use the selected attraction model to compare forecasts, made using our simulation technique, to forecasts obtained with a naive method, that constructs market share forecasts from forecasts of log relative market shares. To illustrate the performance of our model selection strategy, we also use attraction specifications often encountered in the literature to generate forecasts. In the relevant literature, we have found 5 types of attraction specifications. Table 1 presents the specification of these models together with some references. Note that none of the models uses lagged exogenous variables whereas our model selection procedure did indicate these variables to be important. Furthermore in all models the same competitive structure is assumed for every marketing instrument. We consider 13 one-step ahead forecasts, where the current and future levels of the marketing instruments are assumed to be known. Two sets of forecasts are made using simulations, one set while ignoring parameter uncertainty and one where we account for such uncertainty. These forecasts are

***Table 1.*** Model Specifications Used in the Literature.

| Model | Lag | Restrictions on* | | | | Literature |
|---|---|---|---|---|---|---|
| | | Dyn. | Cov. | Exo. | Lag. Exo. | |
| I | 1 | RD | NR | RC | NI | Leeflang and Reuyl (1984)<br>Danaher (1994) |
| II | 1 | CD | NR | RC | NI | Naert and Weverbergh (1981)<br>Brodie and Bonfrer (1994)<br>Brodie and de Kluyver (1984)<br>Chen et al. (1994)<br>Kumar (1994) |
| III | 0 | – | NR | RC | NI | Chen et al. (1994)<br>Ghosh et al. (1984) |
| IV | 1 | CD | NR | RE | NI | Naert and Weverbergh (1981)<br>Brodie and de Kluyver (1984)<br>Leeflang and Reuyl (1984)<br>Chen et al. (1994)<br>Kumar (1994) |
| V | 0 | – | NR | RE | NI | Chen et al. (1994) |

* RD = restricted dynamics, CD = common dynamics, RC = restricted competition, RE = restricted effects, NR = no restrictions, NI = not included

*Table 2.* Forecasting accuracy on 13 one-step ahead forecasts, measured by the log of the determinant of the residual covariance matrix.

| Model* | Naive | Simulated | Bootstrap |
|---|---|---|---|
| I | 9.57 | 9.58 | 9.63 |
| II | 9.99 | 9.95 | 9.97 |
| III | 10.51 | 10.43 | 10.42 |
| IV | 9.82 | 9.80 | 9.80 |
| V | 10.32 | 10.24 | 10.24 |
| VI | 9.13 | 9.05 | 9.05 |

* Models I to V are specified in Table 1, model VI is the model according to our model specification strategy.

referred to as "Simulated" and "Bootstrap", respectively. Both sets of forecasts are based on 25,000 replications. To measure the predictive accuracy of the models one could use the Root Mean Squared Prediction Error [RMSPE] per brand summed over all brands. However, the sum of the forecast errors over all brands is zero as market shares sum up to 1. Simply adding up the RMSPE over the brands is therefore not a good criterion of forecasting accuracy. As an alternative measure we consider the log of the determinant of the covariance matrix of the forecast errors for the first $I-1$ brands, see also Clements and Hendry (1993). This measure is independent of the chosen base brand, due to properties of the determinant operator.

In Table 2 we present the forecasting performance of the models in Table 1 together with the model suggested by our selection strategy. For all but one model, the forecasts obtained through simulation are more accurate than those from the naive method. Comparing the "Simulated" to the "Bootstrap" one sees that, for this market, correcting for parameter uncertainty does not seem to add much to the forecasting accuracy. Finally note that our model selection procedure seems to perform quite well as it yields the best forecasts for this market.

## 9. CONCLUDING REMARKS

In this chapter we have gone through part of the econometrics involved in analyzing market share attraction models. We believe that a systematic strategy enhances the possibility to compare various empirical findings and to understand deficiencies in case model forecasts turn out to be inaccurate.

There are a few more issues that need concern in future work. One of these involves the analysis of possibly differing short-run and long-run effects of marketing efforts, see Dekimpe and Hanssens (1995) and Paap and Franses (2000), among others. In Fok et al. (2001) we provide a first attempt in the context of a market share attraction model. Next, one may want to allow for the event of new brands entering the market or old brands leaving it. In Fok and Franses (2001a) we discuss techniques for doing so. Finally, one would want to allow for endogenous marketing efforts, like pricing strategies, which originate from attraction models.

## ACKNOWLEDGMENTS

We thank Alan Montgomery and an anonymous referee for their comments. The third author thanks the Netherlands Organization for Scientific Research (N.W.O.) for its financial support.

## REFERENCES

Bell, D. E., Keeney, R. L., & Little, J. D. C. (1975). A Market Share Theorem. *Journal of Marketing Research, 12*, 136–141.

Bowman, K. O., & Shenton, L. R. (1975), Omnibus Test Contours for Departures from Normality Based on $b_1^{1/2}$ and $b_2$. *Biometrika, 62*, 243–250.

Brodie, R. J., & Bonfrer, A. (1994). Conditions When Market Share Models are Useful for Forecasting: Further Empirical Results. *International Journal of Forecasting, 10*, 277–285.

Brodie, R. J., Danaher, P., Kumar, V., & Leeflang, P. S. H. (2001), Econometric Models for Forecasting Market Share. In: J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 597–611). Norwell, MA: Kluwer.

Brodie, R. J., & de Kluyver, C. A. (1984). Attraction versus Linear and Multiplicative Market Share Models: An Empirical Evaluation. *Journal of Marketing Research, 21*, 194–201.

Bronnenberg, B. J., Mahajan, V., & Vanhonacker, W. R. (2000). The Emergence of New Repeat-Purchase Categories: The Interplay of Market Share and Retailer Distribution. *Journal of Marketing Research, 37*, 16–31.

Casella, G., & George, E. I. (1992). Explaining the Gibbs Sampler. *American Statistician, 46*, 167–174.

Chen, Y., Kanetkar, V., & Weiss, D. L. (1994). Forecasting Market Shares with Disaggregate of Pooled Data: A Comparison of Attraction Models. *International Journal of Forecasting, 10*, 263–276.

Clements, M., & Hendry, D. (1993). On the Limitations of Comparing Mean Squared Forecast Errors. *Journal of Forecasting, 12*, 617–637.

Cooper, L. G. (1993). Market-Share Models. In: J. Eliashberg & G. L. Lilien (Eds), *Handbook in Operations Research and Management Science* (Vol. 5, Chap. 6, pp. 259–314). Amsterdam: North-Holland.

Cooper, L. G., & Nakanishi, M. (1988). *Market Share Analysis: Evaluating Competitive Marketing Effectiveness*. Boston: Kluwer Academic Publishers.

D'Agostino, R. B. (1970). Transformation to Normality of the Null Distribution of $g_1$. *Biometrika, 57*, 679–681.

Danaher, P. J. (1994). Comparing Naive with Econometric Market Share Models when Competitors' Actions are Forecast. *International Journal of Forecasting, 10*, 287–294.

Dekimpe, M. G., & Hanssens, D. M. (1995). The Persistence of Marketing Effects on Sales. *Marketing Science, 14*, 1–21.

Doornik, J. A., & Hansen, H. (1994). *An Omnibus Test for Univariate and Multivariate Normality*. http://www.nuff.ox.ac.uk/Users/Doornik/index.html.

Fok, D., & Franses, P. H. (2001a). Analyzing the Effects of a Brand Introduction on Competitive Structure Using a Market Share Attraction Model. Unpublished Working Paper, Erasmus University Rotterdam.

Fok, D., & Franses, P. H. (2001b). Forecasting Market Shares from Models for Sales. *International Journal of Forecasting, 17*, 121–128.

Fok, D., Franses, P. H., & Paap, R. (2001). Short-Run and Long-Run Dynamics in the Market Share Attraction Model. Unpublished Working Paper, Erasmus University Rotterdam.

Franses, P. H., Kloek, T., & Lucas, A. (1999). Outlier Robust Analysis of Long-Run Marketing Effects for Weekly Scanning Data. *Journal of Econometrics, 89*, 293–316.

Franses, P. H., & Paap, R. (2001a). *Quantitative Models in Marketing Research*. Cambridge: Cambridge University Press.

Franses, P. H., & Paap, R. (2001b). Selecting a Market Share Attraction Model. Unpublished Working Paper, Erasmus University Rotterdam.

Franses, P. H., Srinivasan, S., & Boswijk, P. (2001). Testing for Unit Roots in Market Shares. *Marketing Letters, 12*, in press.

Ghosh, A., Neslin, S., & Shoemaker, R. (1984). A Comparison of Market Share Models and Estimation Procedures. *Journal of Marketing Research, 21*, 202–210.

Greene, W. H. (1993). *Econometric Analysis*. New Jersey: Prentice-Hall Inc.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hsu, A., & Wilcox, R. T. (2000). Stochastic Prediction in Multinomial Logit Models. *Management Science, 46*, 1137–1144.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Judge, G. G., Griffiths, W., Hill, R. C., Lütkepohl, H., & Lee, T.-C. (1985). *The Theory and Practice of Econometrics* (2nd ed.). New York: John Wiley & Sons.

Kumar, V. (1994). Forecasting Performance of Market Share Models: An Assessment, Additional Insights, and Guidelines. *International Journal of Forecasting, 10*, 295–312.

Leeflang, P. S. H., & Reuyl, J. C. (1984). On the Predictive Power of Market Share Attraction Model. *Journal of Marketing Research, 21*, 211–215.

Leeflang, P. S. H., Wittink, D. R., Wedel, M., & Naert, P. (2000). *Building Models for Marketing Decisions*. Dordrecht: Kluwer Academic Publishers.

Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis* (2nd ed.). Berlin: Springer-Verlag.

Naert, P. A., & Weverbergh, M. (1981). On the Prediction Power of Market Share Attraction Models. *Journal of Marketing Research, 18*, 146–153.

Paap, R. (2002). What are the Advantages of MCMC-Based Inference in Latent Variable Models? *Statistica Neerlandica, 56*, 1–21

Paap, R., & Franses, P. H. (2000). A Dynamic Multinomial Probit Model for Brand Choice with Different Long-run and Short-run Effects of Marketing-Mix Variables. *Journal of Applied Econometrics*, *15*, 717–744.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, *6*, 461–464.

Srinivasan, S., & Bass, F. M. (2000). Cointegration Analysis of Brand Sales and Category Sales: Stationarity and Long-run Equilibrium in Market Shares. *Applied Stochastic Models in Business and Industry*, *16*, 159–177.

Van Heerde, H. J., Leeflang, P. S. H., & Wittink, D. R. (2000). The Estimation of Pre- and Postpromotion Dips with Store-Level Scanner Data. *Journal of Marketing Research*, *37*, 383–395.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias. *Journal of the American Statistical Association*, *57*, 348–368.

# APPENDIX

## *Estimation of Restricted Covariance Matrix*

Recall the log likelihood function (28)

$$\ell(\tilde{\Sigma}) = -\frac{T(I-1)}{2}\log(2\pi) + \frac{T}{2}\log|\tilde{\Sigma}^{-1}| - \frac{1}{2}\hat{\eta}'(\tilde{\Sigma}^{-1}\otimes \mathbf{I}_T)\hat{\eta}, \tag{53}$$

where $\tilde{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_{I-1}^2) + \sigma_I^2 \mathbf{i}_{I-1}\mathbf{i}'_{I-1}$. For $i = 1, \ldots, I-1$ it holds that

$$\frac{\partial \ell(\tilde{\Sigma})}{\partial \sigma_i} = \left(\frac{\partial \ell(\tilde{\Sigma})}{\partial \mathrm{vec}(\tilde{\Sigma}^{-1})}\right)' \frac{\partial \mathrm{vec}(\tilde{\Sigma}^{-1})}{\partial \sigma_i}$$

$$\begin{aligned}\frac{\partial \ell(\tilde{\Sigma})}{\partial \mathrm{vec}(\tilde{\Sigma}^{-1})} &= \frac{T}{2}\frac{\partial \log|\tilde{\Sigma}^{-1}|}{\partial \mathrm{vec}(\tilde{\Sigma}^{-1})} - \frac{1}{2}\frac{\partial \hat{\eta}'(\tilde{\Sigma}^{-1}\otimes \mathbf{I}_T)\hat{\eta}'}{\partial \mathrm{vec}(\tilde{\Sigma}^{-1})} \\ &= \frac{T}{2}\mathrm{vec}(\tilde{\Sigma}) - \frac{1}{2}\mathrm{vec}\left(\begin{pmatrix}\hat{\eta}'_1 \\ \vdots \\ \hat{\eta}'_{I-1}\end{pmatrix}(\hat{\eta}_1, \ldots, \hat{\eta}_{I-1})\right) \qquad (54) \\ &= \frac{1}{2}\mathrm{vec}[T\tilde{\Sigma} - \begin{pmatrix}\hat{\eta}'_1 \\ \vdots \\ \hat{\eta}'_{I-1}\end{pmatrix}(\hat{\eta}_1, \ldots, \hat{\eta}_{I-1})]\end{aligned}$$

and

$$\frac{\partial \text{vec}(\tilde{\Sigma}^{-1})}{\partial \sigma_i} = \text{vec}\left( -\tilde{\Sigma}^{-1} \frac{\partial \tilde{\Sigma}}{\partial \sigma_i} \tilde{\Sigma}^{-1} \right) = -(\tilde{\Sigma}^{-1})^2_{ii} e_{i,I-1}$$

$$\frac{\partial \ell(\tilde{\Sigma})}{\partial \sigma_i} = \frac{1}{2} \text{tr}[-T\tilde{\Sigma}(\tilde{\Sigma}^{-1})^2_{ii} e_{i,I-1} \begin{pmatrix} \hat{\eta}'_1 \\ \vdots \\ \hat{\eta}'_{I-1} \end{pmatrix} (\hat{\eta}_1, \ldots, \hat{\eta}_{I-1})(\tilde{\Sigma}^{-1})^2_{ii} e_{i,I-1}]$$

$$= \frac{1}{2} [-T(\tilde{\Sigma})_{ii}(\tilde{\Sigma}^{-1})^2_{ii} + \hat{\eta}'_i \hat{\eta}_i (\tilde{\Sigma}^{-1})^2_{ii}]$$

$$= \frac{1}{2} (\tilde{\Sigma}^{-1})^2_{ii} [\hat{\eta}'_i \hat{\eta}_i - T(\sigma_i^2 + \sigma_I^2)], \tag{55}$$

where $e_{i,k}$ is a zero vector of size $(k \times 1)$ with the $i$-th element equal to 1. Solving the last equation given $\hat{\sigma}_I^2$ yields

$$\hat{\sigma}_i^2 = \frac{\hat{\eta}'_i \hat{\eta}_i}{T} - \hat{\sigma}_I^2. \tag{56}$$

The concentrated likelihood is obtained by inserting (56) into the likelihood (53). The concentrated likelihood now has to be optimized over just one parameter, that is $\sigma_I$.

# REFLECTING UNCERTAINTY ABOUT ECONOMIC THEORY WHEN ESTIMATING CONSUMER DEMAND

Alan L. Montgomery


## ABSTRACT

*Economic theory provides a great deal of information about demand models. Specifically, theory can dictate many relationships that expenditure and price elasticities should fulfill. Unfortunately, analysts cannot be certain whether these relationships will hold exactly. Many analysts perform hypothesis tests to determine if the theory is correct. If the theory is accepted then the relationships are assumed to hold exactly, but if the theory is rejected they are ignored. In this paper we outline a hierarchical Bayesian formulation that allows us to consider the theoretical restrictions as holding stochastically or approximately. Our estimates are shrunk towards those implied by economic theory. This technique can incorporate information that a theory is approximately right, even when exact hypothesis tests would reject the theory and ignore all information from it. We illustrate our model with an application of this data to a store-level system of demand equations using supermarket scanner data.*


**Econometric Models in Marketing, Volume 16, pages 257–294.**

**ISBN: 0-7623-0857-5**

# I. INTRODUCTION

A basic goal of many marketing analysts and econometricians is the estimation of consumer demand models. More specifically analysts might be interested in estimating price and promotional elasticities that can be used in developing better marketing strategies. Economics provides a large body of theory to guide an analyst in constructing a consumer demand model. Unfortunately, the analyst can never be entirely confident that this theory is correct. In practice many marketing analysts may assume that nothing is known about expenditure and price elasticities due to their uncertainty about whether all theoretical assumptions are met. However, even if the assumptions of these theories are not met exactly the theory might still be approximately correct. It is this notion of approximation that we formalize in this paper.

The focus of many econometric studies is to determine the extent that the data supports a particular theory. Classical approaches to testing lead the analyst to an all or nothing approach. If the data provides strong confirmatory evidence then the analyst usually proceeds under the assumption the theory is correct and estimates the model. However, if the theory is rejected then the analyst simply rejects the theory and ignores all information from the theory. Sharp tests of null hypotheses in large datasets frequently lead to rejection if the tolerance for type I errors is not increased with the sample size. Large datasets can result in very precise tests that often miss the fact that the theory may not be perfect but provides a reasonable approximation to the true process.

In this paper we propose a Bayesian framework in which uncertainty about a theory is directly represented in the model. Our procedure prescribes treating the theory as a prior and follows recent work by Montgomery and Rossi (1999). The prior is centered over the theory, so the mean is what would be expected under a restricted model in which the theory holds exactly. The variance of the prior is allowed to vary depending upon the analyst's confidence about the theory. For example, Slutsky symmetry may require equating two parameters. In our methodology we can represent these two parameters as two draws from a common distribution, which we call the hyper-prior. If we are certain that the theory holds exactly then the variance of this hyper-prior is zero, and the restrictions are implicitly fulfilled. However, we wish to entertain the notion that the theory may only be approximately correct. Hence we allow the variance of the hyper-prior to vary, perhaps substantially. We may be uncertain about the exact values of the parameters of this distribution and place a prior on the parameters of this hyper-prior.

The analyst can incorporate prior beliefs about the adequacy of the theory and gain useful information even if the theory is technically wrong, but is approximately right. It is this notion of approximation that we are especially interested in representing. The estimator proposed results in adaptive shrinkage towards the theory. Adaptivity refers to the ability of the model to decrease the amount of shrinkage if the data disagrees with the prior. As more information or data is observed less shrinkage occurs and we can learn more about how good an approximation the theory provides to the observed data. Our framework allows the flexibility to mimic the estimates of a model achieved by an economist who holds to theory dogmatically, an analyst who ignores theory entirely, or an analyst's whose beliefs fall in between by choosing the prior appropriately. Our framework also contrasts with statistical formulations of shrinkage estimators in marketing that move estimates towards one another due to empirical similarities without any theoretical justification (Blattberg & George, 1991; Montgomery, 1997).

Economic theory provides many possible sources of information. First, it can provide information about relationships that elasticities should satisfy, such as adding up or Slutsky symmetry. Second, specific assumptions about utility may result in more parsimonious demand models. For example, the assumption of additive utility results in a very parsimonious model. Many marketing models, like logit choice models and conjoint models, are based upon the assumption of an additive utility model. Third, elasticity estimates for one economic agent may be similar to those of other agents. Finally, previous empirical research may enable us to directly postulate priors on the parameters, i.e. the elasticity matrix is made up of negative elements on the diagonal (negative own-price elasticities) and small positive cross-diagonal elements (modest direct substitution between products within a category). In this paper we show how these prior sources of information can be parameterized and incorporated into a hierarchical Bayesian framework.

Previous research in marketing has considered economic restrictions in demand models (Berndt & Silk, 1993), restricted relationships between elasticities (Allenby, 1989) in the context of market structure, and the use of hierarchical models to shrink estimates across stores and households (Blattberg & George, 1991; Allenby & Rossi, 1993; Montgomery, 1997). Our framework provides a unifying treatment to these ideas. By evaluating these components together we can appreciate the significant gains in measuring demand that can be had by incorporating theory in a stochastic manner.

The outline of this paper is as follows. First we present our demand model in Section 2 and the restrictions implied by economic theory. Section 3 goes on to show how these restrictions can be incorporated stochastically in a hierarchical Bayesian model. A short example is given to illustrate these restrictions. The estimation of this Bayesian treatment is presented using the Gibbs Sampler in Section 4. Section 5 provides an empirical example of shrinkage of price elasticities towards those restrictions implied by an additive utility model. This example estimates store level demand systems using weekly UPC scanner data for the refrigerated orange juice category at Dominick's Finer Foods (DFF), a major Chicago supermarket chain. Section 6 considers a further application of this framework by considering changes in market structures. We conclude the paper in Section 7 with a discussion of these results along with suggestions for implementing these techniques in other problems.

## 2. SALES RESPONSE MODELING

We begin not with a formal theory of consumer behavior from which we derive a model of demand as in customary in econometrics, but with a sales response model. Both models try to capture the relationship between quantity and price, the essential difference is in terms of interpretation. A sales response model is a model motivated by statistical considerations, for example a logarithmic relationship between quantity and price is commonly observed by marketing researchers, and is not justified on theoretical grounds. For a discussion of sales response modeling from a marketing perspective see Blattberg and Neslin (1990). On the other hand an econometric model places many restrictions upon the functional form and parameters. The strength of the econometric model is our ability to estimate more parsimonious forms, while its weakness is the requirement to make many assumptions that may be suspect or untestable. In contrast, these strengths are reversed for a sales response model. It makes fewer assumptions about demand, but this flexibility comes at the price of an increased number of parameters.

To begin our analysis of demand we choose a double log functional form for our sales response model. This form is chosen since previous empirical work has shown it to be a good one that captures the logarithmic relationship between quantity and price. Our technique is quite general and does not rely upon a logarithmic functional form, in fact it could be applied to many demand models, such as the AIDS, translog, or Rotterdam model. Our sales response model can be written in vector form:

$$\ln(q_{st}) = \alpha_s + \mu_s \ln(x_s) + \mathrm{H}_s \ln(p_{st}) + e_{st},\ e_{st} \sim N(0, \Sigma_s) \tag{1}$$

Where there are $M$ products in store $s$ at week $t$, $q_{st}$ and $p_{st}$ are vectors of movement and price, and $x_{st}$ is store expenditures ($x_{st} = \Sigma_i p_{ist} q_{ist}$, the $i$th subscript denotes the $i$th product in the vector). Our framework is parameterized by the store subscript $s$, although this index can be interpreted quite generally as an index for different households, markets, or industries, depending upon the application.

The basic problem one encounters in estimating model (1) is the large number of parameters. For example, if there are 10 products and 100 stores as would be found in one small category of a moderately sized retailer, this results in more than 10,000 parameters that must be estimated. In a typical supermarket retailing application perhaps two or three years of weekly observations would be available. While this is a large amount of data, if the retailer wishes to estimate demand for each store separately then it may be difficult to estimate store-level demand with any degree of statistical precision. This problem becomes acute if the retailer wishes to formulate an elasticity based pricing strategy, since the high degree of parameter uncertainty may result in strange pricing prescriptions. For example, positive own-price elasticities may result in undefined optimal prices, or erroneously signed cross-price elasticities may result in higher overall levels of prices.

### 2.1. An Economic Interpretation of the Sales Response Model

We can reinterpret our sales response model in (1) as a system of demand equations. The H represents uncompensated price elasticities and the $\mu$ are expenditure elasticities. Usually $x$ would represent income, and demand would be defined over all products consumed. However, we do not have a measure of weekly income for consumers that shop at store $s$. Therefore, we use store expenditures[1] and consider (1) as a subset demand model for the products in store $s$. Subset demand models possess all the usual properties of full demand models, although the income elasticities are now interpreted as store expenditure elasticities. For a further discussion of subset demand models see Deaton and Muellbauer (1983).

A store expenditure elasticity states how product sales are effected as store shoppers purchase more groceries. Specifically $\mu_i$ states the effect of an increase of store expenditures on the movement for product $i$. If $\mu_i < 0$ then product sales decrease as expenditures grow (an inferior product), and when $\mu_i > 1$ product sales garner a larger share of overall sales. Since, this expenditure elasticity is conditional upon store sales, it cannot be used to determine how store traffic is affected by competition and cross-category promotions.

The price elasticity matrix can be decomposed into expenditure and price effects:

$$H_s = E_s - \mu_s w_s' \tag{2}$$

Where the uncompensated cross elasticity ($H_s$) for store $s$ is the sum of a substitution effect, the compensated cross elasticity matrix ($E_s$), and an income effect, which is the outer product of the income elasticities ($\mu_s$) and the budget or market shares ($w_s$). The $i$th element of the market share vector is defined as $w_{is} = p_{ist} q_{ist} / x_{st}$. We use the usual definition of substitutes ($[E_s]_{ij} > 0$), complements ($[E_s]_{ij} < 0$), and independent products ($[E_s]_{ij} = 0$) that rely upon compensated elasticities.

Substituting (2) into (1) yields a demand system in terms of compensated elasticities. We also augment this model with cross- feature and deal variables to control for other marketing mix effects. Finally, we assume that the category employed in our analysis is independent of other categories, so our system only uses the set of products within a category. The final form of the demand model that we employ in this paper is:

$$\ln(q_{st}) = \alpha_s + \mu_s \ln(x_s / P_{st}) + E_s p_{st} + \Theta_s f_{st} + \Psi_s d_{st} + \epsilon_{st}, \ \epsilon_{st} \sim N(0, \Sigma_s) \qquad (3)$$

Where $P_{st} = \exp\{\Sigma_i w_{ist} \ln(p_{ist})\}$ is a Divisia price index, $f_{st}$ and $d_{st}$ are the vectors of feature and display variables for store $s$ during week $t$.

### 2.2. *Economic Theory*

If we interpret (3) not as a sales response model, but as a system of demand equations then economic theory is very informative about the parameters or more specifically the conditions that the price elasticities must satisfy. These restrictions follow as a consequence of underlying assumptions about utility: reflextivity, completeness, transitivity, continuity, and nonsatiation. In our discussion we only express the consequences of these assumptions on demand and do not provide their derivations. For additional reference we refer the reader to Deaton and Muellbauer (1983, pp. 43–46).

*Adding-Up*: The budget constraint imposes the following condition on demand:

$$p_s' q_s = x_s \qquad (4)$$

This equation can be differentiated with respect to price and expenditures to yield the following:

$$w_s' \mu_s = 1 \qquad (5)$$

and

$$w_s' H_s = w_s \Rightarrow w_s' E_s = 0 \qquad (6)$$

These restrictions reduce our demand system by 1 and $M$ parameters respectively.

*Homogeneity*: The assumption of homogeneity implies that if we double all prices and income then the budget shares remain unchanged (no money illusion):

$$H_s\iota = \mu_s \Rightarrow E_s\iota = 0, \quad \text{where } \iota = (1 \ 1 \ \cdots \ 1)' \tag{7}$$

This restriction reduces our demand system by an additional $M$ parameters.

*Symmetry*: The symmetry restriction is derived from the double differentiability of the cost function or the symmetry of the Slutsky matrix (L),[2] and implies that the compensated elasticity matrix when weighted by the budget shares is symmetric:

$$diag(w_s)E_s = E_s' \, diag(w_s) \tag{8}$$

Notice that symmetry results in a large reduction in the order of the demand system, specifically by $\frac{1}{2}M(M-1)$ terms or a 45% reduction in the cross-price elasticities with 10 products ($M=10$).

Many marketers may worry that Slutsky symmetry may be too restrictive. It is well established in marketing (Blattberg & Wiesniewski, 1989; Kamakura & Russell, 1989) that uncompensated price elasticity matrices are asymmetric. For example price changes of higher quality brands effect sales of lower quality brands, but price changes of lower quality brands have only small effects on high quality brands. These asymmetries are consistent with economic theory and can be explained by differences in market shares and expenditure elasticities, and do not require asymmetries in the compensated elasticity matrix. Consider an example with three brands (premium, national, and store brands) and the following parameters:

$$\mu = \begin{bmatrix} 1.5 \\ 1.0 \\ 0.5 \end{bmatrix}, \ w = \begin{bmatrix} 0.4 \\ 0.3 \\ 0.3 \end{bmatrix}, \ E = \begin{bmatrix} -2.0 & 0.50 & 0.50 \\ 0.67 & -2.0 & 0.50 \\ 0.67 & 0.50 & -2.0 \end{bmatrix}$$

Employing (8) we find the uncompensated price elasticity matrix becomes:

$$H = \begin{bmatrix} -2.6 & 0.05 & 0.05 \\ 0.27 & -2.3 & 0.20 \\ 0.47 & 0.35 & -2.15 \end{bmatrix}$$

The asymmetry in the compensated elasticity matrix (E) between the premium and national brands is due to market share differences ($w$), while there is no assymetry between the national and store brands. However, upon evaluation of the uncompensated elasticity matrix (H), we find pronounced price asymmetries between these three brands. The asymmetry in price elasticities is due to expenditure effects ($\mu$), i.e. as expenditures grow people purchase higher quality brands.

*Sign-Restrictions*: Downward sloping demand curves require the Slutsky matrix to possess a negative semi-definite property:

$$\forall\delta \quad \delta' L_s \delta \leq 0 \tag{9}$$

In addition to the usual consequence that the own-price elasticities must be non-positive, it further implies that any linear bundle of products must also have a non-positive elasticity. A common concern in marketing is that price elasticities can frequently be of the wrong sign.

## 2.3. *Weak Separability and Market Structure*

Another component of economic theory that can induce relationships among price elasticities are ones about the relationships between products. Many marketing researchers have suggested a hierarchical structure for market competition (to name just a few see Allenby, 1989; Vilcassim, 1989; Srivastava et al., 1981). This hierarchy is illustrated in Fig. 1. For example, a consumer first decides whether to buy liquid or dry laundry detergent, and then considers which product to buy within the subcategory. Products at the same level within a branch are strong substitutes, while competition between items in different branches is weaker and have the same general pattern.

At the heart of most research on market structure is weak separability of the utility function. Frequently these hierarchical structures are justified by assuming that consumers engage in some type of hierarchical budgeting process. Allocating budget shares to large groups of products like groceries, housing, transportation, etc., and then deciding upon allocations to individual products within each category. This broad budget allocation process allows us to break the problem into smaller units by assuming groups of products within

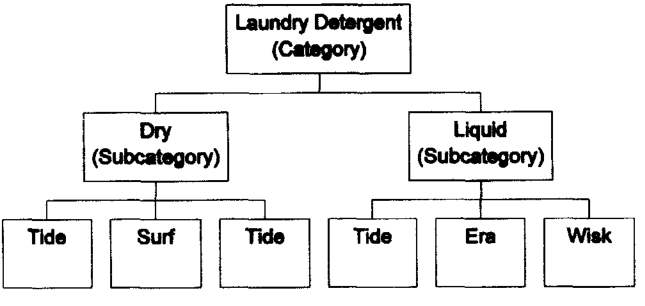


*Fig. 1.* Example of a Hierarchical Market Structure for Laundry Detergent with Dry and Liquid Subcategories.

a category can be weakly separated from one another. Categories can be partitioned into subcategories, until finally we reach the individual product level. The general form of the utility function for an individual household is of the form:

$$utility = \nu(\nu_1(q_1), \nu_2(q_2), \cdots, \nu_C(q_C)) \tag{10}$$

Where $q_i$ is the vector of quantities for all items in the $i$th category of which there are $C$ categories.

The hierarchy in the cost or utility functions naturally imposes a structure in the demand model. It can be shown that weak separability imposes the following restriction on the elasticity matrix:

$$\epsilon_{ijs} = \kappa_{GHs}\mu_{is}\mu_{js}w_{js} \quad \text{if } i \in G \text{ and } h \in H \tag{11}$$

where $\epsilon_{ijs}$ is the $i, j$th element of the matrix $E_s$ and $\kappa_{GHs}$ is a parameter that may depend upon $x$. In other words, the elasticities that capture substitution within a category can take on a general form, but those elasticities representing intra-category substitution must follow a restricted pattern that is common for all items in the subcategories.

### *2.4. Strong Separability and Additive Utility*

The restrictions discussed in the previous subsection hold for many families of utility functions. If the analyst is willing to make stronger assumptions about a specific form of utility then this can also result in much simpler forms to demand. One possibility is to assume utility is additive or strongly separable across products:

$$utility = \nu(\Sigma_i\, \nu_i(q_i)) \tag{12}$$

where $q_i$ is the quantity of the $i$th product consumed. Additivity has a long history in economic models (Lancaster, 1966) and empirical applications in marketing like logit modeling (Guadagni & Little, 1983) and conjoint analysis (Green & Rao, 1971). Often additivity is argued at the attribute level in logit and conjoint applications and not the higher, product level as we have suggested.

Additive utility models result in parsimonious – but restrictive – demand models:

$$E_s = \phi_s\, diag(\mu_s) - \phi_s\mu_s(\mu_s \circ w_s)' \tag{13}$$

Notice that the cross-elasticity matrix is populated solely by the expenditure elasticities ($\mu$), market shares ($w$), and a general substitution parameter ($\phi$).

This restricted elasticity matrix has $M+1$ parameters, not including the market shares, as opposed to $M^2+M$ for the unrestricted form. Additionally, the elasticity matrix in (13) will satisfy the properties of demand models given in the previous subsections. However, the incredible parsimony of the additive model also comes a high penalty. Namely, either all products must be substitutes or complements, and the level of substitution or complementarity is dictated by a single parameter ($\phi$) and the expenditure elasticities.

It might seem odd to many economists to propose an additive utility structure, since many econometric studies have rejected additivity (Barten, 1969; Deaton, 1974; Theil, 1976; Deaton, 1978). However, we are proposing an additive utility structure at a very low-level (e.g. similar products within a single category), while most have considered additivity at high levels in a hierarchical structure (food, clothing, housing). Additive utility implies that the utility gained from one product is unaffected by the utility of other products. For example, there is no interaction in utility from purchasing Minute Maid and Tropicana orange juice together. This makes a great deal of sense for products within a category, which are typically direct substitutes and not used together. However, additivity may not make sense across products from different categories that when combined together can interact, such as bread and peanut butter.

### *2.5. Pooling and Heterogeneity*

The last set of restrictions that we propose are not really theoretical ones, but ones motivated from practice. It is quite common to observe multiple agents, either consumers or stores as in our case. A common assumption is to simply pool the observations across all agents and assume identical elasticities as in the following relationship:

$$\mathrm{E}_s = \mathrm{E},\ \mu_s = \mu \tag{14}$$

Recently there has been a great deal of research in marketing studying heterogeneity, for a recent review refer to Rossi and Allenby (2000). One technique is to capture heterogeneity in a random coefficient model:

$$\mathrm{E}_s = \mathrm{E} + U_s,\ \mu_s = \mu + u_s \tag{15}$$

This specification has been studied extensively starting with the early work by Swamy (1970) from a frequentist perspective and by Lindley and Smith (1972) from a Bayesian interpretation as a hierarchical Bayesian model.

## 3. A BAYESIAN SPECIFICATION

The economic theory proposed in Section 2 is simply that, a theory. As with any theory the assumptions upon which it is based are subject to question. One technique is to inject randomness into the axioms upon which the theory is constructed, namely reflexivity, completeness, transitivity, continuity, non-satiation, and convexity. However, our belief is that the theory is quite reasonable. But we also realize that there are many reasons to believe that this model may not be entirely correct. Our theory is at an individual level, but our data is at an aggregate level. We know that aggregate demand models will satisfy additivity and other economic properties only under certain conditions (see Deaton & Muellabuer, 1983, pp. 148–166) for a discussion of the conditions for exact aggregation to hold). Additionally, these theoretical relationships are abstractions that omit certain effects (intertemporal substitution, savings, nonlinear budget constraints, etc.) or be subject to measurement errors.

Our belief is that the theory should be a reasonable approximation to the observed process, but will hold approximately or in a stochastic manner and not exactly. This contrasts with the usual pre-testing approach which would test whether these effects hold exactly and then totally discard them if they do not meet a specific p-value. An essential difference is that in our framework – even if the theory is not entirely supported by the data – the information implied by the theory will not be completely ignored. To explicitly incorporate the notion of approximation into our model follow the approach proposed by Montgomery and Rossi (1999). First, we assume that the price elasticities have the following prior distribution:

$$\epsilon_s | \beta_s, \phi \sim N(\bar{\epsilon}_s, \Delta) \tag{16}$$

where

$$\beta_s = [\beta'_{1s} \ \beta'_{2s} \ \cdots \ \beta'_{Ms}]', \ \beta_{is} = [\alpha_{is} \ \mu_{is} \ \theta_{is} \ \psi_{is}]' \tag{17}$$

This distribution will be centered around the restrictions implied by our theory, $\bar{\epsilon}_s$, and the variance around these restrictions represents our confidence in this approximation. $\Delta$ can be interpreted as the degree to which an approximation is valid. If $\Delta$ is small then these restrictions will effectively be enforced. Conversely large values of $\Delta$ will result in estimates that may bear little resemblance to the restricted parameter estimates, i.e. unrestricted parameter estimates.

We are not able to assess the parameters of this prior directly, so we place a prior on this prior. To avoid confusion the prior in (16) is called the hyper-prior.

Additionally, we assume that an elasticity matrix that conforms to an additive utility structure is reasonable, which implicitly satisfies all the relationships outlined in section 2. We propose the following relationship:

$$\epsilon_s = \text{vec}(E_s),\ \bar{\epsilon}_s = \text{vec}(\bar{E}_s),\ \bar{E}_s = \phi_s\ \text{diag}(\mu_s) - \phi_s\mu_s(\mu_s \circ w_s)' \tag{18}$$

We place the usual multivariate normal prior on the remaining store parameters:

$$\phi_s \sim N(\bar{\phi}, \lambda_\phi) \tag{19}$$

$$\beta_s \sim N(\bar{\beta}_s, \Lambda) \tag{20}$$

An important reason for expressing the prior on $E_s$ conditionally upon $\mu_s$ and $\phi_s$ is to avoid problems on nonlinearity. Notice that while $E_s$ is conditionally linear upon $\mu_s$, unconditionally our prior is nonlinear in $\mu_s$. Additionally the prior implies that the price elasticity elements will be correlated, which can help counter the effect of multicollinearity in a typical price dataset.

Notice that our priors on $E_s$, $\mu_s$, and $\phi_s$ are exchangeable across stores. It is this exchangeability that will drive the shrinkage of one store's parameter estimates towards another. The store to store variation of the expenditure elasticities ($\mu_s$) is governed by $\Lambda$, and variation in the price elasticity matrix ($E_s$) – both across store and deviations from the theory – is governed by the $\Delta$ matrix. If $\Lambda$ and $\Delta$ are zero then there will be no random variation across stores and the cross elasticity matrix will be held to its restricted pattern, i.e. the estimates will be close to a pooled restricted model. If $\Lambda$ and $\Delta$ are large then the information from the hyper-distribution will be discounted and the parameter estimates will be close to individual store models.

Since we cannot directly evaluate $\Lambda$ and $\Delta$, we formulate a prior on these matrices, and use the data to make inferences about the variation present in the data. In our Bayesian framework we assume independent Wishart priors for each of these matrices:

$$\Phi^{-1} \sim Wishart(\nu_\Delta, V_\Delta^{-1}),\ \Lambda^{-1} \sim Wishart(\nu_\Lambda, V_\Lambda^{-1}) \tag{21}$$

We parameterize the prior on these priors as: $V_\Delta = \nu_\Delta k_\Delta \bar{V}_\Delta$ and $V_\Lambda = \nu_\Lambda k_\Lambda \bar{V}_\Lambda$, so that these priors are centered over $\bar{V}_\Delta^{-1}/k_\Delta$ and $\bar{V}_\Lambda^{-1}/k_\Lambda$, respectively.

The use of independent priors on $\Lambda$ and $\Delta$ as in Montgomery and Rossi (1999) provides an important point of divergence with previous work in marketing research that uses a single joint Wishart prior on these matrices (Blattberg & George, 1991; Montgomery, 1997). The problem with a single inverted Wishart prior on the variance of $\beta_s$ and $\epsilon_s$ is a lack of flexibility. Once the mean of the distribution is set then the dispersion around this mean is controlled by a single scaling parameter. However, we want a prior that will allow for differential degrees of freedom on how tight the prior should be on $\beta_s$ and $\epsilon_s$. Specifically in our problem we wish to have a prior that may allow differential amounts of shrinkage across stores and towards the theory. For example, we may wish to have more cross-store shrinkage than shrinkage towards the theory, i.e. $\Lambda > \Delta$.

To illustrate this problem consider Fig. 2 which illustrates the inverted Wishart prior for two diagonal elements in the corresponding panels. Once the dispersion is set for the first element, the dispersion for the second element is automatically fixed, as denoted by a solid line. If we wish to loosen up the prior on the first element to increase the amount of shrinkage (there is an inverted relationship), this would also increase the shrinkage of the second element, as denoted by the dashed line. However, we wish to have the ability to tighten up the prior on the first element without altering the second element, i.e. choose the dashed line for the first parameter and the solid line for the second parameter. The introduction of two independent priors allows for this type of differential shrinkage.

Recent work by Barnard et al. (2000) on decomposing the prior on the covariance matrix into the standard deviations and correlation matrices can also allow differential shrinkage.

### 3.1. An Example

To illustrate the framework presented in the previous subsection consider an example with three products. We use our demand model from (3) without promotional variables:

$$\begin{bmatrix} \ln(q_{1ts}) \\ \ln(q_{2ts}) \\ \ln(q_{3ts}) \end{bmatrix} = \begin{bmatrix} \alpha_{1s} \\ \alpha_{2s} \\ \alpha_{3s} \end{bmatrix} + \begin{bmatrix} \mu_{1s} \\ \mu_{2s} \\ \mu_{3s} \end{bmatrix} \ln(x_{ts}/P_{ts}) + \begin{bmatrix} \epsilon_{11,s} & \epsilon_{12,s} & \epsilon_{13,s} \\ \epsilon_{12,s} & \epsilon_{22,s} & \epsilon_{23,s} \\ \epsilon_{31,s} & \epsilon_{32,s} & \epsilon_{33,s} \end{bmatrix} \begin{bmatrix} \ln(p_{1ts}) \\ \ln(p_{2ts}) \\ \ln(p_{3ts}) \end{bmatrix}$$

$$+ \begin{bmatrix} e_{1ts} \\ e_{2ts} \\ e_{3ts} \end{bmatrix}, \ e \sim N(0, \Sigma) \tag{22}$$

The hyper-parameters are:

$$\bar{\mu} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \ \bar{\alpha} = \begin{bmatrix} 8 \\ 7 \\ 6 \end{bmatrix}, \ \Lambda = 0.1 I_6, \ \Delta = 0.01 I_9, \ \bar{\phi} = -3, \ \lambda_\phi = 1$$

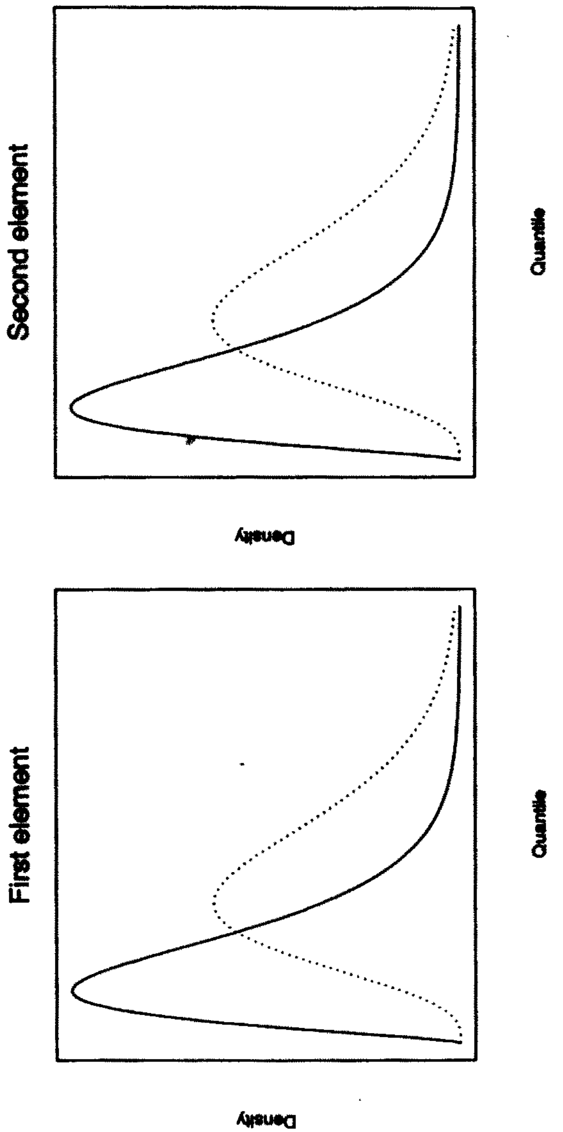


*Fig. 2.* Two Selected Elements of an Inverted Wishart Distribution.

Suppose the draw for an individual store is:

$$\phi_s = -3,\ \mu_s = \begin{bmatrix} 1.4 \\ 0.8 \\ 0.8 \end{bmatrix},\ w_s = \begin{bmatrix} 0.33 \\ 0.33 \\ 0.33 \end{bmatrix},\ \mathrm{E}_s = \begin{bmatrix} -2.3 & 1.1 & 1.1 \\ 1.1 & -1.7 & 0.6 \\ 1.1 & 0.6 & -1.7 \end{bmatrix}$$

The restricted price elasticity implied by this specific model would be:

$$\phi_s = -3,\ w_s = \begin{bmatrix} 0.33 \\ 0.33 \\ 0.33 \end{bmatrix},\ \bar{\mathrm{E}}_s = \begin{bmatrix} -2.3 & 1.1 & 1.1 \\ 1.1 & -1.7 & 0.6 \\ 1.1 & 0.6 & -1.7 \end{bmatrix}$$

Notice that this restricted price elasticity reflects the high own-price sensitivity and small cross-price elasticities that is usually observed in empirical work.

The price elasticity estimates for this individual store will be shrunk towards the restricted price elasticity matrix. This contrasts with Blattberg and George (1991) who propose shrinking all own price terms (deflated by relative prices) to a single value. Their structure would result in price terms being shrunk towards:

$$\begin{bmatrix} -2.0 & 0.5 & 0.5 \\ 0.5 & -2.0 & 0.5 \\ 0.5 & 0.5 & -2.0 \end{bmatrix}$$

Notice that Blattberg and George (1991) can be thought of as a special case of our framework. The shrinkage pattern they suggested is the same as ours when market shares and expenditure elasticities are equal. However, market shares are rarely equal and we may expect some brands to benefit from category expenditures more than others (unequal expenditure elasticities). An advantage of our framework is that we can evaluate the shrinkage of the estimates in terms of theoretical properties of our model, and not rely upon empirical justifications. This is an important distinction since it permits evaluation of shrinkage in terms of utility and not ad hoc empirical justifications.

## 4. ESTIMATION

We rewrite our model in SUR form:

$$y_s \sim N(X_s\beta_s,\ \Sigma_s \otimes I_T),\ \Sigma_s^{-1} \sim W(\nu_\Sigma,\ \bar{V}_\Sigma^{-1}) \tag{27}$$

In this case the $s$ subscript denotes an individual store, and the dimension of the $y_s$ vector is $M$ brands by $T$ weeks. In rewriting the model we have implicitly stacked the vector of observations for each brand on top of one another in the following manner:

$$y_s = \begin{bmatrix} q_{1s} \\ \vdots \\ q_{Ms} \end{bmatrix}, \quad q_{is} = \begin{bmatrix} q_{i1s} \\ \vdots \\ q_{iTs} \end{bmatrix}, \quad X_s = \begin{bmatrix} X_{1s} & & \\ & \ddots & \\ & & X_{Ms} \end{bmatrix},$$

$$X_{is} = \begin{bmatrix} 1 & \ln(x_{1s}/P_{1s}) & p_{11s} & \cdots & p_{M1s} & f_{i1s} & d_{i1s} \\ 1 & \ln(x_{2s}/P_{2s}) & p_{12s} & \cdots & p_{M2s} & f_{12s} & d_{i2s} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 1 & \ln(x_{Ts}/P_{ts}) & p_{1Ts} & \cdots & p_{MTs} & f_{1Ts} & d_{iTs} \end{bmatrix} \tag{28}$$

The second stage of our hierarchical model refers to the hyper-distribution from which the vector of parameters for each store is drawn:

$$vec(\mathrm{E}_s) \mid \mu_s, \phi_s \sim N(vec(\bar{\mathrm{E}}_s), \Delta) \text{ for } s = 1, \cdots, S, \; \Delta^{-1} \sim W(\nu_\Delta, V_\Delta^{-1}) \tag{29}$$

where the expected price elasticity matrix is the restricted one implied by an additive utility model:

$$\bar{\mathrm{E}}_s = \phi_s \, diag(\mu_s) - \phi_s \mu_s (\mu_s \circ w_s)' \tag{30}$$

The remaining parameters are drawn from:

$$\beta_s \sim N(\bar{\beta}_s, \Lambda) \text{ for } s = 1, \cdots, S, \; \Lambda^{-1} \sim W(\nu_\Lambda, V_\Lambda^{-1}) \tag{31}$$

The third stage of our model expresses the prior on the hyper-distribution:

$$\bar{\beta} \sim N(\theta, V_\theta) \tag{32}$$

### *4.1. Estimation Using the Gibbs Sampler*

Our goal is to compute the posterior distribution of the model parameters. The posterior distribution contains all the information from our sample given our distributional assumptions. From the posterior distribution we can compute the means, which are commonly used as point estimates, along with any other measures of the distribution that are of interest. The following data and parameters are supplied by the analyst:

$$X, Y; \theta, V_\theta, \nu_\Lambda, V_\Lambda, \nu_\Delta, V_\Delta, \nu_\Sigma, V_\Sigma, \nu_\phi, V_\Phi, \bar{\phi}, V_{\bar{\phi}} \tag{33}$$

The general procedure for finding the marginal posterior distribution is to compute the joint posterior and then integrate out all parameters except those of interest. In this case the joint distribution of our model can be written as:

$$p(\beta_1, \cdots, \beta_S, E_1, \cdots, E_s, \phi_1, \cdots, \phi_S, \Sigma_1, \cdots, \Sigma_S, \bar{\beta}, \Lambda, \Delta, \bar{\phi}, \lambda_\phi | \text{data, priors})$$
$$\propto \prod_{s=1}^{S} like(\beta_s, E_s, \Sigma_s | \Delta, \Lambda) p(\bar{\beta} | \beta_1, \cdots, \beta_S, \Lambda) p(\Lambda) p(\Delta) p(\theta) p(\bar{\phi}) p(\lambda_\phi) \tag{34}$$

If we wanted to find the marginal posterior distribution of $\theta$ we would need to solve:

$$p(\bar{\beta} | \theta, V_\theta, \nu_\Lambda, V_\Lambda, \nu_\Delta, V_\Delta, \nu_\Sigma, V_\Sigma, \nu_\phi, V_\phi, \bar{\bar{\phi}}, V_{\bar{\phi}})$$
$$= \int p(\beta, \epsilon, \Sigma, \phi, \bar{\beta}, \Lambda, \Delta, \bar{\phi}, \lambda_\phi)\, d\beta\, dE\, d\Sigma\, d\phi\, d\bar{\phi}\, d\lambda_\phi \tag{35}$$

The analytic solution to this integral is not known even with natural conjugate priors. To understand the difficulty in solving this integral, we refer the reader to the simpler case of trying to solve a single stage SUR model (Zellner, 1971, pp. 240–246) for which the analytic solution is not known either. Therefore we will have to rely upon numerical procedures to find the solution. Unfortunately the high dimension of the integral makes it difficult to find a solution using conventional numerical integration techniques.

An alternate method is through the use the Gibbs sampler. The Gibbs sampler requires the solution of the conditional distributions, which can be easily derived due to the hierarchical structure of the model. For a good introduction to the Gibbs sampler see Casella and George (1992). We do not advocate the use of Gibbs sampler based on computational efficiency, instead we advocate its use because of its ease of implementation. The most desirable solution would be an analytical one, but given that this solution does not exist in closed form we satisfy ourselves with a numerical solution.

The Gibbs sampler employed in this paper requires sequentially randomly sampling from each of the conditional distributions. It has been shown by Gelfand and Smith (1990) and Gelfand et al. (1990) that this draws converge to the posterior marginal distributions. The general outline of the procedure is:

(1) Select starting values for the parameters of the marginal posterior distributions. In our practice the least squares estimates of these parameters provide good starting points.
(2) Generate $M_1 + M_2$ sets of random numbers with each set being drawn in the following manner:

$$\beta_s^{(k)} \veebar p(\beta_s \mid E_s^{(k-1)}, \Sigma_s^{(k-1)}, \cdots) \quad \text{for } s = 1, \cdots, S \tag{36}$$

$$E_s^{(k)} \veebar p(E_s \mid \beta_s^{(k)}, \Sigma_s^{(k-1)}, \phi_s^{(k-1)}, \cdots) \quad \text{for } s = 1, \cdots, S \tag{37}$$

$$\phi_s^{(k)} \veebar p(\phi_s \mid \beta_s^{(k)}, E_s^{(k)}, \cdots) \quad \text{for } s = 1, \cdots, S \tag{38}$$

$$\Sigma^{(k)} \veebar p(\Sigma_s \mid \beta_s^{(k)}, E_s^{(k)}, \cdots) \tag{39}$$

$$\bar{\beta}^{(k)} \veebar p(\bar{\beta} \mid \beta_1^{(k)}, \cdots, \beta_S^{(k)}, \Lambda^{(k-1)}, \cdots) \tag{40}$$

$$\bar{\phi}^{(k)} \veebar p(\bar{\phi} \mid \phi_1^{(k)}, \cdots, \phi_S^{(k)}, \lambda_\phi^{(k-1)}, \cdots) \tag{41}$$

$$\Delta^{(k)} \veebar p(\Delta \mid E_1^{(k)}, \cdots, E_S^{(k)}, \beta_1^{(k)}, \cdots, \beta_S^{(k)}, \phi_1^{(k)}, \cdots, \phi_S^{(k)}, \cdots) \tag{42}$$

$$\Lambda^{(k)} \veebar p(\Lambda \mid \beta_1^{(k)}, \cdots, \beta_S^{(k)}, \bar{\beta}^{(k)}, \cdots) \tag{43}$$

$$\lambda_\phi^{(k)} \veebar p(\lambda_\phi \mid \phi_1^{(k)}, \cdots, \phi_S^{(k)}, \bar{\phi}^{(k)}, \cdots) \tag{44}$$

Where the symbol $x \veebar p(x)$ means that the $x$ is a simulated realization or draw from the density $p(x)$ and $k$ denotes the iteration number. The above conditional distributions are understood to also depend upon the prior parameters and the data.

(3) Use the last $M_2$ sets of draws to estimate the posterior marginal distributions.

This means that the problem reduces to solving the conditional distributions of each of the parameters in the posterior distribution. These solutions are readily available due to the hierarchical structure of our model and the affine nature of the normal and Wishart distributions. The solution of the conditional densities are:

(1) Draw the parameter vector in the first-stage in two parts to avoid the nonlinearity induced by the additive separable prior:

(a) Since we know the price elasticities, we can rewrite the model as below:

$$\left[ \ln(q_{its}) - \sum_j \epsilon_{ijs} \ln(p_{jts}) \right] = \alpha_{is} + \mu_{is} \ln(x/P_{ts}) + \theta_{is} f_{its} + \phi_{is} d_{its} + e_{its} \tag{45}$$

The $\beta_s$ vector can be drawn using the usual SUR result.

(b) Since we know the $\beta_s$ vector we can rewrite the model as below:

$$[\ln(q_{its}) - \{\alpha_{is} + \mu_{is} \ln(x/P_{ts}) + \theta_{is} f_{its} + \psi_{is} d_{its}\}] = \sum_j \epsilon_{ijs} \ln(p_{jts}) + e_{its} \tag{46}$$

The $E_s$ matrix can be drawn using the usual multivariate regression result.

(2) Draw the $\phi$ parameter. Notice that conditional upon $E_s$ and $\mu_s$ we have the following univariate regression:

$$\epsilon_{ijs} = (\delta_{ij} - \mu_{js} w_{js})\mu_{is}\,\phi_s + u_{ijs},\; u_s \sim N(0, \Delta) \tag{47}$$

Hence, $\phi_s$ can be drawn using the usual univariate regression result.

(3) $\Sigma_s$ is drawn from an inverted Wishart distribution

$$\Sigma_s^{-1} \sim W(\nu_\Sigma + T_s, (V_\Sigma + \hat{E}_s'\hat{E}_s)^{-1}),\; \hat{E}_s[,i] = y_{is} - X_{is}'\beta_{is} \tag{48}$$

(4) $\bar{\beta}$ is a multivariate regression

$$\bar{\beta} \sim N(H(\Sigma_s \Lambda^{-1}\beta_s + V_\theta^{-1}\theta), H),\; H = (S\Lambda^{-1} + V_\theta^{-1})^{-1} \tag{49}$$

$\bar{\phi}$ is a univariate regression

$$\bar{\phi} \sim N\left(H\left[\frac{\Sigma_s \phi_s}{\lambda_\phi} + \frac{\bar{\phi}}{V_{\bar{\phi}}}\right], H\right),\; H = \left[\frac{S}{\lambda_\phi} + \frac{1}{V_{\bar{\phi}}}\right]^{-1} \qquad (50$$

(5) Since $\Delta$ and $\Lambda$ are independent they can be drawn separately from inverted Wishart distributions:

$$\Lambda^{-1} \sim W(\nu_\Lambda + S,\, V_\Lambda + \Sigma_s(\beta_s - \bar{\beta}_s))(\beta_s - \bar{\beta}_s))') \tag{51}$$

$$\Delta^{-1} \sim W(\nu_\Delta + S,\, V_\Delta + \Sigma_s(vec(\mathrm{E})_s - vec(\bar{\mathrm{E}}_s))(vec(\mathrm{E})_s - vec(\bar{\mathrm{E}}_s))') \tag{52}$$

$$\lambda_\phi^{-1} \sim \mathrm{W}(\nu_\phi + \mathrm{S},\, \mathrm{V}_\phi + \Sigma_s(\phi_s - \bar{\phi})(\phi_s - \bar{\phi})') \tag{53}$$

# 5. APPLICATION TO SCANNER DATA FROM THE REFRIGERATED ORANGE JUICE CATEGORY

We apply our methods to store level scanner data collected from 83 stores from Dominick's Finer Foods chain in Chicago, IL. This data is collected from point-of-sale computers that record quantity and prices of purchased items. Our data is reported at the weekly level for each store. We have 120 weeks of data which is split for the purposes of model validation into a sample for estimation and another for out-of-sample predictive validation. We consider products in the refrigerated orange juice category. Table 1 lists the items under study, average price and market share. The 11 items represent well over 70% of the revenue in this category and cover the range from premium national brands to lower quality store brands. Our expenditure variable ($x$) is calculated from a subset of 26 store categories with over 5,000 UPC's. These categories account for over 25% of total store ACV.

The settings of the priors are chosen to be relatively uninformative relative to the data except for priors on $\Delta$ and $\Lambda$. The prior on $\Delta$ controls the amount

*Table 1.* Listing of the Items Used in the Study, Along with their Average Price and Market Share.

| Item | Abbreviation | Average Price | Market Share |
|---|---|---|---|
| Tropicana Premium 64 | TropP64 | 2.87 | 16.1 |
| Tropicana Premium 96 | TropP96 | 3.12 | 10.7 |
| Florida's Natural 64 | FNat64 | 2.86 | 4.0 |
| Tropicana 64 | Trop64 | 2.27 | 15.8 |
| Minute Maid 64 | MMaid64 | 2.24 | 16.9 |
| Minute Maid 96 | MMaid96 | 2.68 | 5.7 |
| Citrus Hill 64 | CHill64 | 2.32 | 5.1 |
| Tree Fresh 64 | TFresh64 | 2.18 | 2.5 |
| Florida Gold 64 | FGold64 | 2.07 | 2.6 |
| Dominick's 64 | Dom64 | 1.74 | 13.6 |
| Dominick's 128 | Dom128 | 1.83 | 6.9 |

of shrinkage towards the theory, and the prior on $\Lambda$ controls the amount of shrinkage across the stores. A judicious choice of prior settings on these variables can result in estimates that closely proxy the restricted or unrestricted models, or fulfill our desire to fall somewhere in-between these estimates. We evaluate the impact of the prior over a range of settings.

### *5.1. How Good is Our Theory?*

We remind the user that we can actually think of our model as providing two dimensions of restrictions. The first is to employ the restrictions on the price elasticity matrix implied by an additive utility model as described in Section 2. The second is to pool the observations across stores, which would restrict the estimates of one store to be equal to one another. A natural starting point is to perform a classical test to determine whether the restrictions hold exactly. We summarize the number of parameters, in- and out-of-sample MSE, log-likelihood, and Schwarz information criterion (SIC) in Table 2. The restrictions implied by an additive utility model, pooling assumption, or both are all overwhelming rejected ($p < 0.0001$) by standard likelihood ration tests. Upon an initial evaluation it might appear that the neither the theory nor pooling is helpful. An alternative model selection criterion would be to use SIC as an asymptotic argument to justify the choice of models. Using the Schwarz information criterion (SIC) would lead to the choice of restricted store-level models. The out-of-sample predictions imply that the parameter bias induced

***Table 2.*** Comparison of various estimators in terms of number of parameters, log-likelihood, Schwarzinformation criterion (SIC), and in-sample and out-of-sample MSE estimates. The Bayes estimates forseveral prior settings that range between weak and moderate settings of the priors that control shrinkage across stores and towards the restrictions of the additive utility model are provided.

| Approach | Model | Number of parameters | Log-likelihood | SIC | MSE | Predictive MSE |
|---|---|---|---|---|---|---|
| Classical | Unrestricted store | 12,865 | 49560.9 | 40778.2 | 0.170 | 0.379 |
| | Restricted store | 2,905 | 24798.1 | −18005.8 | 0.247 | 0.318 |
| | Unrestricted pooled | 155 | 20850.9 | −40016.3 | 0.314 | 0.385 |
| | Restricted pooled | 35 | 11598.5 | −22816.4 | 0.358 | 0.402 |
| | *Shrinkage across* | | *Stores Shrinkage Towards Additive Utility Restrictions* | | | |
| Bayes | Strong ($k_\Lambda = 0.0001$) | | Strong ($(k_\Delta) = 0.0001$) | | 0.251 | 0.318 |
| | Strong ($k_\Lambda = 0.0001$) | | Weak ($(k_\Delta) = 10000$) | | 0.209 | 0.301 |
| | Weak ($k_\Lambda = 4900$) | | Strong ($(k_\Delta) = 0.0001$) | | 0.182 | 0.337 |
| | Weak ($k_\Lambda = 4900$) | | Weak ($(k_\Delta) = 10000$) | | 0.177 | 0.350 |
| | Moderate ($k_\Lambda = 1$) | | Moderate ($(k_\Delta) = 1$) | | 0.214 | 0.292 |

by the restricted store model is well worth the reduced variance of the parameter estimates.

Table 2 clearly shows that, either in terms of in-sample or out-of-sample fit, pooled models are inferior to more unrestricted models. This is because of the large heterogeneity in this population of stores. It is important to note that the out-of-sample validation results indicate that this is not just the result of over-fitting. The next most important conclusion is that the restrictions of the additive utility theory are useful in improving predictive accuracy. The Bayes model performs the best in out-of-sample predictive validation and offers the flexibility of store level models without the dangers of over-parameterization. In this data set, it appears that the restrictions of additive utility theory hold fairly well. In addition, there are large and detectable store differences so that the Bayes model adapts to something fairly close to the restricted store models. A more formal measure to determine the best model is to compute the posterior odds of our Bayesian models. We follow Newton and Raftery's (1994) technique to compute the posterior odds and we find overwhelming support that a model with a strong prior on the theory and weak prior on commonalities across stores has the highest posterior probability.

### 5.2. *Impact on Price Elasticity Estimates*

In Table 3 we illustrate the similarity and differences in the point estimates for the expenditure and price elasticities of four selected products. First, note the wide variation in the magnitude of the unrestricted store models. A common complaint amongst analysts is that a large number of elasticities may be incorrectly signed and even the magnitudes may be suspect. Notice four of the twelve parameters have unexpected signs and the magnitudes of the own-price elasticities vary widely from –2.2 to –3.7, given the similarity of the products we might expect more similar estimates. In contrast the restricted pooled model which implements pooling across the stores and the exact restrictions as prescribed by an additive utility model eliminates both of these criticisms. However, we have lost all heterogeneity in the estimates across the stores and the theoretical restrictions are rigidly enforced. Both of these assumptions are rejected by standard statistical tests. The estimates from the Bayes model offer a compromise solution in which the only on of the cross-price elasticity is incorrectly signed, and the range of the elasticities are reduced. A judicious choice of our prior can result in estimates that can mimic these restricted estimates, or result in estimates that fall in between these estimates. Again we

***Table 3.*** Expenditure and Cross-price Elasticity Estimates for Selected Products Using Various Estimators.

| | | | Cross-Price Elasticity Matrix Estimates | | | |
|---|---|---|---|---|---|---|
| Description | Product | Expenditure Elasticity Estimate | TropP64 | TropR64 | CHill64 | Dom64 |
| Unrestricted | TropP64 | 1.1 | –2.2 | 0.2 | 0.2 | 0.0 |
| Store Model | Trop64 | 1.7 | –0.4 | –3.7 | 0.6 | –0.2 |
| | CHill64 | 0.9 | 0.2 | 0.1 | –3.1 | –0.2 |
| | Dom64 | 1.2 | 0.8 | 1.5 | –0.4 | –2.3 |
| Restricted | TropP64 | 1.1 | –3.1 | 0.4 | 0.1 | 0.3 |
| Pooled Model | Trop64 | 1.0 | 0.4 | –2.8 | 0.1 | 0.3 |
| | CHill64 | 1.0 | 0.4 | 0.4 | –3.0 | 0.3 |
| | Dom64 | 1.0 | 0.4 | 0.4 | 0.1 | –2.8 |
| Bayes Model | TropP64 | 1.1 | –2.1 | 0.1 | 0.1 | 0.0 |
| | TropR64 | 1.6 | 0.6 | –3.2 | 0.7 | 0.1 |
| | CHill64 | 1.3 | 0.4 | 0.4 | –2.6 | –0.2 |
| | Dom64 | 1.0 | 0.4 | 1.2 | 0.2 | –2.3 |

note that the data provides strong support that a compromise solution is superior both in the form of improved out-of-sample predictions and high posterior odds.

# 6. INCORPORATING INFORMATION ABOUT MARKET STRUCTURE

The analysis of the previous subsection which uses a prior based upon an additive prior may seem overly restrictive. One concern is that a category may have two subcategories that are only weakly related. For example, the laundry detergent category may consist of liquid and powder forms. Substitution within a subcategory may be high, but between these subcategories it may be weak. Additionally, in the orange juice category discussed in the previous section we may have premium orange juice that is fresh versus lower quality juice that is made from concentrate. An additive utility model would not be able to well represent the fact that those in the from concentrate segment would be willing to switch up to the fresh juices but fresh orange juice buyers may not be willing to switch down to orange juice concentrate. These asymmetry effects have long been noted in the marketing literature. To allow increased flexibility we consider combining the strong and weak separability arguments from Section 2 into a single model. If we assume that utility is additive or strongly separable within a category but weakly separable across categories, then utility can take the following structure:

$$utility = \nu(\Sigma_i\, \nu_{1i}(q_{1i}),\ \Sigma_i\, \nu_{2i}(q_{2i}),\ \cdots,\ \Sigma_i\, \nu_{Ci}(q_{Ci})) \tag{54}$$

where $q_{ci}$ is the quantity of the $i$th product in the $c$th category. This will result in the following restrictions on the price elasticities:

$$\epsilon_{ijs} = \begin{cases} \phi_{GG}\,\mu_{is} - \phi_{GG}\,\mu_{it}\mu_{js}w_{js} & \text{if } i \in G \text{ and } j \in H,\ i \neq j \\ \phi_{GG}\,\mu_{is}\mu_{js}w_{js} & \text{if } i \in G \text{ and } j \in G,\ i = j \\ \phi_{GH}\,\mu_{is}w_{js} & \text{if } i \in G \text{ and } j \in H \end{cases} \tag{55}$$

Notice one change from our previous formulation is that we have dropped the store subscript on $\phi$. This change is necessitated by the increased computational requirements of the model. However, we believe this is a sensible restriction, since the $\phi$'s permit differences in market structures and we presume that the market structure in each store is the same.

This structure permits more flexibility in the price elasticity matrix, but still is a fairly parsimonious structure, perhaps overly so for many analysts. If $\phi_{GH} = \phi$ for all $G$ and $H$ then (55) will reduce to the restrictions induced by an additive utility structure in (13). While these structures can be similar, our hope is that by incorporating models that are closer to the true generating process of the data this should result in better approximations and shrinkage patterns. On the other hand, the added flexibility may not be necessary since the model already permits substantial departures from the theory embedded within the prior.

This type of structure has been considered previously in marketing in the context of market structures. Allenby (1989) proposed identifying market structures using a restricted additive utility model – albeit in nested logit form. If we assume that the expenditure elasticities within a market segment are constant, we can derive the same market structure proposed by Allenby. As an illustration suppose there are two submarkets each with 3 brands. The uncompensated elasticity matrix will be:

$$H = \left[\begin{array}{ccc|ccc} \eta_1 & \xi_{aa}w_2 & \xi_{aa}w_3 & \xi_{ab}w_4 & \xi_{ab}w_5 & \xi_{ab}w_6 \\ \xi_{aa}w_1 & \eta_2 & \xi_{aa}w_3 & \xi_{ab}w_4 & \xi_{ab}w_5 & \xi_{ab}w_6 \\ \xi_{aa}w_1 & \xi_{aa}w_2 & \eta_3 & \xi_{ab}w_4 & \xi_{ab}w_5 & \xi_{ab}w_6 \\ \hline \xi_{ba}w_1 & \xi_{ba}w_2 & \xi_{ba}w_3 & \eta_4 & \xi_{bb}w_5 & \xi_{bb}w_6 \\ \xi_{ba}w_1 & \xi_{ba}w_2 & \xi_{ba}w_3 & \xi_{bb}w_4 & \eta_5 & \xi_{bb}w_6 \\ \xi_{ba}w_1 & \xi_{ba}w_2 & \xi_{ba}w_3 & \xi_{bb}w_4 & \xi_{bb}w_5 & \eta_6 \end{array}\right] \tag{56}$$

Where $i = \xi_{ab}w_i - \mu_i\phi$, $\xi_{ab} = -\phi\mu_i\mu_j - \mu_i$, $a$ and $b$ denote the submarket for products $i$ and $j$. The restricted elasticity matrix of (56) is the same as that given in Allenby's (1989) figure 1.

### *6.1. Constructing a Prior on Market Structure*

The first step in constructing a Bayesian model is to develop a prior assessment of the probability for each market structure. For example if we have a category with three products: A, B, and C, then there are five possible market structures: {(A,B,C)}, {(A,B),(C)}, {(A),(B,C)}, {(A,C),(B)}, {(A),(B),(C)}. The most direct solution would be to assume a particular market structure and simply replace the used in (18), which was based upon an additive utility model with the model proposed in (55). In keeping with the theme of this paper we would like to allow some uncertainty about the market structure and allow deviations away from this market structure. Our prior must attach a probability to each of these possible market structures. As the number of products increases there is a combinatorial explosion of possible market structures, perhaps allowing millions of models. Computationally it is not possible to compute the posterior distribution if all these markets must be considered as would happen with a flat prior. Therefore theory or some expertise must be used to guide in identifying likely market structures. If we are totally agnostic then we will not be able to find a solution.

One technique used by Allenby (1989) is to simply enumerate category structures based upon the product attributes, like brand, size, flavor, etc. For example, choose a market structure induced by blocking all brands with the same size together. This technique results in a small number of market structures. Unfortunately, grouping upon individual attributes alone may not be satisfactory. We would like to propose a more flexible approach, that allows grouping based upon multiple attributes, say size and quality. Additionally, we would like to permit some deviations away from this structure. For example, one product that has the same size as those products in one subcategory should be placed with another subcategory due its similarity on quality.

We use a conditional approach to specify our prior that a product belongs to a subcategory. The conditional approach assumes that we know the assignments of the $N-1$ other items in the category and are interested in assigning one additional product. Our problem becomes one of predicting the probability that this unassigned product should be assigned to a new $k+1$ subcategory or one of the existing $k$ subcategories. This conditional specification makes it easy to incorporate it into our Gibbs sampling algorithm. The marginal probabilities of each model can be computed using simulation.

We begin by considering the probability that a new category should be created. We would like this probability to reflect the similarity of the existing groups. If the existing subcategories are quite similar then they will offer low discriminatory value, and we would argue that it is likely that a new category should be opened. On the other hand, if the unassigned product has a high probability of belonging to one subcategory versus the others then this indicates a high discriminatory power of the existing structure, and we would argue that it is less likely that a new category should be created. Additionally, as more categories are created we wish to decrease the probability that a new category should be opened. Another function of this conditional probability is to serve as a penalty function and avoid creating too many subcategories, which would result in an overparameterized model.

Suppose there are $k$ existing subcategories, and the conditional probability that a new product is assigned to subcategory $g$ is $p_g$ and the probability that it is assigned to a new subcategory is $p_{k+1}$. We begin by defining the probability of opening a new subcategory:

$$p_{k+1} = \frac{1}{\gamma} \exp\{\delta o\} \tag{57}$$

where $\gamma$ is a parameter that scales the overall probability and is positive, $\delta$ is a function of the number of categories that currently existing, and is the entropy of the current subcategory classification probabilities. We define entropy as follows:

$$o = -\sum_{g=1}^{k} p_g \log_2(p_g) \tag{58}$$

Entropy is a measure of how much disparity there is in the attributes of the existing categories. If all the probabilities are ($p_g$) are close then entropy is low, as the probabilities diverge entropy increases. Notice that entropy is always positive. Additionally, the scaling function of entropy () is defined as follows:

$$\delta = \delta_1\left(1 + \frac{k}{\delta_2}\right) \tag{59}$$

where $\delta_1$ and $\delta_2$ are scaling parameters and are positive. $\delta_1$ scales the entropy, and $\delta_2$ increases this penalty as the number of existing categories grows.

In constructing the probability that an unassigned item belongs to an existing category we wish to reflect the similarity of the unassigned product with the existing categories. If an attribute of an unassigned product matches those in an existing category then it is likely that this product belongs to this category. We begin by defining the probability that given attribute $i$ the unassigned product belongs to category $g$:

$$p_{gi} = \frac{c_g + \omega}{n_g + \omega} \tag{60}$$

where $c_g$ is the number of products within subcategory $g$ that have the same $i$th attribute and $n_g$ is the total number of products in the subcategory. The role of the parameter $\omega$ is to prevent zero probabilities. If we assume that the $M$ attributes of a product are independent of one another, then the probability that the new product belongs to $g$th group is proportional to:

$$p_g \propto \prod_{i}^{M} p_{gi} \tag{61}$$

It might seem like an independence assumption may be questionable, but since highly correlated attributes can be omitted independence may be a reasonable assumption.

One further extension that we wish to incorporate is to place additional weight on one particular attribute. We modify (61) by raising the probability of the correspond attribute by and raising the other attributes by $1/\tau$. In our problem a priori we are uncertain as to which attribute will be more important, therefore we consider a mixture prior in which an attribute has an equal probability of being the important attribute. In summary our model is:

$$p_g = (1 - p_{k+1}) \frac{\sum_{j=1}^{M} \prod_{i=1}^{M} p_{gi}^{\kappa_{ij}}}{\sum_{l=1}^{k} \sum_{j=1}^{M} \prod_{i=1}^{M} p_{li}^{\kappa_{ij}}}, \quad \kappa_{ij} = \begin{cases} \tau & \text{if } i=j \\ 1/\tau & \text{otherwise} \end{cases} \tag{62}$$

Where $(1 - p_{k+1})$ reflects the probability that a new category is not created or one of the existing categories is selected.

*Example:* Consider the following example to illustrate this prior. Our problem is to determine whether the eleventh product, Minute Maid – Regular – 96 Oz, should be assigned to subcategory A, B, C, or a new subcategory D given the assignments of the other ten products as listed in Table 4. Notice subcategory A appears to be premium products, B is made up of regular products of various brands and sizes, while C is made up of store brands. We set the parameters of this prior as follows: $\omega = 0.001$, $\delta_1 = 0.25$, $\delta_2 = 10$, $\tau = 2$, and $\gamma = 100$. The results indicate that there is a 99% probability that Minute Maid – Regular – 96 Oz. should be assigned to subcategory B, a 1% chance that it should be assigned to a new category, and a negligible probability of

***Table 4.*** The Attributes of Products and a Sample Market Structure and the Product Assignments to Each Subcategory.

| Product | Brand | Quality | Size | Subcategory |
|---|---|---|---|---|
| 1 | Tropicana | Premium | 64 | A |
| 2 | Tropicana | Premium | 96 | |
| 3 | Florida Natural | Premium | 64 | |
| 4 | Minute Maid | Regular | 64 | B |
| 5 | Tropicana | Regular | 64 | |
| 6 | Florida Gold | Regular | 64 | |
| 7 | Citrus Hill | Regular | 64 | |
| 8 | Tree Fresh | Regular | 64 | |
| 9 | Dominicks | Regular | 64 | C |
| 10 | Dominicks | Regular | 128 | |
| 11 | Minute | Regular | 96 | ? |

being assigned to subcategory A or C. This conforms well with our intuition that subcategory B is made up of various national brands. Intuitively the prior strongly predicts that the product belongs to subcategory B because the quality attribute matches perfectly and there one match in the brand category, otherwise category C would have been highly favored. As the parameter is increased to 0.3 the odds of the product being assigned to subcategory C rise significantly to 41%, subcategory B's probability drops to 58%, and the odds of a new category drop to 0.7%. If the parameter is set to zero then unless there is at least one match of the unassigned attribute to the products in the subcategory there is no probability of the unassigned attribute being assigned to that subcategory.

For the 11 products listed in Table 4 there are almost 40 million possible permutations of market structures. However, many of these permutations result in structures that are essentially the same except for the labeling of the subcategories. For example, the market structure {(A,B),(C)} is the same as {(C),(A,B)}. To insure the identifiability of the market structures we only allow those structures in which the lowest product rank as given in Table 1 is less than those of the subcategories that follow it. In the previous example, the permutation {(C),(A,B)} would not be allowed. This identifiability condition results in about 500,000 possible market structures.

We simulate our prior using 100,000 iterations, and list the parameter settings and the number of subcategory structures identified in Table 5.[3] Setting 1 favors those category structures that allow more subcategories and includes the extreme case that all products are assigned to different subcategories. Settings 2 through 6 include most of the usual candidate structures that are blocked by attributes: brand, size, quality, and all products in the same

***Table 5.*** Number of Market Structures Generated by Various Settings of the Prior.

| | Parameter Values | | | Number of |
|---|---|---|---|---|
| Setting | $\gamma$ | $\omega$ | $\delta$ | Market Structures |
| 1 | 0.6 | 0.005 | 2 | 3,120 |
| 2 | 20 | 0.005 | 2 | 16,192 |
| 3 | 20 | 0.005 | 1 | 6,669 |
| 4 | 20 | 0.500 | 5 | 54,756 |
| 5 | 100 | 0.005 | 2 | 6,168 |
| 6 | 100 | 0.001 | 2 | 3,662 |

category. These priors tend to result in subcategories that have more products and result in those subcategories that have similar attributes.

In our subsequent analysis we use the prior that corresponds with setting 6. To acquaint the reader with the types of structures that this model identifies we list the top ten models along with their prior probability in Table 6. A priori the best market structure is the one in which there are two subcategories. One with the store brands (which match on brand and quality) and all others. Many models are slight deviates from one another, in which one product will switch to a different subcategory. These top ten models account for 59% of the probability in the prior. The market structure in which all items are assigned to the same category was ranked 15th.

To better demonstrate the association of the products using our prior we compute the conditional probability that each pair of products will be included in the same subcategory in Table 7.

We readily acknowledge that this prior is only one out of the multitudes that could be constructed. For example, we could imagine using a flat prior, and enumerate all possible models and allow each to have an equal probability of being selected. However, this is computationally infeasible. Another suggestion would be to simply count the number of categories and place a prior that would penalize models based upon the number of parameters. This may result in a prior that yields a penalty function that is that same as the Schwarz information criterion. The benefit of our prior is that it uses brand attribute information and results in model structures that seem plausible without eliminating too many combinations. We conducted many tests of the sensitivity of the prior and found that the information from the likelihood function tends to dominate the information in the prior. Therefore, the basic function of the prior is simply to identify which models are considered, so the censoring property of the prior is its most critical function (i.e. most market structures have zero probability).

### 6.2. *Estimating the Model*

To estimate this model we can create a Gibbs sampler to simulate draws from the marginal posterior distribution. The estimation structure we proposed in Section 4 can be readily adapted to this new structure. We divide the sampler into two components. The first is to simulate the model conditional upon the market structure. The second component is to simulate the market structure conditional upon the parameter estimates. Since this first component is similar to the algorithm described in Section 4, we will not discuss it in depth. The critical difference is that the mean of the hyper-distribution is based on the restrictions given weak separability across the subcategories as described in

***Table 6.*** Top Ten Market Structures and their Probabilities According to the Prior.

| Rank | Subcategory Product Assignments | Prior Probability |
|---|---|---|
| 1 | {TropP64,TropP96,FNat64,Trop64,MMaid64,MMaid96,CHill64,TFrsh64,FGold64} {Dom64,Dom128} | 27% |
| 2 | {TropP64,TropP96,FNat64,Trop64,FGold64} {MMaid64,MMaid96,CHill64,TFrsh64} {Dom64,Dom128} | 11% |
| 3 | {TropP64,TropP96,FNat64,Trop64,MMaid64,MMaid96,FGold64} {CHill64,TFrsh64} {Dom64,Dom128} | 5% |
| 4 | {TropP64,TropP96,FNat64,Trop64,CHill64,FGold64} {MMaid64,MMaid96,TFrsh64} {Dom64,Dom128} | 4% |
| 5 | {TropP64,TropP96,FNat64,Trop64,TFrsh64,FGold64} {MMaid64,MMaid96,CHill64} {Dom64,Dom128} | 3.6% |
| 6 | {TropP64,TropP96,FNat64} {Trop64,MMaid64,MMaid96,CHill64,TFrsh64,FGold64} {Dom64,Dom128} | 2.7% |
| 7 | {TropP64,TropP96,Trop64,MMaid64,MMaid96} {FNat64,CHill64,TFrsh64,FGold64} {Dom64,Dom128} | 2.5% |
| 8 | {TropP64,TropP96,FNat64,Trop64,CHill64,TFrsh64,FGold64} {MMaid64,MMaid96} {Dom64,Dom128} | 1.7% |
| 9 | {TropP64,TropP96,FNat64,Trop64} {MMaid64,MMaid96,CHill64,TFrsh64,FGold64} {Dom64,Dom128} | 1.5% |
| 10 | {TropP64,TropP96,FNat64,Trop64,CHill64,TFrsh64,FGold64} {MMaid64,MMaid96,Dom64,Dom128} | 1.3% |

***Table 7.*** Prior Probability that a Pair of Products will be Assigned to the Same Subcategory.

| | Trop P64 | Trop P96 | Fnat 64 | Trop 64 | MMai d64 | MMai d96 | CHill 64 | TFrsh 64 | FGol d64 | Dom 64 | Dom 128 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TropP64 | 1.00 | 0.95 | 0.86 | 0.85 | 0.47 | 0.49 | 0.45 | 0.45 | 0.74 | 0.05 | 0.05 |
| TropP96 | | 1.00 | 0.82 | 0.82 | 0.47 | 0.50 | 0.43 | 0.43 | 0.70 | 0.06 | 0.06 |
| FNat64 | | | 1.00 | 0.77 | 0.45 | 0.45 | 0.50 | 0.50 | 0.86 | 0.03 | 0.03 |
| Trop64 | | | | 1.00 | 0.55 | 0.56 | 0.55 | 0.54 | 0.78 | 0.03 | 0.03 |
| MMaid64 | | | | | 1.00 | 0.94 | 0.67 | 0.67 | 0.51 | 0.07 | 0.07 |
| MMaid96 | | | | | | 1.00 | 0.64 | 0.64 | 0.50 | 0.08 | 0.08 |
| CHill64 | | | | | | | 1.00 | 0.75 | 0.59 | 0.05 | 0.04 |
| TFrsh64 | | | | | | | | 1.00 | 0.59 | 0.05 | 0.05 |
| FGol64 | | | | | | | | | 1.00 | 0.03 | 0.03 |
| Dom64 | | | | | | | | | | 1.00 | 0.99 |
| Dom128 | | | | | | | | | | | 1.00 |

The product abbreviations are given in Table 1.

(55) and not the restrictions implied by an additive utility model as given in (13). Again our intent is to allow some variation around the restricted model, but induce strong shrinkage towards the theoretical restrictions.

A new component of our algorithm is to simulate the market structure conditional upon the parameter values. The motivation is to randomly select one of the products, compute the probability that it should be remain in the same subcategory, be reassigned to another subcategory, or a new subcategory created. These probabilities form a multinomial distribution from which we simulate a value and reassign the product to the appropriate subcategory and then repeat the first part of the process again which re-estimates all the parameters conditional upon the market structure.

To illustrate this algorithm, suppose that we have four products: A, B, C, and D. At iteration $i$ the market structure is {(A),(B),(C,D)}, and we wish to re-evaluate the assignment of product A. We need to compute the probability of the following models: {(A),(B),(C,D)}, {(A,B),(C,D)}, and {(B),(A,C,D)}. In other words, what is the chance of no change (i.e. product A staying as a separate subcategory) or product A being merged with one of the existing subcategories. The market assignment of product A at iteration $i$ is defined as $M_i$. In our example $M_i$ can take on one of three values: {(A),(B),(C,D)}, {(A,B),(C,D)}; and {(B),(A,C,D)}. Our problem is to compute the posterior probability of $M_i$:

$$p(M_i | \Theta) \propto p(\Theta | M_i) p(M_i) \qquad (63)$$

where $\Theta$ is the set of all parameters in the model to be estimated, $p(\Theta \mid M_i)$ can be computed from the likelihood function given the market structure, and the prior $p(M_i)$ is the prior probability as defined in the previous subsection. Equation (63) will take on a multinomial distribution which can be sampled easily.

We apply this estimation procedure to the same dataset described in Section 5. We evaluate the posterior using three different priors on the amount of shrinkage towards the theory, as captured by $\Delta$, that should be done: strong ($\nu_\Delta = \dim(\Delta) + 3 + 5{*}S$, $V_\Delta = 0.000025$), moderate ($\nu_\Delta = \dim(\Delta) + 3$, $V_\Delta = 0.01$), and weak ($\nu_\Delta = \dim(\Delta) + 3$, $V_\Delta = 0.25$). The purpose is to guage the sensitivity of the posterior to this prior specification. Table 8 provides the out-of-sample forecasting results. The moderate prior performs the best, but all the priors have superior out-of-sample forecasting results compared with the unrestricted models. In comparison to the market structure restricted models the predictive results are similar. However, there are substantial differences in the price elasticity estimates induced by the differences in market structures. Tables 9 through 11 provide the posterior probability of the top ten market structures for the strong, moderate, and weak priors. The most likely market structure in the strong prior contains the 64 ounce cartons, 96 ounce cartons, Tree Fresh, and the store brands. Again it is unlikely that a priori an analyst would have guessed such a structure since this classification cannot be derived from a single attribute. The only question seems to be whether the subcategory with the 96 ounce cartons should be split. As the prior on is weakened the posterior distribution becomes more diffuse and it is more difficult to identify a single market structure. This is quite important since it suggests that if the analyst is unwilling to be aggressive in stating his beliefs that the theory is correct, relying upon the data using a pre-testing method will lead to biased market

***Table 8.*** Comparison of various prior settings for the Bayes model described in Section 6 in terms of in-sampleand out-of-sample MSE estimates. The historical period is different than the previous example,and has in-sample MSE of 0.164 and predictive MSE of 0.395.

| Description | MSE | Predictive MSE |
|---|---|---|
| Strong ($\nu_\Delta = \dim(\Delta) + 3 + 5{*}S$, $V_D = 0.000025$) | 0.268 | 0.330 |
| Moderate ($\nu_\Delta = \dim(\Delta) + 3$, $V_D = 0.01$) | 0.211 | 0.320 |
| Weak ($\nu_\Delta = \dim(\Delta) + 3$, $V_D = 0.25$) | 0.213 | 0.352 |

***Table 9.*** Top Ten Market Structures and their Posterior Probabilities Estimating Using a Strong Prior.

| Rank | Subcategory Product Assignments | Prior Probability |
|---|---|---|
| 1 | {TropP64,FNat64,Trop64,MMaid64,CHill64,FGold64} {TropP96,MMaid96} {TFrsh64} {Dom64,Dom128} | 55% |
| 2 | {TropP64,FNat64,Trop64,MMaid64,CHill64,FGold64} {TropP96} {MMaid96} {TFrsh64} {Dom64,Dom128} | 45% |

***Table 10.*** Top Ten Market Structures and their Posterior Probabilities Estimating Using a Moderate Prior.

| Rank | Subcategory Product Assignments. | Prior Probability |
|---|---|---|
| 1 | {TropP64,Trop64,MMaid96,TFrsh64} {TropP96,Dom64,Dom128} {FNat64,MMaid64,CHill64} {FGold64} | 1.6% |
| 2 | {TropP64,TropP96,FNat64,MMaid64,MMaid96,CHill64,Dom64,Dom128} {Trop64,TFrsh64,FGold64} | 0.6% |
| 3 | {TropP64,FNat64,MMaid64,TFrsh64,FGold64} {TropP96,Trop64,MMaid96,CHill64,Dom64} {Dom128} | 0.6% |
| 4 | {TropP64,TropP96,TFrsh64,FGold64} {FNat64,Trop64,MMaid64,MMaid96,CHill64} {Dom64,Dom128} | 0.6% |
| 5 | {TropP64,TropP96,FNat64,FGold64,Dom64} {Trop64,MMaid64,CHill64,TFrsh64} {MMaid96,Dom128} | 0.6% |
| 6 | {TropP64,TropP96,MMaid64,MMaid96,TFrsh64} {FNat64,FGold64} {Trop64,CHill64} {Dom64,Dom128} | 0.5% |
| 7 | {TropP64,FNat64} {TropP96} {Trop64} {MMaid64,MMaid96} {CHill64,TFrsh64,FGold64} {Dom64,Dom128} | 0.5% |
| 8 | {TropP64,TropP96} {FNat64,FGold64} {Trop64,MMaid64,TFrsh64} {MMaid96} {CHill64} {Dom64,Dom128} | 0.5% |
| 9 | {TropP64,TropP96,FNat64,Trop64,MMaid64,MMaid96,CHill64,Dom64,Dom128} {TFrsh64,FGold64} | 0.5% |
| 10 | {TropP64,MMaid64,MMaid96,CHill64} {TropP96,FNat64} {Trop64,TFrsh64,FGold64} {Dom64,Dom128} | 0.5% |

***Table 11.*** Top Ten Market Structures and their Posterior Probabilities Estimating Using a Weak Prior.

| Rank | Subcategory Product Assignments. | Prior Probability |
|---|---|---|
| 1 | {TropP64,TropP96} {FNat64,FGold64} {Trop64,CHill64} {MMaid64,MMaid96,TFrsh64} {Dom64} {Dom128} | 1.8% |
| 2 | {TropP64} {TropP96,Trop64} {FNat64,CHill64,FGold64,Dom64} {MMaid64,TFrsh64} {MMaid96,Dom128} | 1.1% |
| 3 | {TropP64,TropP96} {FNat64,TFrsh64} {Trop64,MMaid64,CHill64,FGold64} {MMaid96,Dom128} {Dom64} | 1.1% |
| 4 | {TropP64,TropP96,MMaid96,TFrsh64,Dom64,Dom128} {FNat64,Trop64,MMaid64,FGold64} {CHill64} | 0.8% |
| 5 | {TropP64,TropP96,CHill64,FGold64} {FNat64} {Trop64} {MMaid64,MMaid96,TFrsh64} {Dom64,Dom128} | 0.6% |
| 6 | {TropP64} {TropP96,Trop64} {FNat64,CHill64,FGold64,Dom64,Dom128} {MMaid64,TFrsh64} {MMaid96} | 0.6% |
| 7 | {TropP64,TropP96,FNat64,MMaid64,MMaid96,TFrsh64,FGold64} {Trop64,CHill64,Dom64,Dom128} | 0.6% |
| 8 | {TropP64,Dom64,Dom128} {TropP96,MMaid64,MMaid96} {FNat64} {Trop64,CHill64,TFrsh64,FGold64} | 0.5% |
| 9 | {TropP64,Trop64,CHill64} {TropP96,FNat64,MMaid64,MMaid96,FGold64} {TFrsh64,Dom64,Dom128} | 0.5% |
| 10 | {TropP64,TropP96,CHill64} {FNat64,FGold64} {Trop64,MMaid64,MMaid96,TFrsh64} {Dom64,Dom128} | 0.5% |

*Table 12.* Posterior Probability that a Pair of Products will be Assigned to the Same Subcategory Using a Weak Prior.

| | Trop P64 | Trop P96 | Fnat 64 | Trop 64 | MMai d64 | MMai d96 | CHill 64 | TFrsh 64 | FGol d64 | Dom 64 | Dom 128 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TropP64 | 1.00 | 0.69 | 0.49 | 0.42 | 0.21 | 0.25 | 0.24 | 0.24 | 0.32 | 0.20 | 0.19 |
| TropP96 | | 1.00 | 0.41 | 0.36 | 0.20 | 0.30 | 0.15 | 0.19 | 0.25 | 0.25 | 0.26 |
| FNat64 | | | 1.00 | 0.33 | 0.25 | 0.23 | 0.29 | 0.29 | 0.56 | 0.15 | 0.12 |
| Trop64 | | | | 1.00 | 0.30 | 0.26 | 0.43 | 0.38 | 0.41 | 0.14 | 0.12 |
| MMaid64 | | | | | 1.00 | 0.60 | 0.39 | 0.49 | 0.36 | 0.17 | 0.13 |
| MMaid96 | | | | | | 1.00 | 0.27 | 0.38 | 0.24 | 0.25 | 0.25 |
| CHill64 | | | | | | | 1.00 | 0.42 | 0.49 | 0.14 | 0.07 |
| TFrsh64 | | | | | | | | 1.00 | 0.38 | 0.14 | 0.10 |
| FGol64 | | | | | | | | | 1.00 | 0.11 | 0.06 |
| Dom64 | | | | | | | | | | 1.00 | 0.75 |
| Dom128 | | | | | | | | | | | 1.00 |

The Product Abbreviations are given in Table 1.

structure estimates, and hence price elasticity estimates. Regardless of the analyst's beliefs the data has quite a bit of information and can move the prior on the market structures significantly even with a weak prior as the posterior probabilities that pairs of products will be assigned in the same subcategory shows in Table 12. Table 12 can be contrasted with the prior probabilities given in Table 7.

## 7. CONCLUSIONS

We have shown how economic theory can be incorporated into estimators of consumer demand. Our purpose is to represent the notion that a theory is approximately correct. Our estimates can be described as shrinking the unrestricted model estimates without the theory towards the restricted estimates implied by the theory. The amount of shrinkage is adaptive and modified by both an analyst's prior beliefs and the amount of support the data has for the theory. Classical approaches to estimating demand by first pre-testing the adequacy of the theory and then proceeding conditionally upon these estimates will bias the estimates. This will either lead to overconfidence in the estimates when the theory is accepted or underconfidence when the theory is disregarded. An important facet of our shrinkage estimates is that the theory can contribute information even when it is rejected by classical testing procedures, since the theory may be approximately correct. Another benefit of our approach is that it provides the analyst a method for understanding the impact of theoretical assumptions on parameter estimates by varying the degree of confidence in the prior. While we have illustrated our technique using logarithmic demand models, this approach can be applied to any functional form, such as an AIDS or Rotterdam model. Additionally, we hope that this research will encourage applications of Bayes methods to other problems like the estimation of supply and production functions.

## NOTES

1. Our dataset in Sections 5 and 6 consists of 26 categories with over 5,000 UPC's. This dataset accounts for 25% of total store sales. It is this dataset that we use to compute store expenditures. While it would be desirable to have use all products in a store, many products are not scanned, like produce and meat which account for 50% of store sales. Therefore, our expenditure variable can be thought of largely as grocery sales.

2. The $(i, j)$th element of the Slutsky matrix (L) is defined as $l_{ij} = \frac{\partial q_i}{\partial x} q_j + \frac{\partial q_i}{\partial p_j}$.

3. If a market structure does not occur in the simulation we assume that its probability is zero. Effectively, we are truncating our prior.

## ACKNOWLEDGMENTS

The author would like to acknowledge Peter Rossi, Peter Boatwright, and Wei Zhang for their contributions to research upon which this paper is based.

## REFERENCES

Allenby, G. M. (1989). A Unified Approach to Identifying, Estimating and Testing Demand Structures with Aggregate Scanner Data. *Marketing Science*, *8*, 265–280.

Allenby, G. M., & Rossi, P. E. (1993). A Bayesian approach to estimating household parameters. *Journal of Marketing Research*, *30*, 171–182.

Barten, A. P. (1969). Maximum likelihood estimation of a complete system of demand equations. *European Economic Review*, *1*, 7–73.

Barnard, J., McCulloch, R. E., & Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, *10*(4), 1281–1311.

Berndt, E., & Silk, A. (1993). Consistency Requirements and the Specification of Asymmetric Attraction Models of Aggregate Market Share. Working Paper, MIT.

Blattberg, R. C., & George, E. (1991). Shrinkage Estimation of Price and Promotional Elasticities: Seemingly Unrelated Equations. *Journal of the American Statistical Association*, *86*, 304–315.

Blattberg, R. C., & Neslin, S. A. (1990). *Sales promotion: concepts, methods, and strategies*. Prentice Hall: Englewood Cliffs, N.J.

Blattberg, R. C., & Wisniewski, K. (1989). Price-Induced Patterns of Competition. *Marketing Science, 8*, 291–309.

Casella, G., & George, E. (1992). Explaining the Gibbs Sampler. *American Statistician, 46*, 167–174.

Deaton, A. S. (1974). The analysis of consumer demand in the United Kingdom, 1900–1970. *Econometrica, 42*, 341–367.

Deaton, A. S. (1978). Specification and testing in applied demand analysis. *Economic Journal, 88*, 526–536.

Deaton, A. S., & Muellbauer, J. (1983). *Economics and Consumer Behavior.* Cambridge: Cambridge University Press.

Gelfand, A., Hills, S., Racine-Poon, A., & Smith, A. (1990). Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *Journal of the American Statistical Association, 85*, 972–985.

Gelfand, A., & Smith, A. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association, 85*, 398–409.

Green, P. E., & Rao, V. R. (1971). Conjoint Measurement for Quantifying Judgmental Data. *Journal of Marketing Research, 8*(3), 355–363.

Guadagni, P. M., & Little, J. D. C. (1983). A Logit model of brand choice calibrated on scanner Data. *Marketing Science, 2* (Summer), 203–238.

Kamakura, W., & Russell, G. (1989). A Probabilistic Choice Model for Market Segmentation and Elasticity Structure. *Journal of Marketing Research, 26*, 379–390.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy, 10*(2, part 2), 1–92.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes Estimates for Linear Model. *Journal of the Royal Statistical Society: Series B – Statistical Methodology, 34*(1), 1–7.

Montgomery, A. L. (1997). Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data. *Marketing Science, 16*(4), 315–337.

Montgomery, A. L., & Rossi, P. E. (1999). Estimating Price Elasticities with Theory-Based Priors. *Journal of Marketing Research, 36*, 413–423.

Newton, M., & Raftery, A. (1994). Approximate Bayesian Inference with the weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society: Series B, 56*, 3–48.

Rossi, P. E., & Allenby, G. M. (2000). Statistics and Marketing. *Journal of the American Statistical Association, 95*, 635–638.

Srivastava, R. K., Leone, R. P., & Shocker, A. D. (1981). Market Structure Analysis: Hierarchical Clustering of Products Based on Substitution-In-Use. *Journal of Marketing, 45*, 38–48.

Swamy, P. A. V. B. (1970). Efficient Inference in a Random Coefficient Regression Model. *Econometrica, 38*(2), 311–317.

Theil, H. (1976). *Theory and Measurement of Consumer Demand.* Amsterdam: North-Holland.

# A STUDY OF 'SPURIOUS REGRESSION' AND MODEL DISCRIMINATION IN THE GENERALIZED BASS MODEL

Frank M. Bass and Shuba Srinivasan


## ABSTRACT

*Although opinions among time series econometricians vary concerning whether the variables in linear regression models need to be stationary, the majority view is that stationary variables are desirable, if not required, because of the dangers of "spurious regression," (Enders, 1995). Trending, but independent, variables will likely be significantly correlated when combined in a regression analysis. The issue of "spurious regression" and the appropriate manner of including explanatory variables in nonlinear models has not been extensively examined. In this study we examine the issue of model discrimination and "spurious regression" between two nonlinear diffusion models. We use the Generalized Bass Model (GBM) proposed by Bass, Krishnan and Jain (1994) where explanatory variables are included as percentage changes and as logarithms in comparison with the Cox (1972) proportional hazard model with non-stationary variables included as levels. We use simulations to analyze estimation properties and model discrimination issues for the two models.*


**Econometric Models in Marketing, Volume 16, pages 295–315.**

**ISBN: 0-7623-0857-5**

# BACKGROUND AND INTRODUCTION

The Bass model of the diffusion of innovations has become the foremost model for forecasting the sales growth of new products and technologies. The pattern of adoptions that it identifies have been observed with such frequency that the Bass model has been termed an empirical generalization (See Bass (1995) for a discussion of empirical generalizations in marketing and for examples of the Bass model and various extensions and applications.[1]) The original paper by Bass (1969), published in *Management Science*, "A New Product Growth Model for Consumer Durables" has been cited hundreds of times and has spawned a stream of research that has resulted in hundreds of publications. In addition, the Bass model has been applied and utilized for forecasting new product sales with substantial frequency (For a recent example of use of the model for forecasting the sales of a new technology prior to product launch see Bass, Gordon, Ferguson and Githens (2001)).

Robinson and Lakhani (1975) published a paper that modified the Bass model to include price effects on diffusion. Their work led to a series of papers that used a modified form of the Bass model to include demand influencing variables. Among these were papers by Bass (1980), Horsky and Simon (1983), Kalish (1985), Kamakura and Balasubramanian (1988), Jain and Rao (1990), and Horsky (1990).[2] In 1994 Bass, Krishnan and Jain (1994) published a paper showing that earlier models that modified the Bass model to include demand influencing variables do not reduce to the Bass model unless prices and other variables are constant. In the same paper they developed a "higher level" model that reduces to the Bass model under conditions generally found in nature concerning the behavior of prices and other variables and that retains the essential desirable features of the Bass model. This model was termed the Generalized Bass Model (GBM). This model provides an explanation of why the Bass model is an empirical generalization even though it does not include the effects of prices and other demand influencing variables. In empirical studies the Generalized Bass Model provides good fits to the data and the estimated coefficients on the demand influencing variables are statistically significant.

With the introduction of time series variables into nonlinear diffusion models, issues involving the potential for "spurious regression," that have been extensively examined in time series econometrics studies in the context of linear models, arise for nonlinear models as well. Although opinions among time series econometricians vary concerning whether the variables in linear regression models need to be stationary, the majority view is that stationary variables are desirable, if not required, because of the dangers of "spurious regression," (Enders, 1995). Jointly trending, but independent, variables will likely be significantly correlated when combined in a regression analysis. In a well-known simulation study Granger and Newbold (1974) found that when a large number of cases were generated with independent variables and examined by regressions, at the 5% significance level, they were able to reject the null hypothesis of no correlation about 75% of the time. One would expect that the risks of "spurious regression" in nonlinear models would not be unlike those in linear models. A model that includes non-trending explanatory variables would be at less risk than one with trending variables.

In this study we examine the issue of model discrimination and "spurious regression" among two nonlinear diffusion models, the Generalized Bass Model (GBM) and the Cox (1972) proportional hazard model (PHM) as modified by Jain (1992) in the context of the Bass model to include explanatory variables as levels. We shall refer to this model as PHML. These two models have different properties and the explanatory variables enter the models in different ways. GBM has the "carry-through" property in that the impulse response of the model in a single period carries through to future periods, while PHML is a "current effects" model. Input variables in GBM are included as percentage changes and as logarithms of levels, while input variables are included as levels in PHML.

As opposed to growth models that are S-shaped models of a cumulative nature, diffusion models are usually employed as derivatives of cumulative distributions and are bell-shaped. Diffusion models are models of adoption purchases and are usually estimated on data that extend "just past the peak" in order to minimize the contamination of the data with repeat purchases while capturing the essential curvature in the data. A typical example of sales (adoptions) and price data for new products is shown in Fig. 1. The sales curve is generally trending upward with a downturn only in the last period, while prices are declining monotonically.

Figure 2 shows a trending price series and a series of percentage changes in prices that is not trending. When data are generated by a model with trending output but with non-trending inputs, apart from time, and estimated by a "false" model with trending inputs, the general expectation would be that the "false" model would provide a good fit to the data and that the estimates of the coefficients of the explanatory variables would be significant. The trends in input variables and output variables would very likely produce these spurious results. On the other hand, when data are generated by a model with trending output and trending explanatory variables, other than time, and estimated by a model with non-trending inputs one would expect that the fit would not be good and that the estimates of the coefficients would not be significant. The

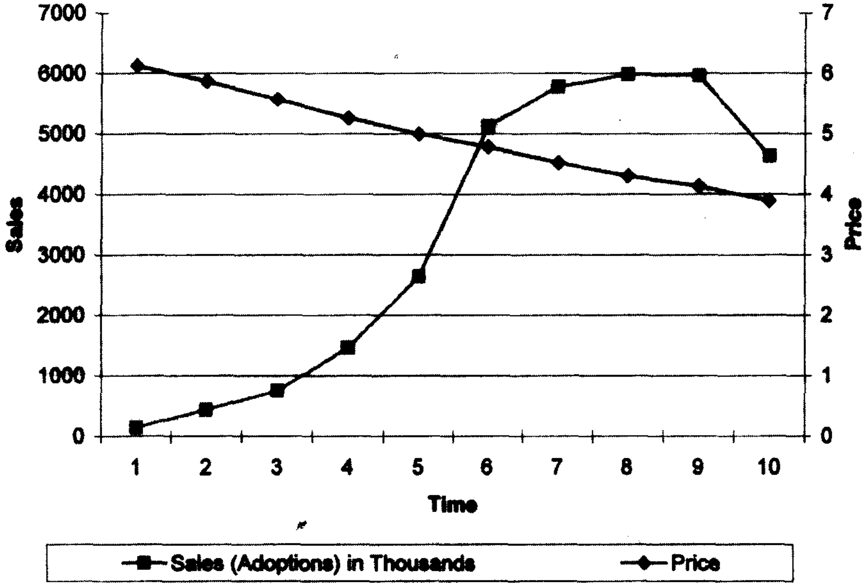


*Fig. 1.* Sales (Adoptions) and Price.

non-trending explanatory variables would be unlikely to be related to the trending output. It would appear, then, that with nonlinear diffusion models the nature of the explanatory variables, non-trending or trending, has a bearing on model discrimination as related to the likelihood of "spurious regression."

When the two models being examined have different properties other than the nature of the explanatory variables the importance of model discrimination is even greater than the case where the only difference between the models is the character of the explanatory variables because the differences in the policy implications of the two models will be influenced by their different properties. The two models that we shall examine here, GBM and PHML, do have very different properties. Each model can provide very good fits to the data, but the nature of response and the policy implications are quite different for the two models.

## THE BASS MODEL

In order to lay the groundwork for the development of the GBM and PHML models we provide here the theoretical development of the Bass model. Unlike

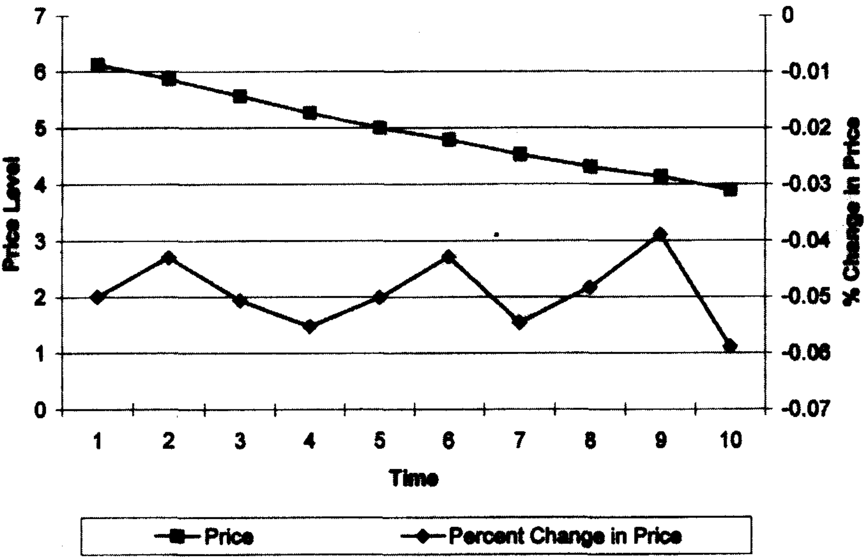


*Fig. 2.* Price Level and Percent Change in Price by Period.

older growth models as customarily applied, such as the Gompertz and Logistic, the purpose of diffusion models is to explain the timing of adoptions (initial purchases) of new products and technologies on the basis of behavioral assumptions that are explicit.

The underlying behavioral assumption of the Bass Model is based largely on the work of Rogers (1962) who specified the following classes of adopters: (1) Innovators; (2) Early Adopters; (3) Early Majority; (4) Late Majority; (5) Laggards. This classification is based upon the timing of adoption by the various groups. A fundamental idea suggested by Rogers is that adopters, other than innovators, are influenced in adoption timing by earlier adopters. The basic idea is that as the number of previous adopters accumulates the pressure on the remaining non-adopters to adopt increases. The increasing pressure of the social system can be explained by "imitation" or "learning." Behaviorally, the "imitation rationale" is similar in certain respects to assumptions employed by Katz and Lazarsfeld (1955) and Mansfield (1961) while the "learning" idea is related to the ideas suggested by Rogers (1962) and Bush and Mosteller (1955). In any case, whether by "imitation" or "learning" or by a combination of the two the pressure to adopt on non-adopters builds up as the number of earlier adopters increases.

Mathematically, the Bass model is related to microscopic models at the individual level such as contagion models that have been applied in epidemiology (Bartlett, 1960) and to personal communication (Taga & Isii, 1959). The fundamental assumption of the Bass model is: *The probability that an initial purchase will be made at time t given that no purchase has yet been made is a linear function of the number of previous adopters.* This assumption may be expressed as a hazard function:

$$f(t)/\{1 - F(t)] = p + qF(t), \tag{1}$$

where f(t) is the likelihood of purchase at t and F(t) is the integral of f(x) from x = 0 to t. If m is defined as the number of initial purchases over the period of interest ("life of the product") and Y(t), the cumulative number of adopters at time t, is defined as mF(t), the right hand side of Eq. (1) may be written as: p + (q/m)Y(t) to demonstrate the linear relationship between the conditional probability of adoption at time t and the number of previous adopters. The parameter p has been defined as the "coefficient of innovation" to represent the influence of innovators in the adoption process (Note that at t = 0, the conditional probability of adoption is p) and the parameter q has been termed the "coefficient of imitation" to reflect the imitation influence of previous adopters on those who have not yet adopted.

Using Eq. (1) f(t) may be written as:

$$f(t) = p + (q - p)F(t) - q[F(t)]^2. \tag{2}$$

Equation (2) is a differential equation and, under the condition that F(0) = 0, may be solved for F(t) to find that:

$$F(t) = (1 - e^{(-(p+q))t})/(1 + (q/p)e^{-(p+q)t}). \tag{3}$$

Equation (3) may be differentiated to find:

$$f(t) = ((p + q)^2/p)e^{-(p+q)t}/(1 + (q/p)e^{-(p+q)t})^2. \tag{4}$$

The sales rate will then be mf(t) or,

$$S(t) = m((p + q)^2/p)e^{-(p+q)t}/(1 + (q/p)e^{-(p+q)t})^2. \tag{5}$$

To find the time, t*, at which the sales rate reaches its peak, Eq. (5) may be differentiated and the result set equal to 0 to find:

$$t^* = 1/(p + q)\ Ln(q/p). \tag{6}$$

Empirically q is usually much larger than p and an interior maximum ordinarily exists and sales curves similar to the one shown in Fig. 1 are typical of new product sales curves. For this reason the Bass model has been termed an empirical generalization.

It is worth noting that if $p=0$ in the hazard function indicated in Eq. (1), the solution for F(t) will be the Logistic function employed by Mansfield (1961) and many others and if $q=0$ the solution for F(t) implies that f(t) is the negative exponential distribution used as employed by Fourt and Woodlock (1960) in the study of the timing of first purchase of grocery products. Hence, the Logistic and negative exponential are special cases of the Bass model.

Although there are many functional forms that could provide good fits to data of the type depicted in Fig. 1, the Bass model has become the model of choice in forecasting sales of new durable products and technologies. The primary reason for this is that the underlying behavioral theory for the Bass model is intuitively appealing and the parameters of the model may be interpreted and understood at an intuitive level. In addition, the empirical estimates of m's, p's, and q's have been cataloged for hundreds of previously introduced products permitting forecasting by the method of "guessing by analogy" under which analysts may try to guess p's and q's by comparing the new product of interest to an analogous product introduced earlier.

## THE GENERALIZED BASS MODEL

The Generalized Bass Model (Bass, Krishnan & Jain, 1994) was developed for the purpose of explaining why the Bass model is an empirical generalization despite the absence of price and other demand influencing variables in the model. Underlying each of the sales curves that have been observed is a sequence of prices and other influences. These variables vary considerably from case to case, but the observed shape of the sales pattern is the same over the large number of products for which sales data have been observed. In order to provide the desired explanation it is necessary that the generalized model reduce to the Bass model under conditions on the explanatory variables that are commonly observed in nature. At the same time it is desirable that the generalized model retain the essential properties of "imitation" or "learning" of the Bass model as indicated in Eq. (1).

The hazard function for GBM is:

$$f(t)/[1-F(t)]=x(t)[p+qF(t)], \tag{7}$$

where x(t) is a dynamic function of dynamic control variables. For purposes of exposition we consider price and advertising although other variables could be easily accommodated. Notice that Eq. (7) is observationally equivalent to Eq. (2) when x(t) is constant.

The next problem is to find a specification for x(t) in such a way that x(t) is approximately constant under circumstances that are ordinarily found in nature.

It is often observed that prices of new technologies fall exponentially with time. The concept of declining costs and prices as expressed in the accumulated experience of a firm has been extensively developed and applied by the Boston Consulting Group (1968). The marginal cost function utilized in experience curve theory is the same as the functional form utilized by Arrow (1962) in his famous paper "The Economic Implications of Learning by Doing." Among others, Bass (1980) found empirical support for exponential price decline for several new product categories. Similar exponential behavior for other decision variables is often the pattern during the growth phase of the diffusion process. Exponential behavior, of course, implies a constant percentage change in a variable. The specification of x(t) is such that exact exponential behavior of the decision variables will result in x(t) being exactly a constant and approximately exponential behavior will result in x(t) being approximately a constant. The choice of the functional form for x(t) is based upon the desire to capture the properties of the Bass model and, at the same time, to have a closed form solution to the resulting differential equation. For these reasons GBM proposed:

$$x(t) = 1 + [\Delta Pr(t)/Pr(t-1)]\beta_1 + [\Delta ADV(t)/ADV(t-1)]\beta_2, \tag{8}$$

where Pr(t) is price at time t and ADV(t) is advertising at time t. Summing x(t) with respect to t yields X(t) so that:

$$X(t) = \sum_{\tau=0}^{t} x(\tau). \tag{9}$$

If time is treated as continuous, then Eq. (8) may be written as:

$$x(t) = 1 + [dPr(t)/dt)/Pr(t)]\beta_1 + [dADV(t)/ADV(t)]\beta_2. \tag{10}$$

Similarly, if time is continuous Eq. (10) may be integrated to find $X(t)^3$:

$$X(t) = t + [Ln(Pr(t)/Pr(0))]\beta_1 + [Ln(ADV(t)/LnADV(0))]\beta_2. \tag{11}$$

The adoption rate (Sales at time t) for the Generalized Bass Model is:

$$S(t) = m((p+q)^2/p)x(t)e^{-(p+q)X(t)}/(1+(q/p)e^{-(p+q)X(t)})^2, \tag{12}$$

where x(t) and X(t) are given by Eqs (10) and (11) respectively.

Equation (12) retains the essential properties of Eq. (5) and is observationally equivalent to Eq. (5) if x(t) is constant. Equation (12) is also capable of explaining deviations of the empirical sales data from the smooth curve produced by the Bass Model on the basis of variations in the decision variables. In an empirical analysis Bass, Krishnan, and Jain (1994) showed that the Generalized Bass Model produced better fits than the Bass Model and that parameter estimates of the β's were statistically significant.

An essential property of both the Bass Model and the Generalized Bass Model stems from the behavioral rationale that underlies these models. The hazard function of these models indicate that the conditional probability of adoption among those who have not yet adopted is influenced by the number of previous adopters. Thus the greater the number of adopters today the greater will be the influence on the remaining potential adopters to adopt at each future time period. A price reduction today will have an influence on adoption today but will also have an impact on adoption in the future because of the way diffusion is captured by the model.

For the Generalized Bass Model the hazard rate may be expressed as:

$$\lambda(t, x) = x(t)(p + q)/(1 + (q/p)e^{-(p+q)X(t)}), \tag{13}$$

where x(t) and X(t) are defined by Eqs (8) and (9). An impulse in a decision variable at time t will have a permanent impact on $X(t+\tau)$ because $X(t+\tau)$ depends on all prior values of x through the summation operation on x that results in X. The hazard rate for the Generalized Bass Model, then, will be permanently shifted by an impulse in a decision variable at time t. Because the hazard rate at $t+\tau$ is influenced by an impulse at t the effect of the impulse will also carry through to the adoption rate at $t+\tau$. The impulse response for the Generalized Bass Model therefore reflects the "carry-through" property of the model.

## THE PROPORTIONAL HAZARD MODEL

The proportional hazard modeling framework was introduced by Cox (1972). This framework was employed by Jain and Vilcassim (1991) in a marketing context in a study of household purchase timing for frequently purchased products. The proportional hazard model has also been studied in a diffusion context in comparison with the Generalized Bass Model by Bass, Jain and Krishnan (2000).

In its most basic form the proportional hazard function may be written as:

$$\lambda(t, Z) = \lambda_0(t)\phi[Z(t)], \tag{14}$$

where $\lambda_0(t)$ is the baseline hazard function and $\phi[Z(t)]$ is a function of explanatory variables. Customarily $\phi[Z(t)]$ is assumed to be $EXP(Z_1(t)\beta_1 + Z_2(t)\beta_2)$, where $Z_1(t)$ is the price level at t, Pr(t), and $Z_2(t)$ is the advertising level at t, ADV(t). If the Bass Model hazard function (Eq. (1)) is used in the baseline function and if the solution to the Bass Model differential equation, $F(t) = (1 - e^{-(p+q)t})/(1 + (q/p)e^{-(p+q)t})$, is substituted into F(t) in the right hand side of Eq. (1) then $\lambda_0(t)$ may be written as:

$$\lambda_0(t) = (p+q)/(1+(q/p)e^{-(p+q)t}). \tag{15}$$

Writing Z(t) as: $Pr(t)\beta_1 + ADV(t)\beta_2$ and Y(t) = cumulative sales at time t = mF(t) it is then possible to write the sales rate for the proportional hazard model as:

$$S(t) = \lambda_0(t)(m - Y(t))e^{Z(t)}. \tag{16}$$

Equation (16) has parameters m, p, q, $\beta_1$, and $\beta_2$ and corresponds to Eq. (12) for the sales rate of the Generalized Bass Model. Unlike x(t) in Eq. (12) the decision variables in Z(t) are included as levels in Eq. (16) and if these variables are trending Eq. (16) appears to be at greater risk of "spurious regression" than Eq. (12). It is also worth noting that Eq. (16), unlike Eq. (12), will reduce to the Bass model only if the decision variables are unchanging.

For the proportional hazard model the hazard rate is: $\lambda(t, Z) = (p+q)/(1+(q/p)e^{-(p+q)t})e^{Z(t)}$, where $Z(t) = Pr(t)\beta_1 + ADV(t)\beta_2$. The function $Z(t+\tau)$ is not affected by an impulse in a decision variable at time t. For this model, then, the hazard rate at $t+\tau$ is not affected by an impulse in Z at t. The proportional hazard model, therefore, does not have the "carry-through" property.

In simulating and estimating the proportional hazard model we shall use the discrete version of the model that is suggested as appropriate for discrete data as suggested by Kalbfleisch and Prentice (1980) as appropriate for grouped data. The model development for this as applied to diffusion is discussed in Bass, Jain and Krishnan (2000). The equation is:

$$S(t) = (m - Y_{t-1})J(t-1, t), \tag{17}$$

where $Y_{t-1}$ is the observed or simulated cumulative sales at t − 1 and where

$$\begin{aligned} J(t-1, t) = 1 - [&Exp(-(p+q))\{(1+(q/p) \\ &\times Exp(-(p+q)(t-1)))/(1+(q/p)Exp(-(p+q)t))\}]^{\wedge(Z(t))}. \end{aligned}$$

## SIMULATION ANALYSIS OVERVIEW

Although the Generalized Bass Model and the proportional hazard model are both based on the Bass Model they differ in important respects: percentage changes vs. levels in the decision variables and the presence and absence of the "carry-through" property. In our simulation analysis we shall simulate data for each of the models and explore the ability to discriminate between the models based on estimates of the simulated data. In our simulation study of the two models we shall refer to the Generalized Bass Model as GBM and the proportional hazard model in levels of the decision variables as PHML. Nine parameters were manipulated for this study:

(i) p, the coefficient of innovation
(ii) q, the coefficient of imitation
(iii) m, the market potential
(iv) $\beta_1$, the price coefficient
(v) $\beta_2$, the advertising coefficient
(vi) the error variance for the price series
(vii) the error variance for the advertising series
(viii) the error variance for the model
(ix) the length of the simulated data series – just past peak versus well beyond peak.

We want to ensure that the parameters selected for the simulation study bear closeness to values that occur in empirical studies. Hence, we choose for starting values of the parameters p, q, m, $\beta_1$ and $\beta_2$ estimates obtained by Bass, Jain and Krishnan (2000) for room air-conditioners, color TVs and clothes dryers. The model parameters are estimated by the non-linear least squares procedure (Bass et al., 2000). We assess the performance of the models in terms of goodness-of-fit (R2), parameter estimates, and standard errors. We generate multiple sets of data for different parameter values with the underlying structures given by GBM and PHML. In order to ensure robustness of our results, we perform 30 replications for each cell and report the mean values of the estimates and the standard error of the mean.

The relative performance of the models is influenced by the model error and hence we manipulate the model error variance to obtain two treatment levels – low and high. Since not using the same error variance introduces a potential confound, we assume that the error distribution is uniform for *both* the high error variance and the low error variance conditions. Further, since the length of the simulated data series may affect the outcome, we control for this by manipulating the length to obtain two treatment levels – just past peak sales and well beyond peak sales. We will also generate sets of simulations in which the decision variables, price and advertising, behave in different ways. In the first set of simulations price and advertising are generated as exponential functions with error. To simulate the price series in levels, we use the following functional form:

$$y_{P,t} = a_0 \exp(at) + \varepsilon_{P,t} \tag{18}$$

where $0 < a < 1$ and $\varepsilon_{P,t}$ is the error term. To simulate the advertising series in levels, we use the following functional form:

$$y_{A,t} = b_0[1 - \exp(bt)] + \varepsilon_{A,t} \tag{19}$$

where $0 < b < 1$ and $\varepsilon_{A,t}$ is the error term. Here too, since not using the same error variance introduces a potential confound, we assume that the error distribution is uniform for *both* the price series and advertising series. While the price series is decreasing exponentially at the rate of $-(1-a)$ with stationary error, the advertising series is increasing exponentially to an asymptote of $b_0$ with a stationary error. Thus, the percentage change in price will be a constant plus stationary error. Similarly, the percentage change in advertising will be a constant plus stationary error.

It then follows that the demand-influencing variable for GBM: $x(t) = 1 + [dPr(t)/dt)/Pr(t)]\beta_1 + [dADV(t)/ADV(t)]\beta_2$ (shown in Eq. 10) will be mean-stationary. In contrast, the proportional hazard model, depends solely on $Z(t) = Pr(t)\beta_1 + ADV(t)\beta_2$ which is not mean-stationary as evident from the data-generating Eqs (18) and (19). Therefore, because the GBM has a stationary demand-influencing variable and the proportional hazard model does not, it would appear that the proportional hazard model would be more susceptible to "spurious regression" than GBM. We empirically investigate these time-series properties of x(t) and Z(t) by testing for the presence or absence of unit roots using the Augmented Dickey Fuller tests (Dickey & Fuller, 1981).

In the second set of simulations the price series is generated with a structural break, a single impulse in which price is sharply reduced in one period. This set of simulations permits us to examine the estimation properties of the two models under a condition in which the adoption curve for GBM is shifted while the position of the PHML curve is not changed.

## SIMULATION ANALYSIS-SMOOTHLY TRENDING DECISION VARIABLES

The results of the simulation study when the decision variables are smoothly trending are summarized in Tables 1–3. Table 1 contains the estimates for a set of simulations with a low model error variance for both models while Table 2 shows the estimates with higher model error variances; the length of the simulated data series is just past peak in both Tables 1 and 2. Table 3 shows the estimates with the length of the data series being longer and well beyond the peak sales.

A summary of the comparisons of statistical results matched with the model generating the data is provided below.

When the data are generated by GBM:

- Estimates of the GBM model provide good fits and the parameter recovery is very good with high levels of significance, as expected, when the model error variance is small. When the model error variance is high the model also

***Table 1.*** **Results from Simulations with Low Model Error Variance* (standard errors are in parentheses).**

| True Model | Estimation | p | q | m | B1 | B2 | R-sq |
|---|---|---|---|---|---|---|---|
| Clothes dryers*** | | | | | | | |
| GBM | Parameters | 0.01 | 0.30 | 17000 | –0.74 | 0.20 | |
| | GBM | 0.01 (0.00) | 0.29 (0.00) | 18491 (303) | –0.70 (0.06) | 0.19 (0.07) | 0.99 |
| | PHML | 0.03 (0.07) | 0.36 (0.01) | 15385 (215) | –0.35 (0.04) | –0.04 (0.02) | 0.97 |
| PHML | Parameters | 0.01 | 0.30 | 17000 | –0.74 | 0.20 | |
| | PHML | 0.01 (0.00) | 0.27 (0.02) | 17626 (99) | –0.67 (0.05) | 0.20 (0.01) | 0.99 |
| | GBM | ** | | | | | |
| Room air-conditioners*** | | | | | | | |
| GBM | Parameters | 0.01 | 0.33 | 19500 | –1.4 | 0.5 | |
| | GBM | 0.01 (0.00) | 0.32 (0.003) | 20564 (108) | –1.35 (0.03) | 0.47 (0.06) | 0.99 |
| | PHML | 0.00 (0.00) | 0.26 (0.02) | 17011 (178) | –0.63 (0.01) | 0.08 (0.01) | 0.97 |
| PHML | Parameters | 0.01 | 0.33 | 19500 | –1.4 | 0.5 | |
| | PHML | 0.01 (0.00) | 0.34 (0.09) | 20043 (44) | –1.31 (0.11) | 0.47 (0.04) | 0.99 |
| | GBM | ** | | | | | |
| Color TVs*** | | | | | | | |
| GBM | Parameters | 0.004 | 0.60 | 40000 | –1.5 | 0.5 | |
| | GBM | 0.003 (0.00) | 0.60 (0.00) | 40665 (38) | –1.49 (0.01) | 0.46 (0.02) | 0.99 |
| | PHML | 0.001 (0.00) | 0.46 (0.16) | 41120 (770) | –0.54 (0.06) | 0.08 (0.09) | 0.97 |
| PHML | Parameters | 0.004 | 0.60 | 40000 | –1.5 | 0.5 | |
| | PHML | 0.003 (0.00) | 0.72 (0.01) | 40537 (42) | –1.40 (0.02) | 0.45 (0.00) | 0.99 |
| | GBM | ** | | | | | |

* The distribution of the model error is uniform with standard deviation of 25 for GBM sales and PHML sales with observations just past peak.
** denotes non-convergence of the estimation routine.
*** All the x(t) series are mean-stationary at the 5% levels while the Z(t) series are not using the ADF unit root tests.

***Table 2.*** Results from Simulations with High Model Error Variance* (standard errors in parentheses).

| True Model | Estimation | p | q | m | B1 | B2 | R-sq |
|---|---|---|---|---|---|---|---|
| Clothes dryers*** | | | | | | | |
| GBM | Parameters | 0.01 | 0.30 | 17000 | –0.74 | 0.20 | |
| | GBM | 0.01 (0.00) | 0.26 (0.00) | 24637 (665) | –0.56 (0.15) | 0.27 (0.19) | 0.97 |
| | PHML | 0.01 (0.00) | 0.23 (0.03) | 21220 (1098) | –0.26 (0.08) | 0.02 (0.02) | 0.94 |
| PHML | Parameters | | | | | | |
| | PHML | 0.02 (0.00) | 0.25 (0.04) | 21308 (531) | –0.65 (0.08) | 0.22 (0.12) | 0.97 |
| | GBM | ** | | | | | |
| Room air-conditioners*** | | | | | | | |
| GBM | Parameters | 0.01 | 0.33 | 19500 | –1.4 | 0.5 | |
| | GBM | 0.01 (0.00) | 0.32 (0.08) | 26497 (2637) | –1.04 (0.22) | 0.31 (0.20) | 0.94 |
| | PHML | 0.00 (0.00) | 0.10 (0.18) | 23876 (7349) | –0.45 (0.22) | 0.12 (0.09) | 0.95 |
| PHML | Parameters | | | | | | |
| | PHML | 0.00 (0.00) | 0.30 (0.00) | 21432 (111) | –0.84 (0.18) | 0. 29 (0.15) | 0.94 |
| | GBM | ** | | | | | |
| Color TVs*** | | | | | | | |
| GBM | Parameters | 0.004 | 0.60 | 40000 | –1.5 | 0.5 | |
| | GBM | 0.01 (0.00) | 0.60 (0.00) | 41050 (390) | –1.15 (0.02) | 0.31 (0.20) | 0.91 |
| | PHML | 0.03 (0.00) | 0.36 (0.07) | 38236 (5718) | –0.30 (0.03) | 0.10 (0.08) | 0.81 |
| PHML | Parameters | 0.004 | 0.60 | 40000 | –1.5 | 0.5 | |
| | PHML | 0.01 (0.00) | 0.58 (0.04) | 42870 (1581) | –0.59 (0.22) | 0.31 (0.19) | 0.92 |
| | GBM | ** | | | | | |

* The distribution of the model error is uniform with standard deviation of 95 for GBM sales and PHML sales with observations just past peak.
** denotes non-convergence of the estimation routine.
*** All the x(t) series are mean-stationary at the 5% levels while the Z(t) series are not using the ADF unit root tests.

***Table 3.*** Results from Simulations with Observations Beyond Peak* (standard errors are in parentheses).

| True Model | Estimation | p | q | m | B1 | B2 | R-sq |
|---|---|---|---|---|---|---|---|
| Clothes dryers*** | | | | | | | |
| GBM | Parameters | 0.01 | 0.30 | 17000 | −0.74 | 0.20 | |
| | GBM | 0.01 (0.00) | 0.30 (0.00) | 17980 (52) | −0.74 (0.05) | 0.10 (0.05) | 0.99 |
| | PHML | 0.02 (0.07) | 0.28 (0.01) | 17817 (145) | −0.11 (0.02) | −0.00 (0.02) | 0.97 |
| PHML | Parameters | | | | | | |
| | PHML | 0.01 (0.00) | 0.31 (0.002) | 17888 (163) | −0.76 (0.06) | 0.19 (0.05) | 0.99 |
| | GBM | ** | | | | | |
| Room air-conditioners*** | | | | | | | |
| GBM | Parameters | 0.01 | 0.33 | 19500 | −1.4 | 0.5 | |
| | GBM | 0.01 (0.00) | 0.32 (0.00) | 20426 (56) | −1.35 (0.05) | 0.40 (0.06) | 0.99 |
| | PHML | 0.04 (0.02) | 0.31 (0.00) | 20316 (87) | −0.21 (0.02) | 0.03 (0.00) | 0.97 |
| PHML | Parameters | | | | | | |
| | PHML | 0.04 (0.02) | 0.31 (0.00) | 20316 (87) | −1.31 (0.11) | 0.52 (0.02) | 0.99 |
| | GBM | ** | | | | | |
| Color TVs*** | | | | | | | |
| GBM | Parameters | 0.004 | 0.60 | 40000 | −1.5 | 0.5 | |
| | GBM | 0.004 (0.00) | 0.60 (0.00) | 40649 (36) | −1.49 (0.03) | 0.486 (0.03) | 0.99 |
| | PHML | 0.11 (0.14) | −0.54 (0.22) | 40909 (117) | −0.74 (0.25) | 0.50 (0.07) | 0.97 |
| PHML | Parameters | 0.004 | 0.60 | 40000 | −1.5 | 0.5 | |
| | PHML | 0.00 (0.00) | 0.64 (0.13) | 40621 (39) | −1.94 (0.32) | 0.54 (0.07) | 0.99 |
| | GBM | ** | | | | | |

* The distribution of the model error is uniform with standard deviation of 25 for GBM sales and PHML sales.
** denotes non-convergence of the estimation routine.
*** All the x(t) series are mean-stationary at the 5% levels while the Z(t) series are not using the ADF unit root tests.

produces good fits and good parameter estimates with low standard errors except for $\beta_2$, the advertising coefficient.
- Estimates of the PHML model provide good fits and all parameter estimates are significant except when the model error variance is high in which case the estimate of the advertising coefficient has larger standard errors.

When the data are generated by PHML:

- Estimates of the PHML model provide good fits and the parameter recovery is good with high levels of significance when the model error variance is small, as expected. However, when the model error variance is large, although the fits are fairly good, the estimates have higher error variances.
- When the GBM model is estimated, the estimation routine does not converge.

These results are valid *even* when the length of the data series is longer and well beyond the peak sales. Further, the results of the ADF unit root tests confirm that the demand-influencing variables for the GBM given by $x(t) = 1 + [dPr(t)/dt)/Pr(t)]\beta_1 + [dADV(t)/ADV(t)]\beta_2$ are mean-stationary while this is not the case for the demand-influencing variables in the PHML (see footnote to Tables 1–3). When the data are generated by smoothly trending decision variables PHML is highly susceptible to "spurious regression." When the data are generated by GBM the PHML model fits the data well and with significant parameter estimates. This result stems from the high correlation of the trending levels of the decision variables with the model trend and is roughly analogous to the "spurious regression" found in linear regression models with trending variables. In contrast to the "spurious regression" of PHML, GBM estimates of data generated by PHML do not converge. The input variable in the GBM, x(t), is mean stationary and uncorrelated with the model trend. As a result GBM rejects the GBM model as being the "true" model when the data are generated by PHML.

## SIMULATION ANALYSIS-IMPULSE IN PRICE SERIES

In order to study the estimation properties of the two models when there is a break in the smoothly trending decision variables we have simulated data in which there is a sharp decline in the price in one time period. The estimates for data simulated for the two models with a single impulse at period 4 ($t = 4$) are shown in Tables 4 and 5.

A summary of the comparisons of statistical results matched with the model generating the data is provided below.

***Table 4.*** Results from Simulations with Impulse in Period t = 4 for the GBM*
(standard errors are in parentheses).

| True Model | Estimation | p | q | m | B1 | B2 | R-sq |
|---|---|---|---|---|---|---|---|
| GBM (with impulse) | Parameters | 0.01 | 0.30 | 17000 | –0.74 | 0.20 | |
| | GBM | 0.01 (0.00) | 0.29 (0.00) | 17456 (665) | –0.69 (0.15) | 0.23 (0.12) | 0.99 |
| | PHML | ** | | | | | |
| GBM (with impulse) | Parameters | 0.01 | 0.33 | 19500 | –1.4 | 0.5 | |
| | GBM | 0.01 (0.00) | 0.33 (0.06) | 22434 | –1.02 (0.12) | 0.21 (0.19) | 0.99 |
| | PHML | ** | | | | | |
| GBM (with impulse) | Parameters | 0.004 | 0.60 | 40000 | –1.5 | 0.5 | |
| | GBM | 0.00 (0.00) | 0.60 (0.00) | 41258 (364) | –1.15 (0.01) | 0.34 (0.19) | 0.99 |
| | PHML | ** | | | | | |

* The distribution of the model error is uniform with standard deviation of 25 for GBM sales with observations just past peak.
** denotes non-convergence of the estimation routine.

***Table 5.*** Results from Simulations with Impulse in Period t = 4 for the PHML*
(standard errors are in parentheses).

| True Model | Estimation | p | q | m | B1 | B2 | R-sq |
|---|---|---|---|---|---|---|---|
| PHML (with impulse) | Parameters | 0.01 | 0.30 | 17000 | –0.74 | 0.20 | |
| | GBM | ** | | | | | |
| | PHML | 0.01 (0.00) | 0.34 (0.01) | 17536 (108) | –0.73 (0.04) | 0.20 (0.01) | 0.99 |
| PHML (with impulse) | Parameters | 0.004 | 0.60 | 40000 | –4.8 | 0.2 | |
| | GBM | ** | | | | | |
| | PHML | ** | | | | | |
| PHML (with impulse) | Parameters | 0.02 | 0.38 | 20000 | –4.8 | 0.2 | |
| | GBM | ** | | | | | |
| | PHML | ** | | | | | |

* The distribution of the model error is uniform with standard deviation of 25 for PHML sales with observations just past peak.
** denotes non-convergence of the estimation routine.

When the data are generated by GBM:

- Estimates of the GBM model provide very good fits to the data and the parameter values are closely recovered with small standard errors.
- The estimation routine for the PHML model does not converge.

When the data are generated by PHML:

- The estimation routine for the GBM model does not converge.
- When the response coefficient for price, $\beta_1$, is small the PHML model produces a good fit to the data and fairly good parameter estimates, but when $\beta_1$ is large the estimation routine for PHML does not converge.

When the data are generated with an impulse in one period by the GBM model and estimated by GBM the fit is good and parameter recovery is good with small standard errors. This result is entirely expected because GBM easily accommodates choppy decision variables. On the other hand, when PHML is estimated with the same data the estimation routine does not converge. This is in sharp contrast to the case when decision variables are smoothly trending where PHML has good fits and parameter estimates with small standard errors. When data are generated by the PHML model and there is an impulse in price, convergence is not obtained by the estimation routine. Putting all this together it appears that the PHML model is not able to handle non-smooth data when there is a strong response to impulses in decision variables. GBM, on the other hand easily accommodates these conditions.

## SUMMARY AND CONCLUSIONS

The issue of whether trending variables in linear regression models should be included as levels or differences is a subject of debate, but the majority view is that because of the dangers of "spurious regression" only stationary variables should be used. In this study we have examined the issue of stationary explanatory variables versus trending explanatory variables in the context of nonlinear models. We have examined the estimation properties of two nonlinear models that have similarities but which are structurally different in that one has a stationary demand-influencing variable and the other does not. One of these models, the Generalized Bass Model (GBM), has a "carry through" property where the effects of a demand-influencing variable in any time period are felt in that time period but also have effects in all future time periods. The other model, the proportional hazard model (PHML), on the other hand, is a "current effects" model in that the effect of a demand-influencing variable in any period is felt only in the current period.

Our simulation analysis indicates that the PHML model is at serious risk of "spurious regression" when the explanatory variables are smoothly trending. When the data are generated by GBM and fitted by PHML the fits are good and the parameter estimates are significant. In contrast, when data are generated by PHML and fitted by GBM the fits are poor. When the data are generated with an impulse in a demand-influencing variable in one period GBM will easily accommodate the impulse and will produce good fits and estimates when GBM is the true model. PHML, however, is unable to recover the parameters when there is an impulse and a strong response to the impulse.

The approach we have taken in this study suggests the possibility of examining the model discrimination properties and "spurious regression" susceptibilities of nonlinear models more generally.

# NOTES

1. For an older literature review of diffusion theory papers see Mahajan, Muller, and Bass (1990) and for a discussion of managerial applications of the Bass model and its extensions see Mahajan, Muller, and Bass (1995). For more recent developments see Mahajan, Muller, and Wind (Eds) (2000).
2. For a review and evaluation of models that use a modified form of the Bass model to include demand influencing variables see Bass, Jain, and Krishnan (2000).
3. When price and advertising are reasonably smooth functions of time x(t) may be approximated by equation (10) and X(t) by Eq. (11). However, when there are discontinuities such as an impulse at some time, the discrete functions indicated in equations (8) and (9) are appropriate.

# REFERENCES


Arrow, K. (1962). The Economic Implications of Learning by Doing. *Review of Economic Studies, 29*, 155–173.

Bartlett, M. (1960). *Stochastic Populations Models in Ecology and Epidemiology.* London: Methuen.

Bass, F. (1980). The Relationship Between Diffusion Rates, Experience Curves, and Demand Elasticities for Consumer Durable Technological Innovations. *The Journal of Business, 53* (pt. 3), S51–S67.

Bass, F., Gordon, K., Ferguson, T., & Githens., M. (2001). DIRECTV: Forecasting Diffusion of a New Technology Prior to Product Launch. *INTERFACES, 31* (pt. 2), S82–S93.

Bass, F. (1969). A New Product Growth Model For Consumer Durables. *Management Science, 15*, 215–227.

Bass, F. (1995). Empirical Generalizations and Marketing Science: A Personal View. *Marketing Science, 14* (pt. 2), G6–G19.

Bass, F., Krishnan, T., & Jain, D. (1994). Why the Bass Model Fits Without Decision Variables. *Marketing Science, 13*, 203–223.

Bass, F., Jain, D., & Krishnan, T. (2000). Modeling the Marketing-Mix Influence. In: V. Mahajan, E. Muller & Y. Wind (Eds), *New-Product Diffusion Models* (pp. 99–122). London: Kluwer, Boston/Dordrech.

Boston Consulting Group (1968). *Perspectives in Experience*. Boston.

Bush, R., & Mosteller, R. (1955). *Stochastic Models for Learning*. New York: Wiley.

Cox, D. (1972). Regression Models with Life Tables. *Journal of the Royal Statistical Society, 13*(34), 187–200.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood Ratio Statistics for Autogressive Time Series with a Unit Root. *Econometrics, 49*, 1057–1072.

Enders, W. (1985). *Applied Econometric Time Series*. New York: Wiley.

Fourt, L., & Woodlock, J. (1960). Early Prediction of Market Success for New Grocery Products, *Journal of Marketing, 26*, 31–38.

Granger, C., & Newbold, P. (1974). Spurious Regressions in Econometrics. *Journal of Econometrics, 2*, 111–120.

Horsky, D., & Simon, L. (1983). Advertising and Diffusion of New Products. *Marketing Science, 9*, 342–385.

Horsky, D. (1990). A Diffusion Model Incorporating Product Benefits. *Price, Income, and Information, 9*, 342–385.

Jain, D., & Rao, R. (1990). Effect of Price on the Demand for Durables: Modeling, Estimation, and Findings. *Journal of Business Economics and Statistics, 8*, 163–170.

Jain, D., & Vilcassim, N. (1991). Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach. *Marketing Science, 10*, 1–23.

Jain, D. (1992). Marketing Mix Effects on the Diffusion of Innovations. Working Paper, Northwestern University.

Kalbfleisch, J., & Prentice, R. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Kalish, S. (1985). A New-Product Adoption Model with Price, Advertising, and Uncertainty. *Management Science, 31*, 1569–1585.

Kamakura, W., & Balasubramanian, S. (1988). Long-Term View of the Diffusion of Durables: A Study of the Role of Price and Adoption Influence Processes via Tests of Nested Models. *International Journal of Research in Marketing, 5*, 1–13.

Katz, E., & Lazarsfeld, F. (1955). *Personal Influence*. New York: The Free Press.

Mahajan, V., Muller, E., & Bass, F. (1990). New-Product Diffusion Models in Marketing: A Review and Directions for Future Research. *Journal of Marketing, 27*, 979–989.

Mahajan, V., Muller, E., & Bass, F. (1995). Diffusion of New Products: Empirical Generalizations and Managerial Uses. *Marketing Science, 14* (pt. 2), G979–989.

Mansfield, E. (1961). Technological Change and the Rate of Imitation. *Econometrica, 29*, 741–766.

Robinson, B., & Lakhani, C. (1975). Dynamic Price Models for New Product Planning. *Management Science, 21*, 1113–1122.

Rogers, E. (1962). *Diffusion of Innovations*. New York: The Free Press.

Taga, Y., & Isii, K. (1959). On a Stochastic Model Concerning the Pattern of Communication-Diffusion of News in a Social Group. *Annals of the Institute of Statistical Mathematics, 11*, 25–43.

# USING STOCHASTIC FRONTIER ANALYSIS FOR PERFORMANCE MEASUREMENT AND BENCHMARKING

Leonard J. Parsons

## ABSTRACT

*Historically standard regression has been used to assess performance in marketing, especially of salespeople and retail outlets. A model of performance is estimated using ordinary least squares, the residuals are computed, and the decision-making units, say store managers, ranked in the order of the residuals. The problem is that the regression line approach characterizes average performance. The focus should be on best performance. Frontier analysis, especially stochastic frontier analysis (SFA), is a way to benchmark such best performance. Deterministic frontier analysis is also discussed in passing. The distinction between conventional ordinary least squares analysis and frontier analysis is especially marked when heteroscedasticity is present. Most of the focus of benchmarking has been on identifying the best performing units. The real insight, though, is from explaining the benchmark gap. Stochastic frontier analysis can, and should, model both phenomena simultaneously.*

**Econometric Models in Marketing, Volume 16, pages 317–350.**

**ISBN: 0-7623-0857-5**

# I. INTRODUCTION

Evaluating performance is an important task for managers. For example, consider the evaluation of salespeople (Jackson, Schlacter & Wolfe, 1995). The evaluation process signals to salespeople those aspects of their jobs that are important and how they are doing on these dimensions. The performance appraisal process entails identifying those factors on which the salespeople will be evaluated, developing performance standards to express the level of performance desired on each factor, monitoring actual performance, and reviewing performance with each salesperson. Performance factors include both results (output) and effort (input) variables.

An inadequate performance appraisal system can cause considerable costs to the firm (Vandenbosch & Weinberg, 1993, pp. 680–682). For example, the system impacts salesforce motivation, morale, and turnover. Yet directly comparing salespeople is difficult to do in practice because territories are not balanced perfectly in terms of environmental characteristics such as workload and potential. To address this, territory response functions have been constructed in which territory sales are a function of salesperson, company, and territory characteristics. These functions are typically estimated using multiple regressions; e.g. Parsons and Vanden Abeele (1981). Significant variables in an estimated response function identify factors that influence salesperson performance. The regression residuals provide an environmental-adjusted measure of differential performance among sales representatives (Vandenbosch & Weinberg, 1993, p. 681).

Regression methods have focused attention on what an average salesperson might achieve in a territory but management may be more interested in what a top-performing salesperson could achieve. This is not an intrinsic problem with regression analysis, but rather a statement about how it is used. Regression models generally require that the functional form of the model and error distribution be correctly specified. If these assumptions are correct, then estimates about a top-performing salesperson could be generated directly from the distribution around the regression line. For example, one can predict what is the chance that a salesperson would exceed sales of a given level. This percentile can be predicted directly, instead of the usual point estimate. If the regression model is correctly specified, one can predict both the "best" the salesperson can do and what the average salesperson should do. Such a regression analysis can be misleading, however, if there is misspecification. The thrust of this chapter is that the error structure in the classical regression model is likely misspecified; in particular, by ignoring technical inefficiency.

This first section begins by looking at the historical approach to performance evaluation: ordinary least squares (OLS) residual analysis. Then it looks at recent developments using hierarchical Bayesian methods. Finally it concludes with coverage of benchmarking. The next section focuses on the concept of economic efficiency. The main section examines the methods for estimating the frontiers indicative of best performance and discusses related empirical marketing studies. The penultimate section addresses additional related methodological advances that should provide further insights into performance measurement in marketing. The last section provides concluding remarks.

## *A. OLS Residual Analysis*

The performance of outlets in a retail chain, members of a sales force, and the like can be assessed using response models. This approach allows the separation of factors under the control of the local store (e.g. employee hours) or sales representative (e.g. calls) from factors not under their control (e.g. national advertising or market potential). The performance models have typically been estimated using conventional regression analysis. Let's look at two prototypical examples of this approach to performance evaluation – one concerning members of a sales force and the other concerning outlets in a retail chain – that will help frame the issues addressed in this chapter.

### *Salesperson Performance Example*

Twenty-five sales territories of the national sales force of a large manufacturer were analyzed in an early performance study (Cravens, Woodruff & Stamper, 1972). These territories ranged from smaller, more congested territories to large territories. Each territory was assigned to a single salesperson. Total unit sales were the measure of performance. Based on a conceptual framework for sales territory performance, six factors were posited to impact territory performance. These were market potential, territory workload, salesperson experience, salesperson motivation and effort, company experience, and company effort. Industry sales were used as a measure of market potential. There were two measures of workload: workload per account and number of assigned accounts. Average workload per account took into consideration annual purchases by accounts and the concentration of accounts. Salesperson experience was the length of time employed by the company. Salesperson motivation and effort was captured by having sales managers provide ratings of the salespeople under their supervision on eight dimensions of performance, which were distilled into one aggregate rating. Company experience also had two measures, average market share and market share change. These were based on the prior four years. Company effort was advertising expenditures in the territory. Thus, there were a total of eight explanatory variables.

A linear relationship was posited and estimated by stepwise regression.[1] The R-square was 0.72. Five variables – number of accounts, industry sales, market share change, workload per account, and performance of the salesperson – were statistically significant. Predicted sales provided a benchmark and were used to analyze the performance level achieved in each territory. The difference between actual sales and predicted sales (the residual) was compared. In addition, the ratio of actual sales to sales predicted by the model was computed. Management indicated that the benchmarks yielded by multiple regression analysis conformed more closely to its beliefs about high- and low-performance territories than did assigned quotas. Indeed, a comparison of benchmarks to quota achievement indicated no significant relationship. The researchers observed "The methodology generates objective standards of performance in terms of *what exists* in the organization rather than *what should exist*" (Cravens, Woodruff & Stamper, 1972, p. 36).[2]

*Retail Store Performance Example*

The performance of IBM Product Centers in the mid-1980s serves as a second illustration. These retail stores served small businesses and individual consumers. Initially set up to sell electronic typewriters, they were undergoing a transition to selling computers at the time of the study.

The original data contained information on 80 Product Centers. However, information on new product centers was considered atypical and analyses were done only on 72 established Product Centers (Parsons, 1992). The database contained information on over 50 variables or combinations of variables. One complication was employee sales. Since making a sale to a commercial customer requires a different level of effort than making a sale to one of IBM's own employees and since IBM's interest was in the commercial customer, employee sales were deducted from revenue. Correspondingly, traffic and total man months of representatives at each Product Center were deflated by the proportion of employee sales at that Product Center.

The primary regression model assumed that revenue was a function of local market conditions and Product Center efforts. Local market conditions included information on potential and competitive activity. Product Center efforts included total man months of representatives. Another key factor explaining sales were traffic, which itself was a function of local market conditions.

Adjusted traffic per month was explained by potential in the standard trade area, potential in the SMSA (Standard Metropolitan Statistical Area) but not in the standard trade area, density of the potential in proximity to the Product Center, whether or not the store was located in the East, and whether or not the store was an urban multiple. Thus, traffic was explained not only by raw potential, but how that potential was distributed across a store's market. The R-square was 0.59.

Adjusted annual revenue was explained by adjusted traffic, total man-months of sales representatives, density of potential in standard trade area, potential in SMSA but not in the standard trade area and *Future Computing* penetration. Thus, potential appears not only directly, but also indirectly because it drives traffic. The R-square was 0.79.

The performance of individual Product Centers were measured on both an absolute and a relative basis. Regression models indicated which factors were salient in predicting performance. The parameter estimates gave the weights assigned to each factor. These weights represented *average* performance. Product Centers that have much lower actual sales than predicted were candidates for the most attention. Here is where the largest gains could be achieved through better efforts. The results implied that simply bringing Product Centers performing below expectation up to an average level of performance could yield significant increases in revenue.

What might explain the differences in performance among Product Centers? The variances of actual revenue from predicted revenue were compared on the basis of the quality of sales effort. Potential sources of success were average sales skill in a Product Center, average skill on DOS, average skill on PCs, average skill on typewriters, percent employees trained in computers, percent employees trained in service, and percent employees trained in typewriters. Further analysis revealed that *training* employees in computers should improve selling effectiveness.

While IBM found the performance study useful,[3] one manager raised some questions "Why should the focus be on average performance? Why should Product Centers above the regression line be left "off the hook?" If the error distribution of the regression model is correctly specified, being over or under the regression line would not be the right question for managers to be asking, as this is uncontrollable random error. The only thing that managers should be focused on, if they believe the regression model, is how can they change the independent variables to control where they fall on the regression line. This is the market response modeler's view of the world. Nonetheless, the questions raised by the IBM manager prompted a look into how to find the line that represents the *frontier* of best performance.[4]

### *B. Hierarchical Bayesian Analysis*

A hierarchical Bayesian (HB) approach has been proposed to address territory and representative heterogeneity in sales force evaluation (Cain, 2001). The HB methodology allows population means to be expressed as functions of explanatory variables; in particular, individual representative effect parameters can be expressed as a function of representative characteristics, such as years of experience. This HB approach could also be used to evaluate store branches and store managers because of the similarity of the problem structure to that which we have already seen.

One advantage of the HB approach over ordinary least squares is that exceptional performance over multiple products can be measured simultaneously. The performance measure in regression models is usually aggregate sales as we have just seen. Moreover, by generating posterior samples, the HB approach cannot only provide an expected rank for a sales representative but also the probability that one sales rep is ranked higher than another. This could also be accomplished by OLS or maximum likelihood. Analytically deriving orders and ranks is generally difficult, if not impossible. However, simulation approaches such as Gibbs sampling that is used to implement many HB approaches are quite conducive to these calculations.

The HB approach was applied to one of a large pharmaceutical company's sales force. The sales force sold 5 well-established prescription products and was comprised of 488 sales representatives. Territories were uniquely assigned to individual sales representatives and covered between 2 and 341 zip codes. There were 6985 zip codes for which there was complete data. The dependent variable was the dollar-weighted number of prescriptions sold for each product in each zip code. The HB approach effectively shared information across sales people and products. Only factors that were beyond the control of the manager were included in the model.[5] The proposed multi-product model had an improved fit over an OLS model and generated quite different parameters and rankings of the sales reps than did the OLS model. Ways to improve upon training decisions were also investigated. The focus was on individual level parameters for each territory characteristic. For example, was a sales rep good at converting market potential for some products and not for others? The researcher noted, "A key theoretical difference between various methods is the implied referent or referent group. In OLS and the proposed hierarchical Bayesian method, sales people are compared to the 'average or typical' representative, while DEA compares them to 'best' performer."

### *C. Benchmarking*

Benchmarking is the search for the best practices that will lead to superior performance of an organization (Camp, 1989, 1995). These best practices may be found within the organization (for example, among outlets of a retail chain), in the industry, or outside the industry. Our focus is on best practices within the organization – *internal benchmarking*. Benchmarking these best practices has two aspects: the practices themselves, i.e. the methods that are used, and the metrics, i.e. the quantified effect of these practices. Benchmark metrics permit identification of the benchmark gap: how much, where, and when; while benchmark practices suggest how to close the gap: improved knowledge, improved practices, and improved processes.

How do you identify which practice should be designated the best? One way is to look for situations in which the benchmark metric is significantly better, and then examine the practices that caused the exceptional achievement. This means that an operation under investigation first must be quantified. An operation can be broken down into inputs, a work process with repeatable practices, and outputs. Then a metric can be constructed based on the analysis of outputs in relation to inputs. The benchmark metric will be based on the best performance among *comparable* operations (for example, among sales people). Further [perhaps qualitative, although here quantitative] investigation of the practices underlying the best performing operation (salesperson) should yield insight into the sources of success, which may come from process practices, management practices, or operational structure. These (sales) practices could then be spread throughout the firm. Thus, training is often key to implementing the findings from a benchmarking study. In addition, insights might be obtained into desirable characteristics of salespeople or retail managers. This information could be used to improve the selection process.

Before proceeding further, let's take a brief look at the language of economic efficiency. The following discussion of economic efficiency and methods draws heavily on Greene (1997) and, especially, Kumbhakar and Lovell (2000). Also see Coelli, Rao and Battese (1998).

## II. ECONOMIC EFFICIENCY

Production is a process for transforming a set of inputs **X** into a set of outputs **Y**. For the most part, we will focus on a single output, $Y$. The transformation process takes place in the context of a body of knowledge called the production function. An idealized production function is given by

$$Y \leq f(\mathbf{X}), \tag{1}$$

where $f(\mathbf{X})$ is the production frontier. Economic efficiency of production has two main components: technical efficiency and allocative efficiency. *Technical efficiency* (TE) focuses on the ability to obtain the maximum output from a given set of resources. Different methods of input application may have different effects on output. An *output-based* Debreu-Farrell style measure of technical efficiency[6] is

$$TE(Y, \mathbf{X}) = \frac{Y}{f(\mathbf{X})}. \tag{2}$$

This ratio of actual output to the optimal value specified by the production function is called total factor productivity (TFP) in the case of a single output. Thus, for a similar bundle of inputs and technology, an economic agent, often called a decision making unit (DMU), that uses the best practice method achieves the maximum possible output, which will be superior to a DMU that does not do the same (Kalirajan, 1990).[7] In our case, the DMUs are typically salespeople or managers of retail outlets in a chain.

Empirical measurement of technical efficiency starts with a model such as

$$Y_i = f(\mathbf{X}_i; \boldsymbol{\beta}) TE_i, \tag{3}$$

where the technology parameters of the production frontier to be estimated are $\boldsymbol{\beta}$ (Greene, 1997, p. 87ff; Kumbhakar & Lovell, 2000, p. 64ff). The technical efficiency of DMU $i$,

$$TE_i = \frac{Y_i}{f(\mathbf{X}_i; \boldsymbol{\beta})}, \tag{4}$$

lies between zero and one. Technical efficiency equals one when observed output achieves its maximum feasible value; otherwise, it provides a measure of the extent to which observed output falls short of maximum feasible output. This entire shortfall is attributed to technical inefficiency since the production frontier is *deterministic*. The actual shortfall indicates the magnitude of the opportunity for improvement. Management will need to judge whether the potential gains indicated are worth pursuing.

Output, however, is likely to be affected as by random shocks not under control of the DMU. A DMU-specific random shock $\exp(V_i)$ can be added to the deterministic part (or kernel) $f(\mathbf{X}_i; \boldsymbol{\beta})$ common to all DMUs to specify a *stochastic* production frontier. Now (3) can be rewritten as

$$Y_i = f(\mathbf{X}_i; \boldsymbol{\beta}) \exp(V_i) TE_i. \tag{5}$$

The observed output for a specific DMU will be greater than the deterministic portion of the frontier if the associated random error is greater than the corresponding inefficiency effect.

*Allocative efficiency* focuses on the ability to maximize profits by equating marginal revenue product with the marginal costs of inputs. Given the technology and price information, DMUs are expected to make adjustments in their levels of application of inputs to achieve allocative efficiency. Allocative efficiency is not addressed here; see Greene (1999, pp. 120–137).

# III. METHODOLOGY

*Econometric frontier analysis* is proposed as one approach to measuring technical efficiency and establishing benchmarks. Another approach would be mathematical programming, in particular, data envelopment analysis (DEA).[8] We first review the traditional response function that incorporates only equation error, next look at deterministic frontier models that incorporate only inefficiencies, and finally discuss stochastic frontier models that incorporate both inefficiencies and random error.

## A. Conventional Response Function

In marketing, the focus is on the response function:

$$Y = r(\mathbf{X}; \boldsymbol{\beta}) \exp(V), \tag{6}$$

where $V$ represents random error and is a given convenient representation. The most common market response model used is the multiplicative model, known as the constant elasticity model (Hanssens, Parsons & Schultz, 2001, pp. 101–102). Economists call it the Cobb-Douglas production function. A [natural] log-log transformation of the structural model creates an estimation model that is linear in its parameters,

$$\ln Y = \beta_0 + \sum_{k=1}^{K} \beta_k \ln X_k + V, \tag{7}$$

where $V$ iid $\mathrm{N}(0, \sigma_V^2)$. This model is known as the double-log model or linear-in-the-logs model. As already noted, the problem for performance measurement is that this traditional regression model is inadequate because it ignores the truncated errors representing technical inefficiency. The model when estimated by OLS represents *average* performance; that is, it fits a line through the middle of the data. One would, however, like to benchmark best performance; that is, fit a line marking the frontier of the data.

### *B. Deterministic Frontiers*

Assuming that one will usually formulate the production function as linear in the natural logs of the variables, the empirical version of (3) is

$$\ln Y_i = \ln f(\mathbf{X}_i\!: \boldsymbol{\beta}) + \ln TE_i. \tag{8}$$

Thus, to assess best performance, a *deterministic frontier* production function with *one-sided errors representing inefficiency* could be formulated:

$$\ln Y_i = \ln f(\mathbf{X}_i, \boldsymbol{\beta}) - U_i, \tag{9}$$

where $U \geq 0$ and is a measure of technical inefficiency. In this model, efficiency can be found as

$$TE_i = \exp(-U_i). \tag{10}$$

Finding the deterministic frontier was first addressed, not econometrically, but by mathematical programming and applied to the Cobb-Douglas production function (Aigner & Chu, 1968). A linear programming (absolute deviations) formulation is

$$\min_{\boldsymbol{\beta}} \sum_{i}^{N} |\ln Y_i - \ln f(\mathbf{X}_i, \boldsymbol{\beta})|$$

subject to

$$\ln f(\mathbf{X}_i, \hat{\boldsymbol{\beta}}) = \ln \hat{Y}_i \geq \ln Y_i$$

A quadratic programming version has also been posited. The main problem with the mathematical programming approach is extreme sensitivity to outliers.

In an attempt to make a more robust model, a probabilistic reformulation,

$$\Pr(\hat{Y}_i \geq Y_i) > \mathrm{P}, \tag{12}$$

in which one discards the top $(1 - \mathrm{P})*100\%$ of efficient observations was proposed (Timmer, 1971). The problem is that this is an ad hoc procedure.

The mathematical programming approaches are not based on a statistical model and so no statistical inferences can be made. To address statistical issues, several tacks have been taken. One is to develop new maximum likelihood estimators. Another is to adjust the ordinary least squares estimators.

Suppose the constant elasticity model (7) is reformulated as containing only technical inefficiency:

$$\ln Y = \beta_0 + \sum_{k=1}^{K} \beta_k \ln X_k + \ln TE_i = \beta_0 + \sum_{k=1}^{K} \beta_k \ln X_k - U, \tag{13}$$

where $U \geq 0$. This formulation violates the usual regression assumption that the mean of the errors is zero. Maximum likelihood estimation (MLE) involves specifying a one-sided distribution for the inefficiencies. Usually $U$ is $u \sim$ half normal but could be $U \sim$ general truncated normal, $U \sim$ exponential, $U \sim$ gamma, or other single-tailed distribution.[9]

The exponential distribution is

$$f(U) = \frac{1}{\sigma} \exp\left( -\frac{U}{\sigma} \right), \tag{14}$$

where $U \geq 0$. MLE yields the same estimator as Aigner and Chu's linear programming. The truncated half-normal distribution is

$$f(U) = \frac{2}{\sqrt{2\pi\sigma}} \exp\left( -\frac{U^2}{2\sigma^2} \right), \tag{15}$$

where $U \geq 0$. MLE yields the same estimator as Aigner and Chu's quadratic programming (Schmidt, 1976). This linkage, however, does not endow the production functions calculated with mathematical programming with a statistical foundation. The problem is that regularity conditions for MLE are violated. Imprecisely speaking, the range of the random variable in question should be independent of parameters and this independence is invoked to prove the general result that ML estimators are consistent and asymptotically efficient. Here, however, the range is not independent of the parameters.

The precise violation of the Aigner-Chu/Schmidt formulation of the above estimators is that the interchange of integration and differentiation of the gradient needed to obtain the asymptotic distributions for these likelihoods by familiar methods is not permissible (Greene, 1980). The non-zero root of the log-likelihoods is a consequence. Not all distributions suffer from this problem, most notably the gamma distribution:

$$f(U) = \frac{\gamma^{\alpha}}{\Gamma(\alpha)} U^{\alpha-1} \exp(-\gamma U), \tag{16}$$

where $U > 0$, $\gamma > 0$, and $\alpha > 2$. While the density is defined for $\alpha > 0$, $\alpha > 2$ is required for a well-behaved log-likelihood function for the frontier model. There is no known mathematical programming problem corresponding to a ML problem with inefficiency distributed as gamma. Moreover, the requirement that all sample residuals must be strictly positive for the resultant estimator to be computable can create practical difficulties for iterative search methods (Greene, 1997, p. 94).

market in which it is operating. Early empirical papers include Kalirajan (1981) and Pitt and Lee (1981).

A two-stage approach is thus adopted. The first stage involves the specification and estimation of the stochastic frontier production function and the estimation of technical inefficiency effects, assuming that these inefficiency effects are identically distributed. The second stage involves the specification of a regression model for predicted technical inefficiency effects. Since technical efficiency is bounded by zero and one, OLS is not the appropriate technique although it has been used in practice. One must either transform the dependent variable or use a limited dependent variable technique, such as tobit.[15] For example, a semi-log relationship was used by Kalirajan (1990):

$$\ln(\exp(-U)) = \alpha_0 + \sum_{k=1}^{K} \alpha_k Z_k + V \tag{25}$$

and a logistic model was used Mester (1997):

$$\hat{E}(U_i | V_i + U_i) = \frac{\exp(\mathbf{Z}'\boldsymbol{\alpha})}{1 + \exp(\mathbf{Z}'\boldsymbol{\alpha}0}, \tag{26}$$

in modeling cost efficiency. While the two-step approach seems reasonable, it contradicts the assumption of identically distributed inefficiency effects on the stochastic frontier.

The solution is to estimate the parameters of the stochastic frontier and inefficiency models *simultaneously*. The relevant technical literature includes Kumbhakar, Ghosh and McGucklin (1991), Reifschneider and Stevenson (1991), Huang and Liu (1994), and Battese and Coelli (1995). For example, the Battesse-Coelli model is

$$\ln Q = \beta_0 + \sum_{k=1}^{K} \beta_k \ln X_k + V - U$$

where $V$ iid $N(0, \sigma_V^2)$, $U$ iid $N(\mu, \sigma_U^2)$ with truncations at zero and

$$\mu = \delta_0 + \sum_{k=1}^{K} \delta_k Z_k.$$

Implementation uses the reparameterizations

$$\sigma^2 = \sigma_V^2 + \sigma_U^2 \text{ and } \gamma = \frac{\sigma_U^2}{\sigma_V^2 + \sigma_U^2}.$$

Additional information and references on incorporating exogenous influences on efficiency can be found in Kumbhaker and Lovell (2000, pp. 261–278).

Parsons (1991), in introducing frontier analysis to marketing, illustrated this approach using sales force data from the inorganic products division of a major German manufacturer [described in Böcker (1986)]. Its industrial cleaning products division was organized with a sales manager, 6 territory managers, 29 representatives, and 6 sales engineers. Because of peculiar situations in some regions, data from the accounting department about the number of sales calls were only available for 19 sales representatives operating under "normal" conditions. The territory managers, moreover, were in charge of all contacts with large customers. Consequently, sales to the giant buyers had to be eliminated to properly appraise the performance of field representatives. The constant elasticity model relating sales to sales effort, measured by calls, and account potential, measured by the number of inhabitants, was

$$\begin{aligned} \ln[\text{Sales}] = {} & 1.45 + 0.59\ \ln[\text{Calls}] + 0.74\ \ln[\text{Inhabitants}] \qquad \bar{R}^2 = 0.57 \\ (s) \quad & (0.23) \qquad\qquad (0.17) \\ (p) \quad & (0.02) \qquad\qquad (0.005) \end{aligned} \tag{21}$$

The estimated sales call elasticity is similar to that generally found for salesforce-sensitive firms of around 0.5 (Hanssens, Parsons & Schultz, 2001, p. 348). The estimated OLS residuals are given in the fourth column of Table 1. Shifting the intercept to make the residuals non-negative yields the COLS solution, shown in the sixth column.

Since only the intercept has shifted, the estimated production frontier is parallel to the OLS regression. However, there is no reason to expect that the structure of "best practices" production technology is the same as the structure of "central tendency" production technology (Färe, Grosskopf & Lovell, 1994, p. 3). But, while the COLS technique is easy to implement, it imposes this very restrictive property.

### *C. Stochastic Frontiers*

The central problem is that the deterministic formation does not allow for the usual random errors or "noise" encountered with any model. In particular, a single unusual observation, that is, outlier, can have serious effects on the estimates. The solution is to build a *composed error* model:

$$Y_i = f(\mathbf{X}_i; \boldsymbol{\beta}) \exp(V_i - U_i), \tag{22}$$

*Table 1.* Dunker Residual Analysis.

| Rep | LSALES | OLS PRED | OLS RESID | COL PRED2 | COL RESID2 |
|---|---|---|---|---|---|
| Hamburg | 5.8889 | 6.1551 | −0.2662 | 6.6894 | −0.8005 |
| Bielefeld | 6.3936 | 6.2105 | 0.1831 | 6.7448 | −0.3512 |
| Krefeld | 6.7901 | 6.7899 | 0.0002 | 7.3242 | −0.5341 |
| Koln | 6.3716 | 6.4453 | −0.0737 | 6.9796 | −0.6080 |
| Dusseldorf | 6.4600 | 6.2953 | 0.1646 | 6.8297 | −0.3698 |
| Wuppertal | 6.3491 | 5.8148 | **0.5343** | 6.3491 | **0.0000** |
| Ludenscheid | 6.0777 | 5.9757 | 0.1019 | 6.5101 | −0.4324 |
| Kassel | 6.1485 | 5.9991 | 0.1493 | 6.5335 | −0.3850 |
| Frankfort | 5.7269 | 5.8954 | −0.1686 | 6.4297 | −0.7029 |
| Weisbaden | 5.6956 | 6.0939 | −0.3287 | 6.6282 | −0.8630 |
| Koblenz | 5.8319 | 6.0768 | −0.2449 | 6.6111 | −0.7792 |
| Mannheim | 6.5294 | 6.1554 | 0.3740 | 6.6898 | −0.1603 |
| Wurzburg | 5.9864 | 6.1463 | −0.1598 | 6.6806 | −0.6942 |
| Nurnberg | 5.5294 | 5.9555 | −0.4261 | 6.4899 | −0.9605 |
| Munchen | 5.9454 | 6.1901 | −0.2447 | 6.7244 | −0.7790 |
| Augsburg | 5.2523 | 5.2553 | −0.0031 | 5.7897 | −0.5374 |
| Stuttgart | 6.7719 | 6.5992 | 0.1727 | 7.1335 | −0.3616 |
| Karlsruhe | 6.1570 | 5.9640 | 0.1930 | 6.4983 | −0.3413 |
| West Berlin | 6.3802 | 6.3376 | 0.0425 | 6.8719 | −0.4918 |

*Source:* Parsons (1991).

which incorporates both technical inefficiency and random error (Aigner, Lovell & Schmidt, 1977; Meeusen & van den Broeck, 1977). The random error $V$ is assumed to be iid and symmetric and distributed independently of $U$. The most common empirical representation is

$$\ln Y = \beta_0 + \sum_{k=1}^{K} \beta_k \ln X_k + V - U. \tag{23}$$

Assuming that $V$ and $U$ are distributed independently of the regressors **X**, OLS estimation of (23) provides consistent estimates of the $\beta_k s$, but not $\beta_0$. While OLS does not generate the desired estimates of DMU-specific technical efficiency, it does provide a basis for a simple test of the presence of technical inefficiency in the data, which is indicated by negative skewness of OLS residuals. Under the null hypothesis of zero skewness of the OLS residuals, the test statistic

$$M3T = \frac{m_3}{\sqrt{\frac{6m_2^3}{N}}}, \tag{24}$$

where $m_2$ and $m_3$ are the second and third sample moments of OLS residuals, is asymptotically distributed as a standard normal random variable (Coelli, 1995, p. 253; Kumbhakar & Lovell, 2000, p. 73).[11] Note that the Dunker OLS residuals in Table 1 have a positive skew. Positive skewness is nonsensical in a composed-error model and suggests that the model is misspecified. One would not proceed to estimate the stochastic frontier production function for Dunker.

Dunker raises a more general issue: one's ability to identify the frontier. Suppose most DMUs are inefficient and so are below the frontier. Perhaps a few are just slightly below, a few far below, and most somewhat below. If, as well might be the case, the distance between the frontier and estimated observations is approximately normally distributed, then the estimated frontier will be essentially identical to the OLS line (Haughton et al., 2000, fn. 2).

If the test supports the existence of technical inefficiency, the maximum likelihood method is used to estimate all the parameters of the model.[12] The random error component is assumed normally distributed. The one-sided error component is usually distributed half normal but could be exponential, truncated normal, or gamma. Conditional on the maximum likelihood estimates of the parameters, technical efficiency is estimated for each DMU by decomposing the maximum likelihood residual term into a noise component and technical inefficiency component. The main computer programs for stochastic frontier estimation are LIMDEP and FRONTIER (Sena, 1999). The sample mean efficiencies will be sensitive to the distribution chosen for one-sided error; however, the ranking of DMUs by their individual technical inefficiencies are not particularly sensitive. The recommendation is to use a relatively simple distribution, such as the half normal, rather than a more flexible distribution, such as the gamma (Kumbhakar & Lovell, 2000, p. 90).[13]

BELGACOM, the state-owned Belgian telecommunications company, found itself undergoing privatization and facing competition from a number of private companies. In an attempt to stay close to its customers, BELGACOM opened outlets all over Belgium. These outlets were known as teleboutiques if they focused on the residential market (Fig. 1) or telebusinesses if they focused on the small and medium-sized enterprise (SME) market. A performance monitoring system was developed assist BELGACOM to evaluate the relative efficiency of its network of about 100 teleboutiques (Sinigaglia, 1997).[14]

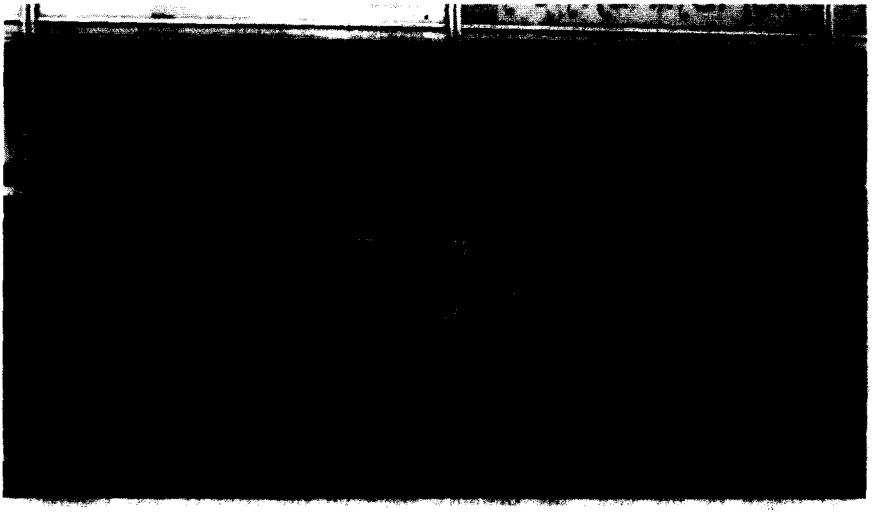

*Fig. 1.* Belgacom Telebotique.

A constant elasticity model with composed error was estimated. Two alternative measures of output were examined: total revenue and gross profit margin. Three controllable inputs were sales personnel, selling (product display) area, and number of opening hours. Uncontrollable factors included economic potential within a store's trading area. All factors were significant and explained 78 to 88% of the variation in the performance factors. The results were compared to output-oriented DEA with non-discretionary inputs. The two techniques identified the same units as being top-performing and also converged in terms of detecting the less efficient units.

Another study that compared a constant elasticity composed-error model to an output-oriented DEA model assessed advertising media spending inefficiency (Luo & Donthu, 2001). The reported results assumed that technical inefficiency followed a half-normal distribution. The truncated-normal and exponential distributions were also tried but were said to generate similar results to the half-normal distribution. The sales revenues of 94 of the leading 100 national advertisers were modeled as a function of three advertising spending variables: print, broadcast, and outdoors. The years 1997 and 1998 were examined separately. The mean inefficiency scores from DEA and SFA were not significantly different and the two inefficiency scores were highly correlated in 1998. However, the opposite results were found for 1997. The mean inefficiency scores were significantly different while the inefficiency scores were not correlated. These conflicting results led the researchers to conclude that both DEA and SFA should be used in all applications.

### *D. Incorporating Heteroscedasticity*

One complication is that the assumption of homoscedasticity of both error components in the stochastic production function may be violated. Either the systematic noise error component or the one-sided technical efficiency error component or both may be heteroscedastic. See Caudill and Ford (1993) and Caudill, Ford and Gropper (1995). In the case of retail outlets, the sources of noise and inefficiency might vary with the size of outlets.

When the symmetric noise error component exhibits heteroscedasticity, one obtains unbiased estimates of the $\beta_k s$ and a downward-biased estimate of $\beta_0$ (Kumbhakar & Lovell, 2000, pp. 116–118). The bias in the estimated intercept can be corrected once $\sigma_U$ is estimated. The problem is that there are now two sources of variation in the estimated technical efficiency: (1) the residual itself and (2) the weight attached to the residual, which has a noise component with nonconstant variance. Suppose that $\sigma^2_{V_i}$ does vary with the size of the DMU. Then the estimates of technical efficiency under the mistaken assumption of homoscedasticity will be biased upward for relatively small DMUs and downward for relatively large DMUs. The reason is that heteroscedasticity is improperly attributed to technical inefficiency. One cannot estimate DMU-specific variance parameters when one only has cross-sectional data. Instead one models $\sigma^2_{V_i}$ as a function of DMU-specific variables, such as size. Estimation can be done by maximum likelihood and then estimates of the technical efficiency of each DMU can be found.

When the one-sided error component exhibits heteroscedasticity, both the estimates of the technology parameters describing the structure of the production function and the estimates of technical efficiency will be adversely affected by mistakenly assuming homoscedasticity (Kumbhakar & Lovell, 2000, pp. 118–121). If heteroscedasticity varies with DMU size, then the estimates of technical efficiency under the mistaken assumption of homoscedasticity will be biased downward for relatively small DMUs and upward for relatively large DMUs. Thus, the impact of ignoring heteroscedasticity in $U$ is in the opposite direction of the impact of ignoring heteroscedasticity in $V$. Once again maximum likelihood estimation can be done and technical efficiencies found. It is important to emphasize that, when the presence of heteroscedasticity is reflected in the inefficiency component of the error term, the MLE line will not be parallel to the OLS line. This is illustrated in Fig. 2.

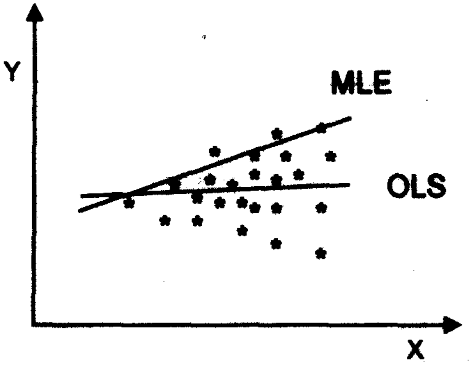


*Fig. 2.* Frontier and Regression Lines in the Presence of Heteroscedasticity.

When there is heteroscedasticity in both error components, one can hope that ignoring heteroscedasticity will result in a small overall bias because of offsetting impacts of heteroscedasticity in these two components (Kumbhakar

***Table 2.*** Estimating Prescription Frontier.

| | MLE | OLS | Mean Value of Variable |
|---|---|---|---|
| *Dependent variable* | | | |
| Log of mg prescribed per physician per year | | | |
| *Independent variables* | | | |
| Arteriosclerosis cases | 5.183 | 6.460 | 4.760 (/1000) |
| Congestive heart failure cases | 0.128 | 0.394 | 505.1 (/1000) |
| (Condition 1 prescriptions * Specialty A) squared | 0.796 | 0.211 | 1132 (/1000000) |
| (Condition 1 prescriptions * Specialty B) squared | 5.950 | 0.327 | 4822 (/1000000) |
| Index of oral contraceptive prescriptions (min 0, max 5) | 0.250 | 0.308 | 1.45 |
| Prescriptions, category 29 | 1.979 | 1.051 | 73.9 (/1000) |
| Prescriptions, category 32 | 0.060 | 0.038 | 406441 (/1000000) |
| Specialist in Specialty A? (Yes = 1) | –0.760 | –5.190 | 5411 (/1000000) |
| Constant | 2.907 | 2.322 | 0.033 |
| Log likelihood | 2.122 | 1.389 | |
| Adj. R-square | –15731 | 0.31 | |
| Number of observations | 9994 | 9994 | |

*Source*: Haughton et al. 2000, p. 39. Republished with permission of John Wiley & Sons, Inc. and the Direct Marketing Educational Foundation from Journal of Interactive Marketing, 14, 2000. Copyright © John Wiley & Sons.

***Table 3.*** Picking 1000 Target Physicians.

| | | Does the frontier method pick the doctor for a direct marketing intervention? | | |
|---|---|---|---|---|
| | | No | Yes | Total |
| Does the standard OLS model pick the doctor for a direct marketing intervention? | No | 8,558 | 436 | 8,994 |
| | Yes | 436 | 564 | 1,000 |
| Total | | 8,994 | 1,000 | 9,994 |

*Source*: Haughton et al. 2000, p. 39. Republished with permission of John Wiley & Sons, Inc. and the Direct Marketing Educational Foundation from Journal of Interactive Marketing, 14, 2000. Copyright © John Wiley & Sons.

& Lovell, 2000, pp. 121–122). A better approach would be to postulate a model that contains heteroscedasticity in both error components represented as functions and then to test the homoscedasticity assumptions.

Stochastic frontier analysis has been applied to determine which physicians should be targeted for a direct mailing for a widely used antiviral drug (Haughton et al., 2000). A stepwise OLS regression was used to distill a large number of doctor characteristics, such as age, specialty, and general prescription behavior, and geo-demographic variables into a more succinct set of independent variables. Then the MLE procedure assuming heterogeneity in both error components was estimated. A comparison of the results for the OLS and MLE methods is given in Table 2. The two approaches give significantly different values and will lead one to target different doctors for a direct mailing or other intervention as shown in Table 3.

### *E. Incorporating Exogenous Influences on Efficiency*

Having benchmarked performance, the next step is to see if systematic departures from the frontier can be explained. Here one is looking for sources of success. The second stage model is technical inefficiency = $g$(explanatory variables). Technical efficiency for a salesperson could be based on the individual's technical knowledge and the socioeconomic environment in which he is working. Technical efficiency for a retail outlet could be based on the size, age, and the other outlet characteristics as well as the characteristics of the market in which it is operating. Early empirical papers include Kalirajan (1981) and Pitt and Lee (1981).

A two-stage approach is thus adopted. The first stage involves the specification and estimation of the stochastic frontier production function and the estimation of technical inefficiency effects, assuming that these inefficiency effects are identically distributed. The second stage involves the specification of a regression model for predicted technical inefficiency effects. Since technical efficiency is bounded by zero and one, OLS is not the appropriate technique although it has been used in practice. One must either transform the dependent variable or use a limited dependent variable technique, such as tobit.[15] For example, a semi-log relationship was used by Kalirajan (1990):

$$\ln(\exp(-U)) = \alpha_0 + \sum_{k=1}^{K} \alpha_k Z_k + V \tag{25}$$

and a logistic model was used Mester (1997):

$$\hat{E}(U_i | V_i + U_i) = \frac{\exp(\mathbf{Z}'\boldsymbol{\alpha})}{1 + \exp(\mathbf{Z}'\boldsymbol{\alpha}0}, \tag{26}$$

in modeling cost efficiency. While the two-step approach seems reasonable, it contradicts the assumption of identically distributed inefficiency effects on the stochastic frontier.

The solution is to estimate the parameters of the stochastic frontier and inefficiency models *simultaneously*. The relevant technical literature includes Kumbhakar, Ghosh and McGucklin (1991), Reifschneider and Stevenson (1991), Huang and Liu (1994), and Battese and Coelli (1995). For example, the Battesse-Coelli model is

$$\ln Q = \beta_0 + \sum_{k=1}^{K} \beta_k \ln X_k + V - U$$

where $V$ iid $N(0, \sigma_V^2)$, $U$ iid $N(\mu, \sigma_U^2)$ with truncations at zero and

$$\mu = \delta_0 + \sum_{k=1}^{K} \delta_k Z_k.$$

Implementation uses the reparameterizations

$$\sigma^2 = \sigma_V^2 + \sigma_U^2 \text{ and } \gamma = \frac{\sigma_U^2}{\sigma_V^2 + \sigma_U^2}.$$

Additional information and references on incorporating exogenous influences on efficiency can be found in Kumbhaker and Lovell (2000, pp. 261–278).

A pilot study was conducted for a particular strategic business unit of the American subsidiary of a leading European chemical company (Parsons & Jewell, 1998). Cross-sectional data were available on 18 salespeople. The dependent variable was sales. The explanatory variables were salesperson effort, firm support, and potential. The associated operational definitions were the number of sales calls per customer, promotional dollars spent per customer, and average customer size. Technical inefficiency was subsequently related to the number of contacts per customer. The Battesse-Coelli model and its parameters were statistically significant. The mean technical efficiency of the sales representatives was 0.79.

## IV. FUTURE DIRECTIONS

Stochastic frontier analysis can be extended to incorporate multiple outputs, cost efficiency, panel data, Bayesian estimation, and allocative efficiency. These advances in econometric methodology provide promising avenues for further work on performance on performance measurement in marketing.

### *A. Multiple Outputs*

Performance evaluation often involves multiple criteria. A nominal advantage of data envelopment analysis over stochastic frontier analysis has been the ability to handle multiple outputs. In a Special Issue of the *International Journal of Research in Marketing* on "Channel Productivity," edited by A. Bultez and L. Parsons, DEA was used to assess the individual stores for a multi-store, multi-market retailer. The researchers noted "DEA is particularly appropriate for this evaluation because it integrates a variety of performance metrics . . ." (Thomas et al., 1998, pp. 487–488). Profit and sales were used as outcome measures. Similarly, in assessing outlets from a fast-food restaurant chain, one of the main advantages listed for DEA-based retail outlet productivity evaluations was "DEA accommodates *multiple* inputs and outputs" (Donthu & Yoo, 1998, p. 95). This work also stressed the importance of including behavioral output measures as well as financial output measures and used customer satisfaction along with sales. However, one does not have to move to DEA to handle multiple-output frontier situations. The single-output frontier model that allows estimation of frontier functions and technical efficiency can be generalized to handle multiple input, multiple output technologies.

Shepard's distance functions provide a characterization of the structure of production technology when multiple inputs are used to produce multiple outputs.

$$\mathrm{D_O}(\mathbf{X}_i, \mathbf{Y}_i; \boldsymbol{\beta}) = \exp(V_i - U_i), \tag{27}$$

An output distance function takes an output-expanding approach to the measurement of the distance from a DMU to the frontier. One property of output distance functions is

$$D_O(\mathbf{X}_i, \lambda\mathbf{Y}_i; \boldsymbol{\beta}) = \lambda D_O(\mathbf{X}_i, \mathbf{Y}_i; \boldsymbol{\beta}), \tag{28}$$

where $\lambda > 0$. A suitable choice for the normalizing variable $\lambda$ leads to an estimable composed error regression model. One possibility is the Euclidean norm of the frontier output vector, $\lambda = |\mathbf{Y}_i|^{-1} = (\Sigma_M Y_{mi}^2)^{-1/2}$, which Kumbhakar and Lovell (2000, p. 94) recommend because it is neutral with respect to outputs.[16] Also see Löthgren (1997). Inserting (27) into (28) with this normalization and rearranging yields

$$|\mathbf{Y}_i|^{-1} = D_O\left(\mathbf{X}_i, \frac{\mathbf{Y}_i}{|\mathbf{Y}_i|}; \boldsymbol{\beta} \exp(U_i - V_i)\right). \tag{29}$$

This is a reciprocal measure of output-oriented technical efficiency. The constant elasticity (Cobb-Douglas) functional form cannot accommodate multiple outputs without violating the requisite curvature properties of output space (Kumbhakar & Lovell, 2000, p. 143). Since the constant elasticity model does not have the correct properties to represent $D_O$, an appropriate flexible form must be selected. This is usually the translog, which provides a second-order approximation to any arbitrary function and thus can capture a wide variety of shapes (Hanssens, Parsons & Schultz, 2001, pp. 114–115). We will indicate shortly that translog might not be the best approximation to use when panel data are available. One potential problem with the distance function approach is that the normalized regressors may not be exogenous. If endogeneity is a serious issue, it should be addressed by a system of equations (Kumbhakar & Lovell, 2000, p. 95).

### *B. Estimation of Cost Efficiency*

Rather than focusing on the production frontier and taking an output-oriented approach, one could focus on the cost frontier and take an *input-oriented* approach. Multiple outputs are more easily handled in the cost frontier framework than the production frontier one. A discussion of estimation and decomposition of cost efficiency is given in Kumbhakar and Lovell (2000, pp. 131–183).

Econometric methods, although not stochastic frontier analysis, have been used in marketing to estimate cost functions for retail outlets within a chain. The most common functional form used has been the translog, which simply extends the log-log model by adding second-order terms.[17] The stochastic cost frontier can be expressed then as

$$\ln C = \beta_0 + \sum_{m=1}^{M} \alpha_m \ln Y_m + \sum_{k=1}^{K} \beta_k \ln P_k + \sum_{j=1}^{J} \beta_j \ln Z_j + \frac{1}{2}\sum_{m=1}^{M}\sum_{m'=1}^{M} \alpha_{mm'} \ln Y_m \ln Y_{m'}$$

$$+ \frac{1}{2}\sum_{k=1}^{K}\sum_{k'=1}^{K} \beta_{kk'} \ln P_k \ln P_{k'} + \sum_{k=1}^{K}\sum_{m=1}^{M} \delta_{km} \ln P_k \ln Y_m + V + U, \qquad (30)$$

where $C$ is the total cost by outlet, $Y_m$ is the units of output $m$ produced, and $P_k$ is the unit price of input $k$, and $Z_j$ is the units of allocative input $j$, an input which is not under the control of the store manager, used by the outlet, $V$ is the two-sided random-noise component, and $U$ is the nonnegative cost inefficiency component.[18] Note the plus sign in front of the cost inefficiency component – with the consequence that the composed error term will be positively skewed. Note further the different information requirements, especially the need to know input prices, for the estimation of cost efficiency. Thus there are two sources of cost inefficiency: input-oriented technical inefficiency and input allocative inefficiency. A single-equation cost frontier model such as (30) provides DMU-specific estimates of cost efficiency but does not permit the decomposition of this cost efficiency into its technical and allocative efficiency components. Such decomposition requires the use of input quantity or input cost share data and a simultaneous-equation model (Kumbhakar & Lovell, 2000, pp. 146–166). Thus, estimation is improved by simultaneously considering the cost-share equations that result if cost-minimizing levels of each input are selected:

$$CS_k = \beta_k + \sum_{k'=1}^{K} \beta_{kk'} \ln P_{k'} + \sum_{m=1}^{M} \gamma_{km} \ln Y_m + W_k, \qquad (31)$$

where $CS_k$ is the cost share of input $k$ out of the total cost for a particular outlet. Note that a behavioral objective – cost minimization – has been imposed on the DMU.

In the estimation of a stochastic frontier production all inputs are treated the same. However, in the estimation of a cost frontier, one may exploit the differences between inputs that are variable and those that are quasi-fixed. In particular, one might use a variable cost frontier rather than a total cost frontier.

To get an idea of the input-oriented approach, consider the estimation of three non-frontier cost functions for branches of banks.[19] The largest commercial bank in Greece had a domestic branch network amounting to over 400 branches, of which 362 were studied (Pavlopoulos & Kouzelis, 1989). A third of the branches were situated in Greece's three major cities. The bank was considered to produce three outputs: new accounts, loans granted, and ancillary services. Each was measured as the number of transactions. Input prices were for capital, labor, and management. Capital was measured as the annual rental cost of office space while labor and management were operationalized as average annual salaries per employee. Two variables not under control of the branch were included. A ratio, constructed from the number of competing firms in the relevant market and their market shares, was designed to capture the influence of competition. A technology dummy variable was used to indicate whether or not a branch was connected with the on-line system of the bank. A measure of long-run total cost was approximated by the annual operating cost of each branch, interest paid to deposits excluded. A translog multiproduct function form with certain restrictions, such as linear homogeneity in input prices, was estimated by restricted OLS. The goodness of fit was high. The competition and technology variables proved not to be statistically significant. As part of the analysis, an appropriate Chow-type test showed that there was no statistically significant difference in the behavior of costs between those branches located in the three major cities and those in the rest of the country.

A similar translog cost function was estimated for a major Canadian bank (Doukas & Switzer, 1991). Branches that were part of a hub and spoke type branch banking network as well as central administrative and marketing units were omitted, leaving a sample 563 branches. These sample branches spanned both urban and rural areas and included both retail and commercial banking operations. The focus was on operating expenditures using total staff costs, including part-time worker costs and benefits, premises expenses, and equipment expenses. Thus factor input prices included the price of labor, the price per physical area of premises employed, and the price of equipment. There were seven outputs covering both deposits and loans: (1) total consumer lending, including consumer installment loans, credit card balances, and overdrafts, (2) mortgage loans, (3) total business credit, including business loans, loan equivalent securities and Banker's Acceptances, (4) total personal demand and savings deposits, (5) total business demand and savings deposits, (6) total term deposits, and (7) total non-interest revenues. The explanatory power of the model was quite high. Further analysis revealed that retail-oriented branches enjoyed greater than average economies of scale but such benefits diminished more quickly as branch size increased.

Performance comparisons among branches were a critical management issue for a commercial bank within a large metropolitan area in Latin America (Kamakura, Lenartowicz & Ratchford, 1996). Four outputs, expressed in monetary units, were (1) cash deposits, (2) other deposits, such as checks and money orders, (3) funds in transit in the branch, which already have their destination, such as tax collection, payment of bills, or pay checks, and (4) service fees charged to customers by the branch to pay for their transfers, checkbooks, statements, and so on. Two inputs were total number of man-hours of direct (clerks) labor allocated at the branch and floor area, the size (in square meters) of the customer services area. Floor area was treated as an allocative input, that is, under control of the bank but not the branch manager. Initially, another allocative input, the total number of teller stations, was also considered for inclusion in the model but it turned out to be highly collinear with floor area.

The researchers noted, "If estimated with standard econometric methods, the translog cost function in (30) does not represent the 'minimum' cost frontier. Rather this function represents the 'mean' or expected cost for any combination of input prices and output volumes. Therefore, $U$ must be viewed as the technical inefficiency (in log-cost) relative to the 'average' outlet operating at the same scale level. This view is compatible with our particular purpose of evaluating multiple retail outlets relative to each other. However, if one is interested in measuring technical inefficiency relative to the minimum cost frontier (see Ferrier & Lovell, 1990), $U$ must be restricted to non-negative values, and assumed to be distributed across outlets as [a specific one-sided distribution]." They then proceeded to identify sets of retail outlets operating under similar conditions while simultaneously estimating multiple cost functions for these classes. Estimates were obtained using a fuzzy-clustering regression procedure, which required longitudinal data on inputs and outputs for each outlet, i.e. panel data. They conjectured that the clusterwise translog cost function approach represents a compromise between the flexible piece-wise linear deterministic frontier in DEA and the stochastic estimation of a single translog function.

There remains a need to employ stochastic frontier analysis to estimate minimum cost functions for retail outlets within a chain. Such an analysis may be enhanced if panel data are available.

### *C. Panel Data*

Applications of stochastic frontier analysis in marketing have focused on cross-sectional data. Econometricians have extended the technique to panel data; that is, repeated observations on DMUs (Kumbhakar & Lovell, 2000, pp. 95–115). Even conventional panel data analysis can provide insights that a single cross section cannot. In particular, technology is unlikely to remain constant over time. The longer the panel time horizon, the more probable it is that technical change occurs. The common practice is to include time among the explanatory variables as a proxy for technical change. For example, panel data was used to estimate returns to scale and productivity change for a medium-sized nine-unit chain of retail book and office supply stores (Ratchford & Stoops, 1988). Data consisted of hours worked, shelf space, and quantity sold for each of four departments: books, office supplies, art supplies, and fine stationery. A non-frontier translog model was used to estimate labor demand as a function of physical output of each department, shelf space, and time trend. All variables, except time (months), were expressed as natural logs. A separate trend term was included for each store. The translog model was found to be statistically superior to the embedded log-log model. The impact of time was negative indicating that less labor was needed as time went on. Although there were no major technological changes adopted by the chain, such as automated checkouts, there were changes in the way labor was organized within stores, including increased customer self-service. The empirical finding of productivity gain was thus judged plausible.

The translog represents a second-order Taylor series approximation of an arbitrary function at a point. However, OLS estimates of a second-order polynomial do not generally correspond to the underlying Taylor expansion of the underlying function at an expansion point and are biased estimates of the series expansion (White, 1980). This inadequacy of the translog is an issue in banking where scale and product mix are often far from the mean. What is needed is a global approximation.

Empirical work on banking, e.g. Mitchell and Onvural (1996), has introduced the Fourier-flexible functional form, in which the translog is a special case, to model cost frontiers:

$$\ln C = \beta_0 + \sum_{m=1}^{M} \alpha_m \ln Y_m + \sum_{k=1}^{K} \beta_k \ln P_k + \frac{1}{2} \sum_{m=1}^{M} \sum_{m'=1}^{M} \alpha_{mm'} \ln Y_m \ln Y_{m'}$$

$$+ \frac{1}{2} \sum_{k=1}^{K} \sum_{k'=1}^{K} \beta_{kk'} \ln P_k \ln P_{k'} + \sum_{k=1}^{K} \sum_{m=1}^{M} \delta_{km} \ln P_k \ln Y_m$$

$$+ \sum_{m=1}^{M} [\phi_m \cos y_m + \theta_m \sin y_m]$$

$$+ \sum_{m=1}^{M} \sum_{m'=m}^{M} [\phi_{mm'} \cos(y_m + y_{m'}) + \theta_{mm'} \sin(y_m + y_{m'})]$$

$$+ \sum_{m=1}^{M} \sum_{m'=m}^{M} \sum_{m''=m'}^{M} [\phi_{mm'm''} \cos(y_m + y_{m'} + y_{m''})\_ + \theta_{mm'} \sin(y_m + y_{m'} + y_{m''})]$$

$$+ U + V \tag{32}$$

where the $y_m$ are adjusted values of $Y_m$ such that they span the interval $[0.1 \times 2\pi, 0.9 \times 2\pi]$, cutting 10% off of each end of the $[0, 2\pi]$ interval to reduce approximation problems near the endpoints (Berger, Leusner & Mingo, 1997, p. 146). The Fourier-flexible form is a semi-nonparametric approach that uses data to infer the relationships among variables when the true functional form of the relation is not known. Specification error is minimized at the cost of approximation error, which arises from having to choose a subset of trigonometric terms to represent the cost function.

The cost efficiency of over 760 branches of a large U.S. commercial bank for a three-year period was examined (Berger, Leusner & Mingo, 1997). Two alternate cost specifications – intermediation and production – were tried. In the intermediation specification, costs include total operating costs plus interest expenses, outputs are measured as the number of dollars intermediated, and both physical and financial input prices are included. In this particular intermediation analysis, there were four outputs: consumer transaction accounts, consumer nontransaction accounts, business transaction accounts, and business nontransaction accounts. Input prices included average wage rate and average rental rate on capital faced by the branch in its local market. In the production approach, costs included operating expenses only, outputs were the number of transactions completed, and only physical input prices were specified. In this empirical study, the six outputs were the numbers of deposit accounts, debits, credits, accounts opened, accounts closed, and loans originated.

While the intermediation approach is more inclusive and captures the essence of a financial intermediary and provides a good indicator of profitability, the production approach seems a more natural one from the point of view of analyzing branch performance. A branch is viewed as a producer of depositor services for the bank. The bank, in turn, then makes the decisions on how to intermediate the funds. The researchers go on to highlight a key consideration, the sensitivity of each approach to the number of transactions per dollar of deposits. Branches in affluent neighborhoods are likely to have customers with fewer transactions per dollar in their accounts, which makes costs per dollar of deposits lower. The intermediation approach might mistakenly treat this as high efficiency. This is not a problem in the production approach as the number of transactions is directly measured as a service output.

Fourier-flexible and translog models were estimated separately for the intermediation and production approaches. The null hypothesis that the nested translog specification (the coefficients of all trigonometric terms were jointly zero) was correct was rejected. Both the intermediation and production results generated similar broad findings. Branch X-inefficiencies, variations in costs ascribed to differences in managerial ability, were much larger than branch scale inefficiencies. Moreover, branches appeared to be, on average, about half of the average-cost-minimizing size for their product mixes. This means that there were twice as many branches as would minimize costs. The dispersion of measured X-efficiency suggested that the bank's management was not able to control fully the costs at its branch offices through its policies and procedures, incentives, and supervision. Thus, the quality of local management was important in determining the performance of branches.

As already noted, the longer the panel, the more likely technical progress occurs and a time indicator should be included in any model. A long panel also means technical efficiency may well change and so a time-varying representation of technical efficiency is required. Indeed, both phenomena must be included in the stochastic frontier production function model so as to be able to disentangle the effect of technical change from that of technical efficiency change (Kumbhakar & Lovell, 2000, pp. 108–115). The importance of including both effects is illustrated by an empirical study by Battese and Coelli (1992). When the year of observation was excluded from the stochastic frontier, the technical efficiencies of the DMUs were time varying. However, when the time indicator was included, not only was it statistically significant, but the corresponding technical efficiencies were time invariant. Moreover, the stochastic frontier was not significantly different from the traditional average response function in this case.

Just as with cross-sectional models, heteroscedasticity can present in either or both error components in panel data models. Heteroscedasticity can be assessed for either the time-invariant or time-varying technical efficiency cases. See Kumbhakar and Lovell (2000, pp. 122–130).

Panel data has several additional advantages when doing stochastic frontier analysis (Kumbhakar & Lovell, 2000, pp. 95–97). First, the strong distributional assumptions on each error component can be relaxed. Second, the technical inefficiency error term need not be independent from the regressors. This is relevant because technical inefficiency might well be correlated with the input vectors DMUs choose.

A distribution-free panel data approach can be used to disentangle inefficiency differences from random errors that temporarily give DMUs high or low costs. It does this by assuming inefficiencies are stable over time and that random error tends to average out over time.

### *D. Bayesian Frontier Estimation*

Bayesian analysis of stochastic frontier models with composed error has been shown to be both theoretically and practically feasible (see, for example, van den Broeck et al., 1994; Osiewalski & Steel, 1998). The paradigm allows for direct posterior inference on DMU-specific efficiencies. The need to choose a particular sampling model for the inefficiency error term is avoided by mixing over different models. Thus, the Bayesian approach leads to the posterior probabilities of these models, indicating which of them is most favored by the data. Additional insights from a Bayesian perspective can be obtained using panel data (Kim & Schmidt, 2000). Work on applying the hierarchical Bayesian method in marketing, such as that by Cain (2001), might be well extended to cover the frontier case.

### *E. Allocative Efficiency*

Technical efficiency has been addressed in isolation from allocative efficiency. Both the technique or manner of applying and the levels of application of inputs, however, determine the economic efficiency of production. Thus simultaneous estimation of the two components should provide more efficient estimates than estimating either the production function alone or the profit function alone (Kalirajan, 1990).

# V. CONCLUDING REMARKS

Management wants to set its expectations of performance based on the best that can be achieved. A frontier production function that specifies the maximum output attainable at given levels of input is the key to understanding performance results and specifying performance standards. On occasion, marketing builds directly on the economic concept of production, that is, the process of transforming labor and capital into goods. This is true in the area of assessing the performance of retail outlets. Other times, marketing is more interested in a more general transformation process that yields performance measures of interest to management. This is the case in sales force performance studies. Whatever the application area, the estimated stochastic frontier model provides a basis for comparing individual DMUs to each other or to the ideal production frontier. In sum, marketing productivity analysis holds the promise of a manager being able to assess marketing performance and then to take steps to improve it. See Parsons (1994) for further discussion.

# NOTES

1. Care should be exercised when extending the analysis to groups of salespeople, such as at the district or region levels, instead of individual salespeople as in this model. The aggregation may not necessarily hold in taking the original model from the individual level to the aggregate level. Even a linear model may have heteroscedasticity if there are differing numbers of individuals in each district or region.
2. Emphasis added.
3. Shortly after completion of this study, IBM decided that it was a technology company, not a retailer, and it sold its Product Centers to NYNEX, a regional Bell operating company arising from the then recent breakup of AT&T.
4. Parsons (1992) reanalyzed the IBM Product Center data using DEA.
5. This is not the case for other sales performance models that have been discussed. Indeed, the very purpose of these models has been to distinguish the impact of factors under control of the DMU from factors not under the control of the DMU.
6. An output-based Debreu-Farrell style measure of technical efficiency focuses on equiproportionate expansion of *all* outputs and is a radial measure, which has desirable properties such as invariance to changes in units of measurement. A more exacting standard is an output-based Koopmans style measure of technical efficiency, which focuses on the increase in *any* output but unfortunately is a nonradial measure (Färe, Grosskopf & Lovell, 1994, pp. 7–9; Kumbhakar & Lovell, 2000, pp. 42–46).
7. Operations management often focuses on minimizing inputs instead of maximizing outputs.
8. For information about DEA, see Coelli, Rao, and Battesse (1998) or Cooper, Seiford and Tone (2000).
9. Representative plots of these various distributions can be found in Kumbhakar and Lovell (2000).

10. A related procedure is modified ordinary least squares (MOLS). The disturbances are assumed to follow an explicit one-sided distribution. After OLS estimation, the intercept is shifted up by the mean of the assumed one-sided distribution. However, this shift may not be large enough for the frontier to bound all DMUs from above. Thus, one or more DMUs may have technical efficiency scores greater than one!

11. The asymmetry of the distribution is given by the third moment of the residuals:

$$m_3 = \sum_{i=1}^{N} (\hat{W}_i^* - E[\hat{W}_i^*])^3.$$

This quantity is estimable with OLS so long as the slope estimators are consistent (Greene, 1997, p. 99).

12. An alternative procedure is the method of moments (Kumbhakar & Lovell, 2000, pp. 90–93).

13. Note, however, that the half normal is rather inflexible and incorporates the assumption that most observations are clustered near full efficiency (i.e. zero mode), with larger values of inefficiency being decreasingly likely. More likely, the factors that relate to managerial efficiency, such as educational training, intelligence, and persuasiveness, are distributed non-monotonically with a non-zero mode. The truncated-normal and gamma have been posited as more plausible alternative models of inefficiency (Stevenson, 1980; Greene, 1980). Note further that the ML estimator should be used in preference to the COLS estimator whenever possible under the half-normal assumption, especially when the contribution of technical efficiency effects to the total variance term is large (Coelli, 1995).

14. This work was done at the Centre for Research on the Economic Efficiency of Retailing under the direction of A. Bultez. See also Singaglia et al. (1995).

15. In an interesting meshing of techniques, DEA was used to calculate radial technical efficiency scores of Indian commercial banks. Then SFA was employed to attribute variation in calculated efficiency scores to three sources: a temporal component, an ownership component, and a random noise component. Publicly-owned Indian banks were found to be more efficient than foreign-owned banks and privately-owned Indian banks (Bhattacharyya, Lovell & Sahay, 1997).

16. One could simply choose an arbitrary output, such as the *m*th output, and set $\lambda = 1/Y_{mi}$. See, for example, Coelli and Perelman (1999) or Fuentes, Grifell-Tajé and Perelman (2001, p. 85).

17. The multiplicative model does not admit *U*-shaped cost curves; consequently, the optimal size of a store cannot be determined from this model.

18. The same distributional assumptions about the error terms can be made as those in the stochastic production frontier model.

19. An early study, which developed and tested a multiplicative model of branch operating costs for a large bank operating in a relatively small country, provides another example (Murphy & Orgler, 1982).

# REFERENCES

Aigner, D. J., & Chu, S. F. (1968). On estimating the industry production function. *American Economic Review, 58*, 826–839.

Aigner, D. J., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, *6*, 21–37.

Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency, and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis*, *3*, 153–169.

Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, *20*, 325–332.

Berger, A. N., Leusner, J. H., & Mongo, J. J. (1997). The efficiency of bank branches. *Journal of Monetary Economics*, *40*, 141–162.

Bhattacharyya, A., Lovell, C. A. K., & Sahay, P. (1997). The impact of liberalization on the productive efficiency of Indian commercial banks. *European Journal of Operational Research*, *98*, 332–345.

Böcker, F. (1986). The Dunker Company. In: D. J. Dalrymple & L. J. Parsons (Eds), *Marketing Management* (pp. 709–719). John Wiley & Sons, New York.

Cain, L. B. (2001). Accounting for territory representative heterogeneity in sales force performance evaluation. Unpublished Ph.D. Thesis, University of Pennsylvania.

Camp, R. C. (1989). *Benchmarking*. Milwaukee, WI: ASQC Quality Press.

Camp, R. C. (1995). *Business Process Benchmarking*. Milwaukee, WI: ASQC Quality Press.

Caudill, S. B., & Ford, J. M. (1993). Biases in frontier estimation due to heteroscedasticity. *Economic Letters*, *41*, 17–20.

Caudill, S. B., Ford, J. M., & Gropper, D. M. (1995). Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business Economic Statistics*, *13*, 105–111.

Coelli, T. J. (1995). Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis. *The Journal of Productivity Analysis*, *6*, 247–268.

Coelli, T. J., & Perelman, S. (1999). A comparison of parametric and non-parametric distance functions: With application to European railways. *European Journal of Operations Research*, *17*, 326–339.

Coelli, T. J., Rao, D. S. P., & Battesse, G. E. (1998). *An Introduction to Efficiency and Productivity Analysis*. Norwell, MA: Kluwer Academic Publishers.

Cooper, W. W., Seiford, L. M., & Tone, K. (2000). *Data Envelopment Analysis*. Norwell, MA: Kluwer Academic Publishers.

Cravens, D. W., Woodruff, R. B., & Stamper, J. C. (1972). An analytical approach to evaluating sales territory performance. *Journal of Marketing*, *36*, 31–37.

Donthu, N., & Yoo, B. (1998). Retail productivity assessment using data envelopment analysis. *Journal of Retailing*, *74*, 89–105.

Doukas, J., & Switzer, L. N. (1991). Economies of scale and scope in Canadian branch banking. *Journal of International Financial Markets, Institutions & Money*, *1*, 61–84.

Färe, R., Grosskopf, S., & Lovell, C. A. K. (1994). *Production Frontiers*. Cambridge: Cambridge University Press.

Ferrier, G. D., & Lovell, C. A. K. (1990). Measuring cost efficiency in banking: econometric and linear programming evidence. *Journal of Econometrics*, *46*, 229–245.

Fuentes, H. J., Grifell-Tajé, E., & Perelman, S. (2001). A parametric distance function approach for Malmquist productivity index estimation. *Journal of Productivity Analysis*, *15*, 79–94.

Greene, W. H. (1980). On the estimation of a flexible frontier production function. *Journal of Econometrics*, *13*, 101–115.

Greene, W. H. (1997). Frontier production functions. In: M. H. Persaran & P. Schmidt (Eds), *Handbook of Applied Econometrics (Vol. II: Microeconomics)* (pp. 81–166). Oxford: Blackwell Publishers.

Hanssens, D. M., Parsons, L. J., & Schultz R. L. (2001). *Market Response Models*. Norwell, MA: Kluwer Academic Publishers.

Haughton, D., Haughton, J., Kelly-Hawke, A., & Moriarty, T. (2000). The use of frontier estimation in direct marketing. *Journal of Interactive Marketing, 14*, 33–42.

Huang, C. J., & Liu, J.-T. (1994). Estimation of a non-neutral stochastic frontier production function. *Journal of Productivity Analysis, 5*, 171–180.

Jackson, D. W. Jr., Schlacter, J. L., & Wolfe, W. G. (1995). Examining the bases utilized for evaluating salespeople's performance. *Journal of Personal Selling & Sales Management, 15*, 57–65.

Kalirajan, K. P. (1981). An econometric analysis of yield variability in paddy production. *Canadian Journal of Agricultural Economics, 29*, 283–294.

Kalirajan, K. P. (1990). On measuring technical efficiency. *Journal of Applied Econometrics, 5*, 75–85.

Kamakura, W. A., Lenartowicz, T., & Ratchford, B. T. (1996). Productivity assessment of multiple retail outlets. *Journal of Retailing, 72*, 333–356.

Kim, Y., & Schmidt, P. (2000). A review and empirical comparison of Bayesian and classical approaches to inference on efficiency levels in stochastic frontier models with panel data. *Journal of Productivity Analysis, 14*, 91–118.

Kumbhakar, S. C., Ghosh, S., & McGuckin, J. T. (1991). A generalized production function approach for estimating determinants of inefficiency in U.S. dairy farms. *Journal of Business Economic Statistics, 9*, 279–286.

Kumbhakar, S. C., & Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.

Löthgren, M. (1997). A multiple output stochastic ray frontier production model. Working Paper Series in Economics and Finance No. 158, Department of Economic Statistics, Stockholm School of Economics, Stockholm, Sweden.

Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review, 18*, 435–444.

Mester, L. J. (1997). Measuring efficiency at U.S. banks, accounting for heterogeneity is important. *European Journal of Operations Research, 98*, 230–242.

Mitchell, K., & Onvural, N. (1996). Economies of scale and scope at large commercial banks: Evidence from the Fourier flexible functional form. *Journal of Money, Credit, and Banking, 28*, 178–199.

Murphy, N. B., & Orgler, Y. E. (1982). Cost analysis for branching systems: Methodology, test results, and implications for management. *Journal of Financial Analysis, 5*, 181–188.

Osiewalski, J., & Steel, M. F. J. (1998). Numerical tools for Bayesian analysis of stochastic frontier models. *Journal of Productivity Analysis, 10*, 103–117.

Parsons, L. J. (1991). Estimation of a frontier production function for a sales force. In: *TIMS Marketing Science Conference*. Newark, DE: University of Delaware/ Dupont.

Parsons, L. J. (1992). Retail outlet performance study. In: *TIMS Marketing Science Conference*. London Business School, London.

Parsons, L. J. (1994). Productivity and relative efficiency: Past and future? In: G. L. Lilien, G. Laurent & B. Pras (Eds), *Research Traditions in Marketing* (pp. 169–196). Norwell, MA: Kluwer Academic Publishers.

Parsons, L. J., & Jewell, S. (1998). Measuring the performance of an industrial sales force. In: *INFORMS Marketing Science Conference* Fontainebleau, France: INSEAD, July 1998.

Parsons, L. J., & Vanden Abeele, P. (1981). Analysis of sales call effectiveness. *Journal of Marketing Research, 18*, 107–113.

Pavlopoulos, P. G., & Kouzelis, A. K. (1989). Cost behavior in the banking industry: Evidence from a Greek commercial bank. *Applied Economics, 21*, 285–293.

Pitt, M. M., & Lee, M.-F. (1981). The measurement sources of technical inefficiency in the Indonesian weaving industry. *Journal of Developmental Economics, 9*, 43–64.

Ratchford,. T., & Stoops, G. L. (1988). A model and measurement approach for studying retail productivity. *Journal of Retailing, 64*, 241–263.

Reifschneider, D., & Stevenson, R. (1991). Systematic departures from the frontier, a framework for the analysis of firm inefficiency. *International Economic Review, 32*, 715–723.

Sena, V. (1999). Stochastic frontier estimation: A review of the software options. *Journal of Applied Econometrics, 14*, 579–586.

Sinigaglia, N. (1997). Measuring retail units efficiency, a technical approach. Unpublished Ph.D. Thesis, Facultés Universitaires Catholiques de Mons, Belgium.

Sinigaglia, N., Zidda, P., Panier, V., & Bultez, A. (1995). Looking for rules: Retail units linked-up efficiency standards. In: M. Bergadaà (Ed.), *Marketing Today and for the 21st Century: Proceedings of the 24th Annual Conference of the European Marketing Academy.* Cergy-Pontoise, France: ESSEC.

Schmidt, P. (1976). On the statistical estimation of parametric frontier production functions. *Review of Economics and Statistics, 28*, 238–239.

Stevenson, R. E. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics, 13*, 57–66.

Thomas, R. R., Barr, R. S., Cron, W. L., & Slocum, J. W. Jr. (1998). A process for evaluating retail store efficiency: a restricted DEA approach. *International Journal of Research in Marketing, 15*, 487–503.

Timmer, C. P. (1971). Using a probabilistic frontier production function. *Journal of Political Economy, 79*, 776–794.

Vandenbosch, M. B., & Weinberg, C. B. (1993). Salesforce operations. In: J. Eliashberg & G. L. Lilien (Eds), *Handbooks in Operations Research and Management Science (Vol. 5: Marketing)* (pp. 653–694). Amsterdam: North-Holland.

van den Broeck, J., Koop, G., Osiewalski, J., & Steel, M. F. J. (1994). Stochastic frontier models: A Bayesian perspective. *Journal of Econometrics, 61*, 273–303.

White, H. (1980). Using least squares to approximate unknown regression functions. *International Economic Review, 21*, 149–169.

Winsten, C. B. (1957). Discussion on Mr. Farrell's paper. *Journal of the Royal Statistical Society: Series A, General, 120* (Part 3), 282–284.