ELSEVIER

Intern. J. of Research in Marketing 17 (2000) 203–208

International Journal of Research in Marketing

www.elsevier.com/locate/ijresmar

Marketing data, models and decisions

Michel Wedel [a,b,*], Wagner Kamakura [c], Ulf Böckenholt [a,d]

[a] Faculty of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, Netherlands
[b] University of Michigan, MI, USA
[c] College of Business Administration, University of Iowa, Iowa City, IA 52242, USA
[d] Department of Psychology, University of Illinois, Urbana-Champaign, IL 61820, USA

Abstract

Our comments about the paper by Leeflang and Wittink [Internat. J. Res. Marketing, 17 (2000) 105] comprise of two components: first, we address two issues on which we disagree with Leeflang and Wittink: soft versus hard data, and individual-level versus segment-level models. Secondly, we supplement their paper by attempting to predict how marketing model building will change in the future.

Keywords: Mixture models; Hierarchical Bayes models; Soft data; Transaction databases; Integrated marketing models

1. Introduction

Leeflang and Wittink (2000) provide a concise overview of past and present trends in model building. To add to such a comprehensive presentation is not a trivial task since we agree with many of their observations. Our comments focus on two aspects that are of much importance in empirical applications and theoretical developments of marketing models. We address two issues on which our views differ from Leeflang and Wittink (2000) — soft versus hard data, and individual-level versus segment-level models. Then we predict trends in modeling with a focus on customer transaction databases and supplement the Leeflang and Wittink (2000) paper with our own vision of how marketing model building will change in the future.

2. Hard versus soft data

Leeflang and Wittink (2000) posit normatively several conjectures on the preferred data for marketing modeling. They state, for example that “Models of past actual outcomes can overcome the biases that are inherent in subjective judgments. [...] models of actual outcomes are favored over models of judgments for the prediction of future occurrences”, (Section 3.5, point (1)). Leeflang and Wittink’s (2000) objection to soft judgmental data seems to derive from the fact that “subjective judgments are subject to numerous biases such as prominence effects, anchoring effects and overconfidence”, (Section 4.1). We do not share their view. In our opinion, Leeflang and Wittink (2000) too lightly sidestep the issue that biases are present in “hard data”. Scanner data are problematic for numerous reasons including effects

* Corresponding author. Tel.: +31-50-3633735; fax: +31-50-3633720.
E-mail address: m.wedel@eco.rug.nl (M. Wedel).

0167-8116/00/$ - see front matter
PII: S0167-8116(00)00010-0

of aggregation, incompleteness in the scanning process, selection of heavy users in analyses, missing price and sales information, and lack of variations in price and other marketing control variables. Similarly, click-stream data are difficult to deal with because it is often unclear for what populations the customer samples are representative and there is little control over web page choice and visit times. However, rather than being reasons to discard any of these data sources, we feel that they represent research opportunities in Marketing Science that may aim to understand such biases and devise technologies that reduce their impact.

2.1. Judgmental bias

Leeflang and Wittink's (2000) low opinion about the quality of subjective data is reflective of recent research paradigms in economics and psychology that focus on "traps" and "mistakes" in human decision making (Russo and Shoemaker, 1990). Much of this work attempts to detect and describe judgmental biases but does not provide remedies or explanations for the causes of the studied phenomena. In our view, the notion that the human mind is burdened with numerous detectable but difficult-to-explain anomalies and biases is not constructive. For example, it seems odd to attribute the lack of explanatory power of linear and compensatory models of human attitudes (preference, satisfaction, etc.) to biases in the decision-making process rather than to the incompleteness of the models. Rather than dismissing anchoring, assimilation, contrast and effects of emotions on decisions as "biases", it is more useful to embrace these phenomena as potentially useful adaptations that have evolved to improve human decision making by rendering it more effective or efficient. There is a strong need for future research that includes those effects into models of human judgments.

Consider, for example, the progress that has been made in social psychology on the conditions under which assimilation and contrast effects occur in surveys, as caused by question sequencing, wording and scale usage (e.g., Sirken et al., 1999). Many of the misconceptualizations of biases in judgmental data seem to derive from an incorrect "file drawer" model of human attitudes. For a long time, research in psychology and marketing has assumed that consumers hold stable attitudes that they simply retrieve from memory when needed. Instead, recent evidence suggests that consumers may form evaluative judgments "on the spot" based on information provided by the decision context. This psychological process, which may have evolved to facilitate human interactions, leads to strong context (assimilation and contrast) effects in responses to survey questions.

We conjecture that contrast effects are particularly prominent in satisfaction, conjoint and MDS studies (Kim et al., 1999). In particular, the extent to which attitudes and behavior converge is a function of the degree to which the decision context agrees with the context in which attitudes are assessed. This extends to Leeflang and Wittink's (2000) proposal to improve prediction by using past behavior in predicting current behavior, since the predictive value of past behavior is limited by the comparability of the contexts in which the decisions/judgments were made. Since these contexts are unobserved, as it is the case for scanner data, much care is needed in correctly modeling this source of heterogeneity. To be clear, we do not want to conclude that consumers never hold stable attitudes or values. However, we believe that there is a need to develop models that explicitly distinguish between consumers that evaluate products and services on the spot or by memory retrieval.

2.2. Integration of data sources

To summarize, a deeper understanding of both the consumer decision process and context effects interacting with these decisions is needed. This applies to data-gathering contexts provided by surveys, experimental contexts such as the ones occurring in conjoint and in eye-tracking studies, new media contexts provided by, e.g., the Internet, and also to actual decision making contexts such as retail outlets, etc. The effects identified need to be included in models for judgmental data. To this end, a much closer collaboration between researchers working in the currently disparate fields of consumer behavior and quantitative modeling is required. We agree with Leeflang and Wittink (2000) that data sources need to be integrated, such that in a single study comparative advantages of several data sources can be exploited simultaneously.

3. Segment- versus individual-level models

We do not contest that the Internet facilitates customization of marketing across households (Leeflang and Wittink, 2000, Section 4.3), but it seems unnecessarily restrictive to take the individual-level approach as a central paradigm for marketing and thus marketing modeling in the future, as Leeflang and Wittink (2000) seem to do. It has been argued that with the advent of direct marketing, micro marketing, database marketing and e-commerce individual-level response parameters need to be estimated for the optimal implementation of marketing strategies. Leeflang and Wittink (2000) provide examples where the individualization paradigm has been successful. However, another trend should not remain unnoticed: companies such as Proctor and Gamble and Unilever have scaled down product assortments to target larger segments with a much more limited variety of products. Even direct marketing and catalog selling companies do not tailor their marketing effort to each customer specifically, but use a finite set of marketing stimuli, i.e., letters, folders, catalogues, to target all prospects. Although it is now possible to reach customers individually through mass production and direct media, companies may choose not to do so.

Cost is one reason for deciding not to pursue one-to-one marketing. Although the Internet has reduced the costs of targeting customers individually, those costs are not negligible and will not become so in the near future. The availability of new media has reduced marketing costs, but when there are scale advantages in production, distribution or advertising, individual-level approaches may still not be profitable and segmentation strategies may be called for (Wedel and Kamakura, 1999).

3.1. Heterogeneity in marketing models and Bayesian inference

The above mentioned developments in practice have profound implications for accommodating customer heterogeneity in marketing models. Heterogeneity has been accommodated in marketing models through either continuous or discrete mixing distributions of parameters (Wedel et al., 1999). Discrete mixing distributions are elegantly associated with market segmentation and continuous heterogeneity distributions with direct and one-to-one marketing. Some have argued that the assumption of a limited number of homogeneous segments in the finite mixture models that implement discrete heterogeneity is overly restrictive (Allenby and Rossi, 1999). To those authors, market segmentation yields an artificial partition of an underlying continuous distribution into discrete segments. In addition, discrete mixing distributions (with an incorrectly specified number of components) may lead to inconsistent parameter estimates.

The continuous heterogeneity distribution is often equated with hierarchical Bayes (HB) models. Bayesian statistics have gained popularity predominantly because of the Markov Chain Monte Carlo (MCMC) sampling techniques that allow for the analysis of complex models (Gelman et al., 1995). As an applied science, rather than isolating a small component of a substantive problem for which a model can be more easily constructed, Marketing has a history of modeling complex substantive problems holistically by compounding a series of simple (e.g., OLS) heuristic estimation procedures. Recently, however, researchers have discovered the potential of MCMC for estimating such complex models, by allowing the full posterior distribution of parameters to be split up into a series of full-conditional distributions that can be iteratively cycled through. However, although the application of the Bayesian estimation machinery to analyze problems that could not be solved otherwise is an important asset in marketing, that motivation is a pragmatic one. Although recent advances in MCMC may provide Bayesians with an edge in estimating complex models, non-Bayesian methods, such as simulated likelihood and stochastic EM are rapidly closing the performance gap.

Data are bound to violate assumptions of any model to a certain extent, since no model can be perfectly true. Empirically, the parameter distribution may not be normal and segments may not be completely homogeneous. The identification of a heterogeneity distribution from empirical data is an ill-posed problem that is further complicated by the confounding of heterogeneity with state dependence (Ailawadi et al., 1999). It is therefore unlikely that either theoretical or empirical evidence, even if state dependence is appropriately accounted for, will ever decisively favor one representation over the other. Therefore, we advocate the specification of models that include both continuous and discrete heterogeneity. This approach overcomes the standard assumptions in HB modeling of unimodality and of normal heterogeneity distributions. An attractive feature of these models derives from the fact that strictly discrete or continuous heterogeneity models are special cases of the "full" model that combines both representations of heterogeneity. MCMC estimation methods are well suited to estimate these models. However, widespread adoption of the MCMC methodology by practitioners is contingent upon the availability of user friendly software, some of which has become available.

3.2. The normative approach

Whether consumers should be grouped into segments or be targeted individually is a marketing rather than a statistical question. Therefore, when dealing with heterogeneity, marketing modelers should look also at the segmentation problem from a normative point of view: What level of aggregation of the data produces the most profitable marketing strategy? In other words, the decision on the appropriate aggregation level should be based on marketing and financial rather than on statistical criteria alone. Therefore, revenues and costs need to be included in the specification of the model. So far, only statistical "costs" have been incorporated in the majority of marketing models, which is unsatisfactory: rather than maximizing the likelihood function we might maximize the profit function. It seems desirable to include cost factors directly into models of segmentation to simplify decision-making processes.

We expect, in line with the opinion expressed by Leeflang and Wittink (2000), that such normatively-oriented approaches will become one of the mainstreams of marketing modeling in the near future. Emphasis should shift from inference on the heterogeneity distribution towards the entire decision problem. Efforts to construct models with better predictive power than existing ones will only not realize their full potential until we utilize them in the context of the decision-making process. This necessitates approaches for eliciting priors and utilities from decision-makers and techniques for examining the sensitivity of decisions to their specification. Clearly, improving marketing decisions requires better procedures to address the decision problem itself. As a case in point, game theory has already made its way to the marketing journals for close to a decade, but, recently, the empirical estimation of game-theoretic modeling has begun in marketing as well. Marketing data (scanner data, in particular) lend themselves very well to this arduous task. This approach leads to models that Leeflang and Wittink (2000) refer to as endogenizing marketing variables and we agree that treating those variables as exogenous has been one of the major limitations in marketing modeling of competitive reactions to date (Kadiyali et al., 1999).

4. The future of marketing modeling

In our view, one of the most important and currently underdeveloped area of research in our field is that of database marketing. Many firms routinely compile databases containing transactions with their customers, which can be of great value for direct-, micro- and database marketing (most of the automatic customization observed in e-commerce can be viewed as database marketing applied in real time). While the data are widely available, many companies still lack the skills and knowledge to adequately mind the data, extract relevant information and take competitive advantage of this information. Marketing researchers in practice seem to have focused mostly on the gathering and analysis of primary data, having neglected to sufficiently address the needs of firms for the analysis of their own customer transaction databases. Several challenges will face marketing modelers as they take on those tasks.

4.1. Large transaction databases

The size of the database has major impact on the use of many of the currently available statistical methods, if one wants to process all the data. Information on the entire population of customers is available, but data processing limitations force one to concentrate the estimation effort on smaller samples and to draw upon statistical inference for conclusions about the population (Balasubramanian et al., 1998). The issue of optimal designs for sampling from those databases in order to optimize the properties of the estimates has not been addressed sufficiently. Since a new sample can be drawn for each new analysis of the data, a much closer link between sample design and models can be achieved, where for each particular analysis, a sample design can be used that optimizes the properties of the particular model being used.

A further issue is that of linking various databases to customer transaction data. This is nowadays commonly done at the ZIP-code level, using geo-demographic data collected by specialized companies. Clearly, not all information on customers can be obtained based on the transaction context or by links with ZIP-code data. Often, one needs supplementing survey research, as indicated by Leeflang and Wittink (2000) for the important case of customer and employee satisfaction studies. However, such surveys are costly and usually cannot be conducted among all customers: data are typically obtained only from a sample of customers and are lacking for the remaining customers and need to be imputed. Fortunately, the design by which data are missing is under the control of the researcher, enabling model-based imputation methods that produce imputed values of all variables for every customer in the transaction database. Whereas imputation models are available (Little and Rubin, 1987), the structure of the missing data can be exploited to a larger extent than has been done heretofore. Rather than producing one or more fused data sets and working with those, data fusion may be done "on the fly" and much more precisely, by utilizing the model structure and the data to be modeled in question.

4.2. Integration of models

One of the great challenges for marketing model building in the next one or two decades is the integration of models that have until now been applied disjointedly into broad modeling frameworks. Since Marketing Science sets out to understand and predict the behavior of customers in response to marketing stimuli from multiple competing firms, it does not benefit from the reduction of those substantive problems to forms that are convenient from a mere mathematical or statistical perspective. Therefore, there is a growing need for models that deal with all potentially relevant aspects of consumer behavior within integrated frameworks.

Both theory, data and models are now favoring these efforts. Since the 1970s, many attempts have been made to (partially) integrate marketing theory. Much progress has been made in integrating models because of new numerical and statistical estimation techniques, as well as the growing availability of individual-level data. Mixture and hierarchical Bayes models for heterogeneity have already been integrated with generalized linear models, hazard models, MDS models, factor models, structural equation models, tree structure models, choice models, and so on. Similarly, factor and time series models have been integrated with several limited dependent variable models such as Count data, Tobit and logit models.

We therefore see platforms for model construction (rather than statistical packages) emerging. Those platforms will enable automated choices of probability distributions and facilitate flexible specifications of a model from constituent components such as consideration, perception, preference and utility formation, choice, heterogeneity, competitive reactions. They would provide suggestions for model improvement, diagnostic tests and would be organized by substantive marketing areas, such as segmentation, positioning, market structure, pricing, advertising and promotion effectiveness, direct marketing, satisfaction and e-commerce (rather than by statistical techniques). To that end, we see, both within the confines of our field and beyond its boundaries a vast untapped potential of collaboration of researchers in consumer behavior, psychology, psychometrics, economics, econometrics, statistics, and management to further the theoretical understanding and mathematical representation of exchange behavior among parties in the marketplace and hope that *IJRM* will play an important role in that process.

References

Ailawadi, K.L., Gedenk, K., Neslin, S.A., 1999. Heterogeneity and purchase event feedback in choice models: An empirical analysis with implications for model building. International Journal for Research in Marketing 16 (3), 175–177.

Allenby, G.M., Rossi, P.E., 1999. Marketing models of consumer heterogeneity. Journal of Econometrics 89 (1–2), 57–78.

Balasubramanian, S., Gupta, S., Kamakura, W.A., Wedel, M., 1998. Modeling large datasets in marketing. Statistica Neerlandica 52 (3), 303–324.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian data analysis. Chapman Hall, London.

Kadiyali, V., Vilcassim, N., Chintagunta, P., 1999. Product line extensions and competitive interactions: An empirical analysis. Journal of Econometrics 89 (1–2), 339–364.

Kim, J., Chatterjee, R., DeSarbo, W.S., Bijmolt, T.H.A., 1999. Incorporating context effects in the multidimensional scaling of "pick-any/N" choice data. International Journal for Research in Marketing 16 (1), 35–55.

Leeflang, P.S., Wittink, D.R., 2000. Building models for marketing decisions: Past, present and future. International Journal for Research in Marketing 17, 105–126 (this issue).

Little, R.J.A., Rubin, B.D., 1987. Statistical analysis with missing data. Wiley, New York.

Russo, J.E., Schoemaker, P.J.H., 1990. Decision traps: Ten barriers to brilliant decision making and how to overcome them. Simon and Schuster, New York.

Sirken, M.G., Herrmann, D.J., Schechter, S., Schwartz, N., Tanur, J.M., Tourangeau, R., 1999. Cognition and survey research. Wiley, New York.

Wedel, M., Kamakura, A.W., 1999. Market Segmentation: Conceptual and Methodological Foundations. Kluwer Academic Publishing, Dordrecht, Netherlands.

Wedel, M., Kamakura, W.A., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., Johnson, R., Lenk, P., Neslin, S., Poulsen, C.S., 1999. Discrete and continuous representation of heterogeneity. Marketing Letters 10 (3), 217–230.