Statistica Neerlandica (1998) Vol. 52, nr. 3, pp. 303–323

Modeling large data sets in marketing*

S. Balasubramanian

Department of Marketing, University of Texas at Austin, Austin, TX 78712, USA

S. Gupta

508 Kris Hall, Graduate School of Business, Columbia University, New York, NY 10027, USA

W. Kamakura

University of Pittsburgh, 318 Mervis Hall, Pittsburgh, PA 15260, USA

M. Wedel†

Faculty of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands

In the last two decades, marketing databases have grown significantly in terms of size and richness of available information. The analysis of these databases raises several information-related and statistical issues. We aim at providing an overview of a selection of issues related to the analysis of large data sets. We focus on the two important areas: single source databases and customer transaction databases. We discuss models that have been used to describe customer behavior in these fields. Among the issues discussed are the development of parsimonious models, estimation methods, aggregation of data, data-fusion and the optimization of customer-level profit functions. We conclude that problems related to the analysis of large databases are far from resolved, and will stimulate new research avenues in the near future.

Key Words and Phrases: response models, single source data, customer transaction data.

1 Introduction

Market research data form the core of marketing strategy. The "marketing concept" (as opposed to, for example, the "product concept") involves a customer focus and the integrated use of marketing instruments. In order to implement the marketing concept and to develop products that are tailored to customer needs, marketing data are essential. In the last two decades marketing databases have grown significantly in terms of size and richness of available information. Telephone interviewing, for

* The authors are very grateful to Philip Hans Franses for the initiative to organize the RIBES conference on large data sets and are indebted to him for the invitation to write this paper.

† m.wedel@eco.rug.nl

example, has facilitated the collection of both judgmental and factual data from consumers. The advent of computer assisted interviewing has increased the speed and reduced the costs of data collection. In particular, such systems have provided advantages in international marketing research in Europe and elsewhere. In such studies, samples are often stratified by country and large numbers of respondents are interviewed for each country, leading to large cross-national databases. In other fields of research, unobtrusive measurement of consumer and firm behavior has likewise reduced the costs of data collection, thus enabling sample sizes to increase. For example, in the measurement of television exposure, devices called ''people meters'' are connected to the television sets in households. These devices monitor the channels to which the television sets of household panels are tuned into, on a continuous basis. Specialized firms keep track of advertising media and convert advertising data to monetary equivalents to register advertising expenditures. In evaluating the effectiveness of advertisements, packaging design and shelf-space allocation, both print and TV, consumers' eye movements are recorded in very small time intervals when looking at shelves, products, ads or commercials. In addition, recently the Internet as a source of marketing intelligence has begun to yield substantial amounts of consumer information.

However, much of the impetus for the increase in the size of commercial databases has come from the use of scanning devices at checkout counters in supermarkets to collect sales data by U.S. marketing research companies such as A. C. Nielsen and IRI. Each time a product passes the checkout counter, information such as brand, package size, price, and other marketing variables including display, coupons and feature ads are registered and stored in the retailer's computer. For outlets belonging to the retail panels of these research agencies, this information is downloaded overnight and processed for marketing research purposes. Whereas sales data had been previously collected by store audits conducted by trained personnel, the utilization of scanning devices has greatly facilitated the collection of sales data without human intervention. Rapidly this system of data-collection was introduced in most developed countries in Europe, America, Australia and Asia. The unobtrusive measurement of consumer shopping behavior was received very positively by firms, especially in the sector of fast moving consumer goods. Scanner data developed into an important input in the development of marketing strategy, and influenced strategic decisions regarding market segmentation, price setting and distribution, as well as tactical day-to-day decisions on the use of promotions such as price discounts, coupons, displays, and feature ads. These decisions are aimed at inducing brand switching, or at preempting or retaliating to, competitive actions. Subsequently, the advent of customer identification cards enabled the collection of household level scanner panel data. These data contain detailed accounts of the purchase behavior of individual households. With the automated collection of sales data using scanning, the amounts of data that could be stored and processed increased. The sizes of data matrices expanded in all directions. First, automated data collection led to the recording of decisions from large samples of consumers, sometimes covering the

entire population of interest (but this need not always be the case, since the population of interest may, for example, include customers of other brands, firms, retail outlets, or non-buyers). Second, data were collected on a daily or weekly instead of monthly basis. Third, due to continuous new product introduction, the number of brands in most markets increased as well. For example, in the US, there are more than 1200 SKUs (or stock-keeping units) in the toothpaste category.

A second marketing area where very large databases exist is direct marketing. Many companies, especially in the mail-order business, nowadays keep track of their transactions with customers. This is increasingly true of companies in other sectors as well, such as firms in retailing, banking and communications sectors. The feasibility of collecting such individual transaction-level data has increased through the introduction of information systems, database systems, personal customer credit and identification cards. Since long, companies have compiled databases containing the details of all transactions (e.g. type of product, date, mode of payment) with all customers for accounting purposes. More recently, companies have realized that such internal lists can be used not only for administrative and accounting purposes, but for marketing purposes as well. In addition, specialized firms collect data that is sold for direct marketing purposes. Typically, such databases include geo-demographic and life-style characteristics at the ZIP code level (e.g. age, education, income, occupation, dwelling characteristics, car ownership, subscriptions, etc.). These data are purchased by firms and either used as stand-alone data for direct marketing purposes or linked to their own customer databases.

Obviously, the availability of these large databases in various fields offers opportunities for the improvement of marketing planning and implementation. Central to the marketing concept is a thorough understanding of customer needs. The data-revolution in marketing has facilitated such understanding. However, the analysis of these databases poses specific problems. In many cases, the amounts of data available renders traditional statistical analysis problematic. In practice as well as in academia this has led people to resort to data mining: semi automated data analysis based on techniques for evolutionary computation, including genetic algorithms and neural networks. In the present article, however, we take the stance that rigorous analysis based on statistical theory has not lost its worth, but needs to be adapted to meet the requirements posed by the new conditions of large amounts of data. We provide an overview of a selection of issues related to the statistical analysis of large data sets. In particular, we choose to focus on the two areas we identified above: single source databases and customer transaction databases. We review issues related to modeling such data and provide an outlook to the future. This review is cast against two of the major paradigms that have characterized progress in academic marketing research. First, in the context of substantive relevance, the approaches developed should potentially contribute to the solution of management problems. Second, in terms of statistical and analytical contribution, the models proposed should demonstrably outperform existing procedures in terms of theoretical foundation, robustness, fit to actual data, cross-validation and prediction of hold out data.

2 Marketing models applied to single-source databases

The ability to combine store-level data about the store environment (prices, sales promotions, displays, etc.) collected through scanning, with household-level data into what is commonly known as single-source databases led to the development of models that provide more detailed insights into consumer response to marketing stimuli. Before these single-source databases were available, managers had to be content with either a market-level assessment of response to their marketing policies, or with a mere descriptive analysis of household-level behavior. With single-source databases, researchers are now able to directly relate marketing policies to a household's decision to purchase in a product category, of how much volume to buy, and which brand to choose. These analyses produce a better understanding of the competitive structure in a given market and of the various consumer segments in that market. This understanding helps managers assess the profitability of promotions and plan marketing actions for long-term results.

2.1 *Brand choice*

Random-utility choice models have been widely applied in marketing to measure the impact of price and sales promotions on brand choice. The basic formulation of these models is

$$P_{ijt} = P(y_{ijt} = 1 \mid \alpha, \beta) = P\left(U_{ijt} > \max_{j' \neq j} U_{ij't}\right)$$

where

$$U_{ijt} = \alpha_j + x'_{ijt}\beta + \varepsilon_{ijt}$$

and

$j' = 1, \ldots, J$ are brands (j represents the chosen one),
i is a consumer,
t is a choice occasion,
x_{ijt} are attributes of brand j at occasion t, and/or individual characteristics and their interactions with brand attributes, and

α and β are parameter vectors capturing the brand specific intercepts and the effects of the attributes respectively.

Different assumptions about the random component ε_{ijt} lead to various models, of which the multinomial logit (MNL) and multinomial probit (MNP) are the best known. Many of the choice models that have been applied in marketing have been developed to overcome the conceptual limitations of the multinomial logit model (i.e., ILA) or the implementation problems of more appealing choice models such as the multinomial probit (i.e., solving high-dimensional integrals).

The major limitation of the multinomial logit model is that its *Independence from Irrelevant Alternatives* (IIA) property leads to an unrealistic pattern of promotion

cross-elasticities, in which each brand draws market shares from every competitor in proportion to its own market share (RUSSELL and BOLTON, 1988). This pattern of brand competition with *proportional-draw* contradicts beliefs that a particular brand competes more closely with certain brands than with others. It has been argued that when IIA does not hold at the aggregate level, this might be caused by unobserved heterogeneity. One of the first attempts to overcome this problem was to incorporate predictors accounting for individual differences in preferences. This was for example, done by using the exponential smoothing of each consumer's previous choices as indicators of his brand preferences (GUADAGNI and LITTLE, 1983), or the consumer's choice shares in a pre-calibration period as a measure of his brand preferences (KRISHNAMURTHI and RAJ, 1988). In the first case, the utility function is re-formulated as:

$$U_{ijt} = \alpha_j + \lambda y_{ijt-1} + (1 - \lambda) S_{ijt-1} + x'_{ijt}\beta + \varepsilon_{ijt}$$

where S_{ijt} is a "brand-loyalty" variable indicating household i's preference for brand j at occasion t. This simple formulation presents a considerable improvement over the estimation of the standard MNL in terms of fit and holdout predictions. However, by using previous choices as an indirect predictor of future choices, this formulation creates a structural dependence in the random component of utility. Unfortunately, existing approaches for estimation of choice models with structural dependence (e.g. HECKMAN, 1981) are still not feasible for the large number of choice occasions and choice alternatives typically found in marketing problems.

Another common practice in marketing for handling the spurious effects of unobserved heterogeneity in preferences is the use of choice models with random-coefficients. Some researchers specified a discrete distribution of the brand intercepts α_j across consumers (JONES and LANDEWEHR, 1988; CHINTAGUNTA, JAIN and VILCASSIM, 1991; GONUL and SRINIVASAN, 1993) while estimating a common set of response coefficients β across all consumers. This "point-mass" semi-parametric approach accounts only for individual differences in brand preferences. Consumers might also differ in how they respond to marketing stimuli such as prices and promotions. In order to account for both these sources of unobserved heterogeneity, some authors extended the MNL model by specifying either a discrete or continuous distribution for all parameters. Several studies in the marketing literature have shown that accounting for unobserved heterogeneity in this manner may result in dramatic improvements of in-sample and predictive fit of choice models (see, e.g. WEDEL and KAMAKURA, 1997).

The latent-class logit model developed by KAMAKURA and RUSSELL (1989) can be viewed as a finite mixture of multinomial logit models, in which each latent-class represents one of the mixture components. The likelihood function for this model is

$$L = \prod_i \sum_s \pi_s \left\{ \prod_t P(y_{ijt} = 1 \mid \alpha_s, \beta_s) \right\}$$

where j is the chosen alternative by household i at occasion t, and s is a latent class with size π_s. A gradient search or the EM algorithm can be used to estimate this model. Applying it to a single source database, KAMAKURA and RUSSELL (1989) show that while the model implies an undesirable proportional draw pattern of brand competition within each segment of consumers, it produces a complex pattern of price and promotion cross-elasticities at the market level, in which brands that are preferred by the same consumers have higher cross-elasticities than those preferred by different segments of consumers. Their results also show a price-tier market structure in which high-price brands are more likely to draw shares from lower-priced brands than vice-versa.

By accounting for unobserved heterogeneity in preferences and response to marketing stimuli, the models discussed so far produce a non-IIA pattern of brand competition at the aggregate level, but still have the undesirable IIA property at the consumer level. The finite-mixture of nested logit models proposed by KAMAKURA, KIM and LEE (1996) avoids this undesirable assumption, by allowing a nested structure of choice alternatives. In their application to single source data, they uncover market segments that differ not only on their preferences for brands of peanut butter and response to price, but also on their choice processes. Some segments have brands nested under the type (crunchy vs. creamy) of peanut butter, some have the types nested under each brand, and others make choices according to a MNL model. With these nested structures, the model avoids the counter-intuitive proportional draw pattern of cross-elasticities even at the consumer level.

The latent-class logit model assumes that preferences and response to marketing stimuli are heterogeneous only to a certain degree so that consumers can be grouped into relatively homogeneous classes. A more extreme account of heterogeneity is made by ROSSI, MCCULLOCH and ALLENBY (1996). These authors assume a unimodal continuous distribution for the parameters of the MNL across consumers. They argue that finite-mixtures lead to an artificial partition of this continuously distributed population into groups that are not necessarily homogeneous. The likelihood for a continuous mixture is

$$L = \prod_i \oint \left\{ \prod_t P(y_{ijt} = 1 \mid \alpha, \beta) \right\} \phi(\alpha, \beta) \partial\alpha \partial\beta$$

where $\phi(\alpha, \beta)$ is the (multivariate normal) distribution of the random coefficients.

For most applications in marketing, this likelihood function would require integration over a multivariate distribution. The authors circumvent this problem by developing hierarchical Bayes choice models with a MNP at the individual level, and a multivariate normal distribution of the MNP parameters across individuals. The authors specify pairs for all (hyper) parameters in the model, and use Markov Chain Monte Carlo methods for estimating the posterior distributions of the parameters. One main advantage of their approach is that aside from estimating the parameters of the random distribution of the MNP parameters, it also produces samples from the

posterior distribution of functions of the parameters for each household. For example, price and promotion cross-elasticities can be estimated for each household in the sample, which can be of great help for micro-marketing purposes. Applying their model to a single-source database of canned tuna, ROSSI et al. (1996) demonstrate how it can be used to identify target households for sales promotions, in order to maximize expected net revenues from the promotion campaign. They show that such targeted campaigns could potentially produce net revenues 155% greater than a simple blanket distribution of coupons.

A compromise between the continuous, unimodal mixtures and the finite-mixtures is proposed by ALLENBY and NIRAJ (1998), who propose a finite-mixture of multivariate normals as the mixing distribution. The likelihood for this discrete-continuous mixture is

$$L = \prod_i \sum_s \pi_s \oint \left\{ \prod_t P(y_{ijt} = 1 \mid \alpha_s, \beta_s) \right\} \phi(\alpha_s, \beta_s) \partial\alpha_s \partial\beta_s$$

Estimation of this complex model is possible in a Bayesian framework with the use of Monte Carlo Markov Chain methods. The authors argue that because it combines finite and continuous mixtures, their model is a latent-class model with heterogeneity within each class. They compare the performance of their model with a finite-mixture of MNL, and show that by allowing for within-class heterogeneity their model produces better in-sample fit and predictive fit to single-source data.

A major impediment for the implementation of these choice models to single-source databases is the magnitude of the typical problem, with dozens of brand-size combinations as choice alternatives. This holds for both the finite-mixture of MNL and the continuous mixture of MNL, with a multivariate normal distribution of the random coefficients. For a random-coefficients MNP model, the number of parameters expands even faster with the number of brands and time periods. A majority of the parameters are required to specify the covariance matrix of the MNL or MNP coefficients. Considerable improvements in parsimony can be attained if a factor structure can be imposed to this covariance matrix. As shown by ELROD (1988), KATAHIRA (1990), ELROD and KEANE (1994) and CHINTAGUNTA (1994), in addition to potentially reducing the number of parameters to be estimated, the factor structure has useful implications for brand positioning. A plot of the factor loadings for the brands produces a map in which the angles between the vectors representing each brand represent the degree of direct competition among those brands. From a managerial point of view such market maps are particularly useful, since they enable the identification of new brand opportunities and the assessment of competitive intensity, amongst others.

2.2 *Purchase incidence, timing, quantity and brand choice*

Although choice models provide valuable insights to managers regarding the impact of their policies on consumers' choice behavior, they do not consider the effect of price

and sales promotion on the incidence or timing of a purchase, nor on the quantity bought. These issues were first considered by GUPTA (1988) who suggested a framework to decompose the sales increase during promotion period into sales increase due to brand switching, due to purchase time acceleration, and due to stockpiling. This is accomplished by developing independent models for inter-purchase time, brand choice and purchase quantity. The timing of purchases is modeled by Gupta as a hazard function using an Erlang-2 model, where the scale parameter of this model is allowed to be a function of household and market specific characteristics (e.g., household inventory, price, promotion). Brand choice is modeled using the GUADAGNI and LITTLE (1983) multinomial logit formulation. Purchase quantity is modeled by an ordered-response logit model using household and market-specific covariates. The models were estimated on scanner panel data for coffee. The results indicated that about 84% of the sales increase due to promotion comes from brand switching, purchase acceleration accounts for about 14% of the sales increase, while stockpiling accounts for less than 2% of the overall sales increase.

As demonstrated by HECKMAN and SINGER (1984), unobserved heterogeneity produces the same spurious effects on models of inter-purchase time as those discussed earlier for choice models. These spurious effects were considered by GUPTA (1991) and JAIN and VILCASSIM (1991). GUPTA (1991) considers the exponential and Erlang-2 hazard functions, with a compound (gamma) distribution for unobserved heterogeneity. If inter-purchase time is distributed Erlang-2 for consumer-i, the likelihood function for this consumer is

$$L_i(t \mid \lambda_i) = \left[\prod_{j=1}^{n_i} \lambda_i^2 t_{ij} \exp(-\lambda_i t_{ij})\right](1 + \lambda_i t_{ic})\exp(-\lambda_i t_{ic})$$

where

n_i = number of complete observations for consumer-i,
t_{ij} = inter-purchase time for jth observation of consumer-i,
t_{ic} = censored observation of consumer-i,
λ_i = scale parameter for Erlang-2 distribution.

Consumer heterogeneity is accounted for by allowing $\lambda_i \sim$Gamma (r, α). Likelihood for any randomly chosen consumer then becomes

$$L(t) = \int_0^\infty L_i(t \mid \lambda_i) g(\lambda_i \mid r, \alpha) \partial \lambda_i$$

Allowing for covariates, such as price and promotion, which change every week, makes the model more complex (see GUPTA, 1991 for details). The results show that an Erlang-2 baseline hazard with Gamma heterogeneity and time-varying covariates produces the best fit and predictive performance when applied to inter-purchase data from a scanner panel. This model accounts for duration dependence, heterogeneity, time-varying covariates, switching patterns among brands being non-stationary, and right-censored data.

Following COX (1972) and HECKMAN and SINGER (1984), JAIN and VILCASSIM (1991) specify a proportional hazard model of interpurchase time. A flexible Box–Cox formulation is used for the baseline hazard. Unobserved heterogeneity is modeled using both parametric and non-parametric distributions. Jain and Vilcassim apply these hazard functions to inter-purchase time data for ground coffee, and show that ignoring unobserved heterogeneity can lead to biased estimation of the baseline hazard and of the impact of price and promotions on inter-purchase times. They also show that the baseline hazard has significant duration dependence, implying the existence of household inventory effects.

Several attempts have been made subsequently to integrate the decisions of whether and when to purchase and of what and how much to buy into a single model. BUCKLIN and LATTIN (1991) propose a two-stage nested logit model for purchase incidence and brand choice. VILCASSIM and JAIN (1991) use a continuous-time semi-Markov approach to analyze purchase time and brand switching decisions in a single framework. In this model, brand switching is captured by a Markov process with finite discrete-state space. However, unlike a Markov chain formulation in which the time between transitions is fixed, this model treats the time between transitions as a random variable that follows some probability distribution.

Descriptions of consumers' decisions of how much to buy in a product category and what brand to choose have been integrated in a single model by DILLON and GUPTA (1996). These authors combine a Poisson model for the category volume decision with a multinomial logit model for the distribution of this volume across brands. Heterogeneity is accounted for through latent classes of loyal and switching consumers. A different Poisson model is used by WEDEL et al. (1995) to study the timing and quantity of purchases for repeat purchases and for brand switches. They propose a discrete time, piecewise-linear exponential hazard model, where the weekly changes in the hazard rate for switching from brand j to brand k are explained by preferences for the brands and by current prices and promotions. Their formulation of the hazard function includes non-proportionality effects for price, while unobserved heterogeneity is considered by a finite-mixture specification.

In order to answer the question of whether, what and how much to buy, CHIANG (1991) develops a discrete/continuous random utility model. A distinctive characteristic of Chiang's model is its theoretical development, based on the assumption that consumers maximize the utility derived both from quantity and perceived quality of the purchased goods, subject to budget and non-negativity constraints. His theoretical model leads to a nested model of purchase incidence and brand choice, as well as a demand system for purchase quantity. CHINTAGUNTA (1993) builds on Chiang's work by using a parametric and a semi-parametric random-effects formulation to account for unobserved heterogeneity.

BUCKLIN, GUPTA and SIDDARTH (1998) develop a model that jointly segments consumers based on their response to price and promotion in the brand choice, purchase incidence, and purchase quantity decisions. Brand choice is modeled by multinomial logit, incidence by nested logit, and quantity by Poisson regression.

Response segments are determined using a finite mixture approach across the three behaviors simultaneously.

In sum, significant progress has been made in modeling (a) consumers' decisions of when, what, and how much to buy, (b) consumers' choice process and market structure, and (c) consumer heterogeneity and segmentation in the analysis of large scanner based data sets. Moreover, the application of these models has led to increased insights in the behavior of consumers, including insights into the choice, timing and quantity decisions of consumers, into what consumer and marketing characteristics affect the purchase probabilities and purchase quantities of brands, and into the existence of segments of consumers that respond differently to marketing actions.

2.3 *Dynamic models of choice*

Most of the models discussed previously focus on the short-term effects of marketing policies on brand performance. Increasingly, researchers as well as managers are concerned about the long-run effects of these policies, and the long term viability of brands as early as during their introduction. In the business world, there is a lot of uncertainty about the long run effects of promotions and the optimal mix of advertising and sales promotions. Many practitioners claim that consistent use of price promotions hurts a brand in the long run, while others contend that this claim is incorrect. Recent research in marketing has begun addressing these long-run issues.

Using aggregate sales and advertising data, DeKimpe and Hanssens (1995a, 1995b) use time series models to address whether there is a persistence of marketing effects on sales. They use unit-root tests to assess if the sales series is stable or evolving over time. If there are no permanent or long run components in the sales series and brand sales fluctuate around a fixed mean-level, then DeKimpe and Hanssens argue that there cannot be any long run effects. Next, a univariate persistence measure determines how much of a brand's long-run performance should be updated when its current performance is lower than expected. Finally, a multivariate persistence measure estimates the sources (e.g., advertising) of this long-run persistence. Using data on 213 sales and marketing mix series from 44 studies published between 1975 and 1994, DeKimpe and Hanssens (1995b) find that 68% of the sales series, 22% of the market share series, and 48% of the marketing mix series show a long-run trend. They also find a higher proportion of sales series showing a long-run trend in the North America (72%) compared with Europe (29%).

Mela, Gupta and Lehmann (1997) argue that offsetting competitive activities may stabilize sales or share series over time, but these activities change consumer behavior in the long run. For example, escalation in price discounting by all major players in an industry may make consumers more price sensitive, yet market shares for brands may not show any major shifts over time. To examine this, Mela et al. use a two-stage modeling approach for a panel data set spanning eight years. In the first stage, latent

class logit models of brand choice are estimated for each quarter of the data, producing quarterly estimates of price sensitivities by segment. In the second stage, they assess the impact of long-run changes in advertising and promotion on these quarterly price sensitivities. Their results show that, in the long-run, advertising makes consumers less price sensitive and also reduces the size of the non-loyal segment, while promotions increase consumers' price sensitivity.

JEDIDI, MELA and GUPTA (1997) extend this approach through a dynamic model of brand choice and purchase quantity with time-varying parameters. Brand choice is modeled by a heteroscedastic, varying-parameter probit, and purchase quantity is modeled by a heteroscedastic, varying-parameter regression with selectivity bias. The model is estimated using panel data of 528 households who made about 11,000 purchases over an eight year period. ERDEM and KEANE (1996) propose a dynamic choice model with forward looking consumers. In this model, usage experience and advertising exposure give consumers noisy signals about brand attributes. Consumers use these signals to update their expectations of brand attributes in a Bayesian fashion. This framework is used to derive brand choice probabilities for both myopic and forward-looking consumers who maximize the expected present value of utility over a finite planning horizon. Thus, only relatively recently have researchers begun to make use of the increasing time dimension of scanner data. The development and application of dynamic models to those data to describe consumer behavior in the long run supports the evaluation of the long-run effects of marketing policies.

3 Customer transaction base analysis

Customer databases are an unusually rich source of longitudinal and cross-sectional information on purchase patterns. Direct marketers have traditionally applied a range of heuristics to predict and manage customer responses. The most common heuristic (with popular acronym "RFM") relies on the *recency*, *frequency* and *monetary value* of past transactions. RFM analysis ranks customers by some measure of their value, providing guidelines for resource allocations over the customer base. RFM is ubiquitous, but its heuristic nature can mislead. The explicit incorporation of substantive insights through formal modeling is important for the following reasons. First, the decisions emerging from the models are potentially applied across millions of customers and have significant implications for future cash flows. Second, direct marketers, and particularly catalog marketers, are uniquely equipped to craft and deliver the marketing mix at the level of the individual customer. The returns to individual level data collection and analysis are greater than those for conventional marketers who use mass media for information dissemination. Third, such formal analysis provides the tools to estimate and manage the cash inflows and outflows at a customer level across time, formalizing the concept of "customer equity". Fourth, unlike traditional marketers who hold significant fixed assets, the core asset of the direct marketer is the inherent value of the customer base. A statistically reliable estimate of this value helps raise working capital in a cash deficient industry.

3.1 *Customer order process modeling*

The specific technique chosen for modeling the customer database depends crucially on the nature of the underlying purchase pattern. SCHMITTLEIN, MORRISON and COLOMBO (1987, SMC henceforth) provide a simple, but widely applicable classification of such patterns. The SMC (1987) classification, with minor adaptations, is displayed in Table 1.

The Type I purchase pattern is typically captured using a model that allows customers to place orders at any time, but also permits customers to 'exit' the re-order process. Such exit is not explicitly signaled, and must be inferred from the pattern of purchases. A model of the Type I process typically involves an assumption about the individual customer's purchase and exit processes and an assumption about the underlying heterogeneity in purchase rates and exit rates across customers. For example, SMC (1987) present a NBD/Pareto model of the Type I process, where individual customers make Poisson purchases with rate l and have exponential lifetimes with death (exit) rate m. These parameters are assumed independent and each is gamma distributed across the population. A priori, there is little reason to postulate more structure and the choice of flexible distribution forms allows the data to speak for itself. SCHMITTLEIN and PETERSON (1994) provide an empirical application of the NBD to a customer base of an office supplies firm. Note here that to convert a probabilistic model of customer purchases patterns into a model of expected cash flows, the expectations from the estimated purchase pattern model must be integrated with a separate estimation of expected cash flows.

The Type I process restricts the role of the firm in actively soliciting orders, and there is little emphasis on the role of costs internal to the firm. The case of a catalog marketing firm is different in that orders are, by and large, *primed* by catalogs. Catalog marketing, therefore, can be considered a Type III process, where the decision to mail a catalog can be coordinated with either a response or a non-response. One can reasonably expect that, in the absence of catalogs, orders from a particular customer cease over time. The interesting aspect here is that the catalog marketer can "force" death (exit) by deciding to drop a customer from an active mailing list.

Table 1. Purchase pattern classification (adapted from SMC 1987)

Dimension 1: Opportunities for Transactions	Dimension 2: Time at which Customers become Inactive	Examples of Products and Services
Continuous/Unobserved	Unobserved	TYPE I (e.g., Brokerage transactions, Office supplies)
	Observed	TYPE II (e.g., Telephone service)
Discrete/Observed	Unobserved	TYPE III (e.g., Catalogs, TV home-shopping)
	Observed	TYPE IV (e.g., Magazine subscriptions, Insurance)

In the context of catalog marketing, BALASUBRAMANIAN (1997) defines the concepts of "passive" and "managed" equity. While some customers (e.g., industrial clients) are *potentially* capable of "living" infinitely, in reality they are very likely to drop out within a finite time horizon. Consider a catalog marketing firm that discounts future earnings with discount rate β and receives a constant percentage k of order size θ, as contribution. The firm mails catalogs at a marginal cost c per unit, which includes production and mailing expenses. Consider first the case where the firm mails catalogs to customers at intervals $t = 1, 2, \ldots, \infty$. The "passive" equity of a customer, who exits the purchase process (or "dies") at time T_d, is defined as:

$$E_p = \sum_{t=1}^{\infty} \beta^{t-1}(k\theta_t - c)$$

where $(\theta_t = 0 \forall t \geq T_d)$.

In contrast, a firm that actively manages customer equity may "force" a death at time T_f by taking a customer off the active mailing list. In this case, the "managed" equity of the customer is defined by:

$$E_m = \sum_{t=1}^{\infty} \beta^{t-1}(k\theta_t - c)$$

where $(\theta_t = 0 \forall t \geq T_d)$ and $((\theta_t = 0, c = 0) \forall t \geq T_f)$.

The importance of managing customer equity is clear from a straightforward comparison of the two equations above. The customer can be thought of as an *investment opportunity*. The firm "invests" in the customer by mailing a catalog. A firm that keeps mailing catalogs beyond T_d incurs a marginal cost c per mailing, with no purchases made by the customer. Note that the firm may decide to stop mailing to a customer if the expected contribution is too low, *even* if the customer is alive. Customer equity, therefore, integrates the response rates of the customer, the exit process, the dollar order volume, the contribution percentage, the operational costs of the firm for catalog production and mailing *and* the policies adopted by the firm in deciding who remains on the active mailing list. The task of statistical analysis of the customer base in this context is two-fold. First, using existing records of customer purchases, projections of future purchase frequencies need to be made. Second, the cost of information dissemination must be integrated into the decision to retain/drop customers from the mailing list with the overall objective of maximizing the net present value of the customer base. BALASUBRAMANIAN and SCHMITTLEIN (1997) present an economic model of the catalog marketing process. The model combines an opportunity-based version of the SMC/Pareto model (SMC, 1987) with a dynamic programming framework to indicate appropriate mailing decisions and estimate customer equity.

Researchers have used a variety of models to capture customer response to direct mail. These approaches, which at times combine heuristic procedures and formal

modeling, include Automatic Interaction Detection (AID) (Sonquist, 1970), CHi-squared Automatic Interaction Detection (CHAID), logit and log–linear models (Bult, van der Scheer and Wansbeek, 1997), finite mixtures of Poissons (Wedel, DeSarbo, Bult and Ramaswamy, 1993), semi-parametric classification approaches (Bult, 1993), and other methods. The reader is referred to e.g. Magidson (1988) and Bult (1993) for some comparative details.

A few recent papers have addressed the mailing decisions of the catalog marketing firm. Rao and Steckel (1995), in seeking to analyze the selection and evaluation of direct mail prospects, assume that the response probability of customer i to a mailing is distributed Beta(a, b) across the customer base. The parameters are exponential functions of descriptor variables. The beta provides a flexible prior distribution, and accommodates a Bayes updating of individual level estimates as information on a particular customer builds up over time. The longer the record of the individual customer, the more that information is weighted relative to the information in the rest of the database. Broadly speaking, over time, such updating shifts the informational emphasis from the cross-sectional breadth of the database to the longitudinal length at the level of individual customers. The idea of a ''large data set'' in marketing often seems to invoke stronger allusions to the number of customers in the database, rather than the length of the transactional history.

Bult and Wansbeek (1995) present an approach to determine target selection for mailings. They capture the relationship between the fraction of the customer base receiving mailings and the profits derived from the fraction responding to them. An econometric analysis of the relationship between profits and response probability reveals the optimal set of targets to be selected. Bitran and Mondschein (1996) model mailing and inventory policies under a capital availability constraint. Gönül, Kim and She (1996) and Gönül and She (1996) investigate the timing, frequency and spacing of the catalog or comparable stimulus. They propose that the mailing decision should not rest on whether it generates absolute profit, but rather on a comparison of profits between the mail/no-mail cases. The underlying theme is that customers have a baseline propensity to order even in the absence of the catalog. In addition, researchers have studied specific issues relating to the operations of direct marketing firms. For example, Morwitz and Schmittlein (1996) present a model that complements managerial insights in designing roll-out tests for direct marketers. Basu, Basu and Batra (1995) study the pattern of response to a direct marketing offering over time. Early indications of incoming demand can help managers plan the appropriate level of inventory holding.

In summary, the Type I and Type III scenarios from Table 1 are most amenable to the models discussed here. Type II and Type IV scenarios often involve direct marketing mechanisms for customer acquisition and retention. These scenarios, however, are probably best tackled from the viewpoint of customer satisfaction/dissatisfaction (e.g., Oliver, 1980), evaluation of service and performance levels (e.g., Bolton and Drew, 1991) and the adoption/disadoption process (e.g., Lemon, Barnett and Winer, 1997).

A serious concern with consumer databases of direct marketers is the lack of information on the context of choice. Little is known from the database about whether consumers are choosing between direct marketers, or between conventional retailers and direct marketers. In the emerging multiple channel environment, the design and implementation of databases that merge transactional data with data on the competitive context could be the next frontier in direct marketing.

4 Discussion and conclusion

Many of the models described above for single source and customer transaction databases are special cases or extensions of the generalized linear model (McCullagh and Nelder, 1989). Let $i = 1, \ldots, I$ denote individuals, and $j = 1, \ldots, J$ brands and $t = 1, \ldots, T$ time. The dependent variable of interest, y_{ijt}, is distributed according to a member of the exponential family $p(y_{ijt} \mid \mu_{ijt}, \sigma)$, with expectation $E(y_{ijt} \mid x_{ijt}) = \mu_{ijt}$, conditional upon a set of P covariates, x_{ijt} (and possibly with dispersion parameter σ). The expectation is related to the covariates through: $g(\mu_{ijt}) = x'_{ijt}\beta$, with $g(\bullet)$ a link function, and β a $(P \times 1)$ vector of parameters, that may in addition vary over j and/or t. Based on the above review, we identify a number of issues that are becoming increasingly important as the sizes of databases further expands, in the direction of either of the modes: individuals, brands and time.

4.1 *Increase in the number of parameters estimated*

Most of the models described are estimated with iterative search algorithms to maximize some function of the parameters, most importantly likelihood functions, and related functions such as simulated, quasi, pseudo, and conditional likelihood (cf. Lindsey, 1996). For the generalized linear model, assuming independence across i, the likelihood is:

$$L(\beta, \sigma \mid y) = \prod_{i=1}^{I} p(y_i \mid \mu_i, \sigma)$$

Algorithms that are commonly applied include search algorithms such as Newton and Newton-like algorithms, linear programming, the EM-algorithm and its variants (cf. Thisted, 1988), and more recently, simulated annealing and genetic algorithms. In general, these methods exhibit linear (EM) or quadratic (Newton-Raphson) convergence. Even quadratic convergence may be problematic for large data sets, especially if the number of parameters estimated increases with one or more of the dimensions of the data set. This occurs (a) when parameters are to be estimated for each subject to account for heterogeneity (i.e. the dimension of β depends on i), (b) in models that capture time dependence in long time series of scanner data (β depends on t) and (c) for large numbers of choice alternatives (β depends on j). An advantage of the expanding size of databases is that asymptotic properties of the estimates are obtained for either $i \to \infty$ if β is constant across i, for $t \to \infty$ if β is constant across t,

or for $i \to \infty$ and $t \to \infty$. In large data sets either of these conditions may hold approximately.

On the one hand, the increase in the number of parameters has been met by the development of more parsimonious models. For example, instead of models that include a parameter for each consumer i to account for heterogeneity, continuous or discrete heterogeneity distributions have been assumed. In the former case, it is usually assumed that $\beta_i = N(\beta, \Sigma)$ is multivariate normal. These models have gained increased attention with the advent of Markov Chain Monte Carlo estimation methods in Bayesian statistics that enable the approximation of the integrals involved. Here, imposing hyper-parameter distributions on β enables simulation of the posterior distributions of individual level parameters $p(\beta_i/y)$ that are of much interest in particular in direct marketing, where such estimates enable firms to take the individual customer as the focus of their marketing action. The assumption of a discrete heterogeneity distribution leads to finite mixture models, where for s unobserved mixture components in proportions π_s:

$$p(y_{ij} \mid \beta, \sigma, \pi) = \sum_{s=1}^{S} \pi_s p_s(y_{ij} \mid \beta_s, \sigma_s)$$

Finite mixture models have gained great popularity in marketing because they connect to the important concept of market segmentation (cf. WEDEL and KAMAKURA, 1997). As mentioned above, recent approaches combine discrete and continuous distributions of heterogeneity.

To describe time dependence, researchers have either included functions of past observations on the dependent variable among the predictors, or assumed the coefficient vector to vary over time: $\beta_t = N(\beta, \Sigma)$, where examples are provided by time-series and dynamic choice models. While the former approach in general has undesirable statistical properties, the latter approach has given rise to problems of solving high dimensional integrals in the likelihood function. Those problems have been solved through the application of simulation methods, but these models may require the evaluation of prohibitively high dimensional integrals if either the size of the choice set or the time domain expands. Such dimensional problems have been avoided by imposing factor-structures on the covariance matrix (cf. ELROD and KEANE, 1995). The issue of how to model markets with many items (i.e. brand-size combinations) is still largely unexplored. Usually researchers take the approach of either selecting a potentially relevant subset of the available items, or aggregating brand-size combinations into aggregate categories. A recent approach to modeling competition in retail scanning data with many items, based on nested models, was proposed by FOEKENS, LEEFLANG and WITTINK (1997).

4.2 *Increase in computational requirements*

Unfortunately, even the estimation of the described parsimonious models may require very long computation times for large data sets. Especially in marketing research

practice, where models have to be applied routinely for a large number of product categories, the computational problems posed by large data sets has led to renewed interest in estimation procedures. For example, conjugate gradient techniques can be used to speed up convergence of the EM algorithm in the estimation of mixture models and factor analysis (cf. McLachlan and Krishnan, 1997). Evolutionary computation does not overcome the problem of computational burden imposed by large data sets, rather the opposite holds true. Nevertheless, these techniques do have their merits, in particular in the optimization of highly complex objective functions to find the parameter estimates of assumed statistical models. In the future, much of the worth of those techniques may come from integrating them with existing statistical methodology (cf. Barndorff-Nielsen, J. L. Jensen and Kendall, 1993). The accumulated experience in modeling marketing phenomena thus would carry over to the application of such evolutionary computation methods, which would lend them the theoretical basis of interpreting the estimated parameter values, rather than seeing them as black-box approaches.

Alternatively, simple models and estimation methods for which likelihood or moment estimators have closed form expressions are gaining popularity. An example is provided by Sikkel and Hoogendoorn (1995) who estimate a variety of models of purchase incidence and purchase timing based on the Poisson, zero-inflated Poisson and negative binomial distributions using methods of moments. As an alternative solution, researchers have aggregated databases across one or more of the modes before analysis. Aggregating across consumers, for example leads to vast reductions in data, which then lend themselves to estimation by time-series methods, for example (cf. DeKimpe and Hanssens, 1995a,b; Franses, 1994). However, the effects of aggregating data where the underlying response process is expected to be nonlinear has just begun to be explored (cf. Foekens, 1995).

4.3 *Conclusion*

As the size of data sets expands, information on the population (of all customers) rather than on a sample is often obtained, especially for customer transaction databases. In such cases, one can no longer start from the usual assumption that the model defined by $g(\mu_{ijt})$ is true in the population. At best, one may consider it as a "working" model, which is not necessarily correct. (Alternatively, one may assume that the data are a sample from the population of all time points or from possible outcomes of a random variable subject to measurement error.) Often, samples are drawn from the customer database for the purpose of analysis and model building. Truly random samples may be drawn and repeated samples are available for model testing and predictive validation. The interest in statistical approaches to estimating models based on maximizing goodness of fit seems to decline somewhat due to the availability of data on each individual. Increasingly, researchers recognize that firms need to act not on the basis of statistical fit of some response model, but on the basis of the profit that results from the response. Throughout the literature, and particular in

the area of modeling customer databases, approaches are emerging where instead of maximizing statistical fit of the response model, profit functions are formulated that embeds costs and benefits to the decision maker and explicitly maximized. Typically, expected profit functions take the form:

$$\Pi = \oint (\Sigma p(y \mid \mu, \sigma) P y - c) \mathrm{d}y$$

with P price per unit y and c marginal costs (cf. BULT and WANSBEEK, 1995). Maximizing such profit functions requires non-traditional maximization approaches, such as smoothed maximum score (BULT and WITTINK, 1996).

An additional opportunity provided by the increase in amounts of data, in particular in the area of customer transaction databases, is that of data-fusion. Whereas not all data are typically available in a single database collected on a population of customers, combining databases from several sources greatly increases their potential for marketing purposes. Whereas previous work in the statistics literature has primarily dealt with concatenation of files with a small to moderate number of variables, KAMAKURA and WEDEL (1997) proposed (finite mixture based) models for the concatenation databases with many variables. Other issues that require increased attention in modeling large scanner databases are problems of measurement error and outliers. Those issues have recently attracted attention. FRANSES, KLOEK and LUCAS (1998) have developed robust estimation approaches to deal with outliers in time series models of aggregate level scanner data.

To summarize, the sheer size of available data places severe demands on computing power. While the capacity of computers in terms of speed and storage has increased at a very high rate during the last two decades, the size of marketing databases seems to have increased even faster. Envisioning the databases as a three-way matrix of subjects-by-brands-by-time, the size of the databases has been expanding in each of those dimensions. This has revived interest in speeding up traditional numerical and statistical estimation methods and for estimators that take simple closed forms, while statistical and econometric models for which the parameter expands with any of those dimensions have started to lose much of their appeal. Parsimonious models that entail a compression of the data in either direction have received great interest. In addition, issues of how to aggregate data in a meaningful way by collapsing over certain entries in the databases has become an important problem, along with the study of the effects of aggregation on parameter estimates. The increasing availability of data also increases the importance of more efficient use of those data by combining data sources, for example through data-fusion. In addition, the data collected nowadays often pertain to each individual customer in a population, instead of a mere sample, which changes the nature of statistical inference. Such individual level data on populations of customers enable the formulation and optimization of profit functions at the individual level, rather than maximizing the statistical fit of the models. However, with the expanding size of the databases, problems of control over measurement error and missing observations have seemed to increase, while efficient solutions provided in the

areas of multiple imputations and MCMC estimation often do not yet lend themselves to routine application in practice. Although problems related to the analysis of large databases in marketing have begun to be recognized, and a literature has begun to emerge, the problems seem to be far from resolved, and are expected to stimulate new research avenues in the near future.

5. References

ALLENBY, G. and J. GINTER and NEERAJ ARORA (1998), On the heterogeneity of demand, *Journal of Marketing Research*, forthcoming.

BALASUBRAMANIAN, S. (1997), Two essays in direct marketing, Ph.D. thesis, Yale University.

BALASUBRAMANIAN, S. and D. C. SCHMITTLEIN (1997), Customer equity and relationship management for a catalog marketing firm, Working Paper, University of Texas at Austin.

BARNDORFF-NIELSEN, O. E., J. L. JENSEN and W. S. KENDALL (1993), Networks and chaos-statistical and probabilistic aspects, Chapman and Hall, London.

BASU, A. K., A. K. BASU and R. BATRA (1995), Modeling the response pattern to direct marketing campaigns, *Journal of Marketing Research* **32**, 204–212.

BERRY, M. J. A. and G. D. LINOFF (1997), Data mining techniques: for marketing sales and customer support. Wiley, New York.

BITRAN, G. R. and S. V. MONDSCHEIN (1996), Mailing decisions in the catalog sales industry, *Management Science* **42**, 1364–1381.

BOLTON, R. N. and J. H. DREW (1991), A multistage model of customers' assessments of service quality and value, *Journal of Consumer Research* **17**, 375–384.

BUCKLIN, R. E. and J. M. LATTIN (1991), A two-state model of purchase incidence and brand choice, *Marketing Science* **10**, 24–39.

BUCKLIN, R., S. GUPTA and S. SIDDARTH (1998), Determining segmentation in sales response across consumer purchase behaviors, *Journal of Marketing Research*, forthcoming.

BULT, J. R. and T. WANSBEEK (1995), Optimal selection for direct mail, *Marketing Science* **14**, 378–394.

BULT, J. R. (1993), Semiparametric versus parametric classification models: an application to direct marketing, *Journal of Marketing Research* **30**, 380–390.

BULT, J. R., H. VAN DER SCHEER and T. WANSBEEK (1997), Interaction between target and mailing characteristics in direct marketing, with an application to health care fund raising, *International Journal for Research in Marketing* **14**, 301–308.

CHIANG, J. (1991), A simultaneous approach to the whether, what and how much to buy questions, *Marketing Science* **10**, 297–315.

CHINTAGUNTA, P. K. (1993), Estimating a multinational probit model of brand choice using the method of simulated moments, *Marketing Science* **11**, 386–407.

CHINTAGUNTA, P. K. (1994), Heterogeneous logit implications for brand positioning, *Journal of Marketing Research* **31**, 304–311.

CHINTAGUNTA, P. K., D. C. JAIN and N. J. VILCASSIM (1991), Investigating heterogeneity in brand preferences in logit models for panel data, *Journal of Marketing Research* **28**, 412–28.

COOPER, L. G. and M. NAKANISHI (1988), *Market share analysis*, Dordrecht, Kluwer.

COX, D. R. (1972), Regression models and life-tables, *Journal of Royal Statistical Society B* **34**, 187–200.

DEKIMPE, M. and D. HANSSENS (1995a), The persistence of marketing effects on sales, *Marketing Science* **14**, 1–21.

DEKIMPE, M. and D. HANSSENS (1995b), Empirical generalizations about market evolution and stationarity, *Marketing Science* **14**, part 2, G109–G121.

DILLON, W. R. and S. GUPTA (1996), A segment-level model of category volume and brand choice, *Marketing Science* **15**, 38–59.

ELIASHBERG, J. and G. L. LILIEN (1993), *Marketing*, Handbooks in operations research and management science, Amsterdam, Elsevier.

ELROD, T. (1988), Choice map: inferring a product-market map from panel data, *Marketing Science* **7**, 21–40.
ELROD, T. and M. P. KEANE (1995), A factor-analytic probit model for representing the market structure in panel data, *Journal of Marketing Research* **32**, 1–16.
ERDEM, T. and M. KEANE (1996), Decision-making under uncertainty: capturing dynamic brand choice processes in turbulent consumer goods markets, *Marketing Science* **15**, 1–20.
FAHRMEIER, L. and G. TUTZ (1991), *Multivariate statistical modeling based on generalized linear models*, Springer Verlag, New York.
FOEKENS, E. W., P. S. H. LEEFLANG and D. R. WITTINK (1997), Hierarchical versus other market share models for markets with many items, *International Journal for Research in Marketing* **14**, 359–378.
FOEKENS, E. W. (1995), Scanner data based modeling: empirical applications, Unpublished Ph.D. Thesis, University of Groningen, The Netherlands.
FRANSES, P. H. (1994), Modeling new product sales; an application of cointegration analysis, *International Journal for Research in Marketing* **11**, 491–502.
FRANSES, P. H., T. KLOEK and A. LUCAS (1998), Outlier robust analysis of long-run marketing effects for weekly scanning data, *Journal of Econometrics*, forthcoming.
GONUL, F. and K. SRINIVASAN (1993), Modeling multiple sources of heterogeneity in multinomial logit models: methodological and managerial issues, *Marketing Science* **12**, 213–29.
GONUL, F., B.-D. KIM and M. SHI (1996), Optimal mailing policy for catalog customers, Working Paper, Carnegie Mellon University.
GONUL, F. and M. SHI (1996), Optimal timing and spacing of newsletters: a new methodology using estimable structural dynamic programming models, Working Paper, Carnegie Mellon University.
GUADAGNI, P. and J. LITTLE (1983), A logit model of brand choice, *Marketing Science* **2**, 203–38.
GUPTA, S. (1988), Impact of sales promotion on when, what, and how much to buy, *Journal of Marketing Research* **25**, 342–56.
GUPTA, S. (1991), Stochastic models of interpurchase time with time-dependent covariates, *Journal of Marketing Research* **28**, 1–15.
HECKMAN, J. J. (1981), The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process, in: C. F. MANSKI and D. MACFADDEN (eds.), *Structural analysis of discrete data with econometric applications*, MIT-Press, Cambridge, MA, 179–95.
HECKMAN, J. J. and B. SINGER (1984), A method for minimizing the impact of distributional assumptions in econometric models for duration data, *Econometrica* **52**, 271–320.
JAIN, D. and N. VILCASSIM (1991), Investigating household purchase timing decisions: a conditional hazard function approach, *Marketing Science* **10**, 1–23.
JEDIDI, K., C. MELA and S. GUPTA (1997), The dynamic effects of advertising and promotions on brand choice and purchase quantity, Working Paper, Columbia University.
JONES, J. M. and J. T. LANDEWEHR (1988), Removing heterogeneity bias from logit model estimation, *Marketing Science* **7**, 41–59.
KAMAKURA, W. A. and G. J. RUSSELL (1989), A probabilistic choice model for market segmentation and elasticity structure, *Journal of Marketing Research* **26**, 379–390.
KAMAKURA, W. A., B. KIM and J. LEE (1996), Modeling proference and structural heterogeneity, *Marketing Science* **15**, 152–72.
KAMAKURA, W. A. and M. WEDEL (1997), Statistical data-fusion for cross-tabulation, *Journal of Marketing Research* **34**, 485–498.
KATAHIRA, H. (1990), Joint space market response analysis, *Marketing Science* **36**, 13–27.
KRISHNAMURTHI, L. and S. P. RAJ (1988), A model of brand choice and purchase quantity price sensitivities, *Marketing Science* **7**, 1–20.
LEMON, K. N., T. BARNETT and R. S. WINER (1997), The challenge of letting go: the effect of the past, present and future on the consumer's keep or dispose decision, Working Paper, Duke University.

LINDSEY, J. K. (1996), *Parametric statistical inference*, Clarendon Press, Oxford.
MAGIDSON, J. (1988), Improved statistical techniques for response modeling, *Journal of Direct Marketing* **2**, 6–18.
MCLACHLAN, G. J. and T. KRISHNAN (1997), *The EM algorithm and extensions*, Wiley, New York.
MELA, C., S. GUPTA and D. LEHMANN (1997), The long-term impact of promotion and advertising on consumer brand choice, *Journal of Marketing Research* **34**, 248–261.
MORWITZ, V. G. and D. C. SCHMITTLEIN (1997), Testing new direct marketing offerings: the interplay of models and management judgment, *Management Science*, forthcoming.
OLIVER, R. L. (1980), A Cognitive model of the antecedents and consequences of satisfaction decisions, *Journal of Marketing Research* **17**, 460–469.
RAO, V. R. and J. H. STECKEL (1995), Selecting, evaluating and updating prospects in direct mail marketing, *Journal of Direct Marketing* **9**, 20–31.
ROSSI, P. E., R. E. MCCULLOCH and G. M. ALLENBY (1996), The value of purchase history data in target marketing, *Marketing Science* **15**, 321–340.
RUSSELL, G. J. and R. BOLTON (1988), Implications of market structure for elasticity structure, *Journal of Marketing Research* **25**, 229–241.
SCHMITTLEIN, D. C., D. G. MORRISON and R. COLOMBO (1987), Counting your customers: who are they and what will they do next?, *Management Science* **33**, 1–24.
SCHMITTLEIN, D. C. and R. A. PETERSON (1994), Customer base analysis: an industrial purchase process application, *Marketing Science* **13**, 41–67.
SIKKEL, D. and A. W. HOOGENDOORN (1995), Models for monthly penetrations with incomplete panel data, *Statistica Neerlandica* **49**, 378–391.
SKINNER, C. J., D. HOLT and T. M. F. SMITH (1989), *Analysis of complex surveys*, Wiley, New York.
SONQUIST, J. N. (1970), *Multivariate model building*, Institute for Social Research, Ann Arbor, MI.
STECKEL, J. H. and W. R. VANHONACKER (1988), A heterogeneous conditional logit model of choice, *Journal of Business and Economic Statistics* **6**, 391–398.
THISTED, R. A. (1988), *Elements of statistical computing*, Chapman and Hall, New York.
VILCASSIM, N. and D. JAIN (1991), Modeling purchase-timing and brand-switching behavior incorporating explanatory variables and unobserved heterogeneity, *Journal of Marketing Research* **28**, 29–41.
WEDEL, M. and W. A. KAMAKURA (1997), *Market segmentation: conceptual and methodological foundations*, Kluwer, Dordrecht.
WEDEL, M., W. S. DESARBO, J. R. BULT and V. RAMASWAMY (1993), Latent class Poisson regression model for heterogeneous count data with an application to direct mail, *Journal of Applied Econometrics* **8**, 397–411.
WEDEL, M., W. A. KAMAKURA, W. S. DESARBO and F. TER HOFSTEDE (1995), Implications for asymmetry, nonproportionality and heterogeneity in brand switching from piece-wise exponential hazard models, *Journal of Marketing Research* **32**, 457–62.