A Factor-Analytic Probit Model for Representing the Market Structure in Panel Data

Elrod, Terry and Keane, Michael

February 1995

TERRY ELROD and MICHAEL P. KEANE*

Internal market structure analysis infers both brand attributes and consumer preferences for those attributes from preference or choice data. The authors exploit a new method for estimating probit models from panel data to infer market structures that can be displayed in few dimensions, even though the model can represent every possible vector of purchase probabilities. The result outperforms each of several other models, including Choice Map, SCULPTRE, and Chintagunta's latent class model in terms of goodness of fit, predictive validity, and face validity for a detergent data set. Because theirs is the only market structure model to outperform the structureless Dirichlet-multinomial stochastic brand choice model, the other methods cannot claim to have recovered market structure for these data.

A Factor-Analytic Probit Model for Representing the Market Structure in Panel Data

Market structure, as understood by marketers, is the explanation of consumer brand preferences in terms of the attributes of brands and consumer preferences for those attributes. *Internal* market structure analysis infers both the brand attributes and consumer preferences for those attributes from preference or choice data (Elrod 1991). Inferring market structure from preferences or choices among stimuli under the researcher's control has a long history and is still advancing. Methods for such data that recently have appeared in the literature include, for preferences, GENFOLD2 (DeSarbo and Rao 1986); for pairwise preferences, PROSCAL (MacKay and Dröge 1990; MacKay and Zinnes 1986), and for pairwise choices, DeSarbo, De Soete, and Eliashberg (1987); DeSarbo et al. (1988); and Carroll, De Soete, and DeSarbo (1990).

In recent years a number of methods have also been developed that can provide information about market structures from *real world* choice data. Such data typically entail repeated purchases in settings not under the researcher's control, so there are limits to the inferences that can be made. We cannot, for example, extend the set of attributes characterizing choice among brands by adding new concepts to the choice set. Yet, models that successfully recover market structure from real world data help managers understand current behavior. For example, such models allow managers to verify the positioning of existing brands and suggest to them attractive positions for new or repositioned brands, at least within the set of attributes characterizing existing choices. Ultimately, we can hope to adapt successful models for real world data so they can be estimated from data under the researcher's control. Then managers could use a single modeling framework to infer the attributes that govern choice in a product category (and the positions of brands on those attributes) as well as predict the performance of new concepts that entail new attributes.

MODELS FOR THE INTERNAL ANALYSIS OF MARKET STRUCTURE FROM REAL WORLD CHOICE DATA

A review of recently developed methods for the internal analysis of market structure from all types of preference and choice data is provided by Elrod (1991). A comparison of methods that can be applied to *real world* choice data is provided in Table 1. The methods differ in terms of the type of data they require, whether differences in market shares are explained by the structure (rather than captured by brand-specific constants), whether they can assess the effects of nonproduct marketing variables (such as price and display), the type of structure inferred, and how they treat consumer heterogeneity.

Models for Market Share Data

The first three models in Table 1 analyze changes in *market shares* over time, which are presumed to be caused by fluctuations in a marketing mix variable, such as price. Structure is inferred from the constraints imposed on cross-

*Terry Elrod is Associate Professor of Marketing, Faculty of Business, University of Alberta. Michael P. Keane is Associate Professor of Industrial Relations and Economics, Industrial Relations Center and Department of Economics, University of Minnesota, and Research Department, Federal Reserve Bank of Minneapolis. The authors acknowledge the support of the University of Alberta's Faculty of Business and the Federal Reserve Bank of Minneapolis, the computing assistance of Donna Boswell and Pradeep Chintagunta, and comments by Gary Russell. The data analyzed in this study were provided by the A. C. Nielsen Corporation, whose longstanding support of academic research in marketing is much appreciated.

Journal of Marketing Research
Vol. XXXII (February 1995), 1-16

Table 1
A COMPARISON OF MODELS FOR INFERRING MARKET STRUCTURE FROM REAL WORLD CHOICE DATA

Model	*Response Data*	*Explains Market Shares?*	*Includes Nonproduct Marketing Variables?*	*Model Structure*	*Consumer Heterogeneity*
DEFENDER (Shugan 1987; Waarts, Carree, and Wierenga 1991)	Market shares	Yes	Price (required)	Spatial (two dimensions)	Uniform or beta (one dimension)
Competitive Maps (Cooper 1988)	Market shares	No	Price, etc. (one required)	Spatial	Implicit
Allenby (1989)	Market shares	No	Price, etc. (one required)	Nonoverlapping clusters	Implicit
Grover and Srinivasan (1987); Jain, Bass, and Chen (1990)	Brand switching	No	No	Cluster	Latent class
DeSarbo and Hoffman (1986, 1987); DeSarbo and Cho (1989); Jedidi and DeSarbo (1991)	Pick-any/n (binary)	Yes	No	Spatial	Idiosyncratic or explained
Deterministic methods (correspondence analysis; CGS scaling; pick-any)	Purchase frequencies	No	No	Spatial	Idiosyncratic
SCULPTRE (Ramaswamy and DeSarbo 1990)	Purchase frequencies	Yes	No	Ultrametric tree	Latent class
Choice Map (Elrod 1988a, b)	Purchase frequencies	Yes	No	Spatial	Reduced-rank (bivariate) normal
Chintagunta (1994)	Disaggregate panel	Yes	Price, etc. (any number, optional)	Spatial	Reduced-rank latent class
Factor-analytic probit (introduced in this paper)	Disaggregate panel	Yes	Price, etc. (any number, optional)	Spatial	Full-rank multivariate normal

price elasticities. The advantages of these models are: (1) they include the effects of one or more marketing mix variables and (2) they avoid the problem brand choice models have of handling multiple purchases on single shopping occasions by analyzing only aggregate data.

However, the ability to infer consumer heterogeneity from aggregate data is limited. The models by Allenby (1989) and Cooper (1988) do not characterize consumer heterogeneity. In addition, because brand-specific intercepts are included in these models, differences in market shares are not accounted for by the market structures they infer. This limits their usefulness for product repositioning and new product development, because a brand's share cannot be predicted without knowing its intercept, and the market structure has nothing to say about what that intercept will be. The DEFENDER model (Hauser and Shugan 1983) *does* explain differences in market shares. Shugan's (1987) adaptation of DEFENDER allows it to be estimated solely from scanner data, but the researcher must prespecify a distribution for the household ideal vectors, which are confined to one dimension. More recently, Waarts, Carree, and Wierenga (1991) allow the shape of the univariate distribution to be estimated from the data.

Models for Brand Switching Data

A second approach has been to analyze *brand switching matrices* using latent class models. Brand switching matrices are constructed by tallying a one in cell (i, j) whenever a household is observed to purchase brand i followed by brand j. Latent class models assume that consumers belong to one of several segments (or classes). The segments are latent because households are not assigned to segments a priori.

Grover and Srinivasan (1987) and Jain, Bass, and Chen (1990) both fit switching matrices by postulating that the population of consumers is a mixture of segments. In Grover and Srinivasan's model, some segments are loyal to a single brand, whereas others switch symmetrically among all brands. Although the model is consistent with heterogeneity within switching segments, the degree of heterogeneity cannot be estimated from switching data. The model by Jain, Bass, and Chen (1990) is identical, except that it can allow for and estimate the degree of heterogeneity within switching segments because it excludes brand-loyal segments.

These versatile models have been shown to attain very good fits. Switching matrices require only two purchases per respondent, so they can be constructed for infrequently purchased items while retaining information about consumer heterogeneity. However, these models possess two important limitations that have to do with their structure and the type of data analyzed. First, these methods do not explain differences in market shares, which limits their usefulness. The only means for predicting share for a new or repositioned brand is to specify its share for each of the inferred segments, which is not particularly satisfactory.

Second, although brand switching matrices preserve *some* of the information about consumer heterogeneity, they do so imperfectly and at a price. Because brand switches are aggregated over time, information about the effects of marketing mix variables on brand choice is lost. Furthermore, households that are observed to buy more than twice pose a problem for the construction of switching matrices. For example, a household's ten consecutive purchases can be tallied as nine switches or as five independent pairs of purchases. Either way, this single household that switched numerous times is misrepresented as numerous *different* households that each switched *once*, which is statistically quite different. An extreme solution to this problem is to record a single switch for each household, but this discards most of the data for frequently purchased categories.

Models for Pick-Any/n (Binary) Data

A number of models have been developed for analyzing pick-any/n data. Pick-any/n data are binary preference or choice data. As with brand choice models, the number (n) and identity of alternatives evaluated by consumers is assumed to be known. A one (or pick) signifies the preference or choice of a brand, whereas a zero signifies rejection. There is no constraint on the number of brands that can be picked.

DeSarbo and Hoffman (1986, 1987) and DeSarbo and Cho (1989) developed models for pick-any/n data for which each respondent either does or does not prefer or choose a brand. Jedidi and DeSarbo (1991) extend these models by allowing a third mode to the data in addition to respondents and brands, such as choice or usage occasion.

All these models have two methods for treating consumer heterogeneity. Ideal points or vectors can be estimated separately for each household (referred to in Table 1 as idiosyncratic treatment of heterogeneity). Alternatively, the ideal points or vectors can be constrained to be linear functions of (i.e., be explained by) observed consumer characteristics. In addition, all the models allow brand attributes to be either known a priori or else inferred by the model.

Although published applications of these models have been to questionnaire data, the models could also be applied to panel data. For example, DeSarbo and Cho (1989) ask consumers which of 11 popular brands of soft drink they purchase and consume at least every other week. Comparable information is easily culled from scanner or diary panel data. However these models require binary data, which retain little information about frequency of purchase.

The Jedidi and DeSarbo (1991) model does best in this regard because the third mode can be shopping occasion, but the number of occasions analyzed must be the same for all households. Furthermore, the models also assume that each consumer's response for one brand is made independently of responses for the other brands. Such an assumption is often appropriate for the type of data for which these models were designed—checking responses on a questionnaire. However, the decision to purchase one brand, particularly on a single shopping occasion, surely affects whether other brands are also purchased.

The models also suffer from a difficulty in estimation when household-specific parameters such as ideal points or vectors are estimated. Because the models are estimated by maximum likelihood, the parameter estimates are not consistent in the number of households. In fact, for some households unique maximum likelihood estimates of their parameters do not exist,[1] and the models can obtain estimates of those parameters only by terminating the iterative estimation process before attaining the maximum likelihood. When household ideal points or vectors are constrained to be linear functions of observed household characteristics, then all parameter estimates will generally exist and be consistent, but there is no guarantee that consumer preferences are adequately explained by such variables.

Models for Purchase Frequencies

A number of models use information about purchase frequencies, the number of times each household purchased each brand during an observational period. Thus, purchases are aggregated over time but not over households. As with brand switching data, information about marketing mix effects is lost in return for model simplification and speedier estimation. Any model for such data can also analyze brand switching matrices simply by recoding the switching data as purchase frequencies and assuming two purchases for each household.

Deterministic models. Purchase frequencies are amenable to analysis by deterministic methods such as correspondence analysis (Hoffman and Franke 1986), CGS scaling (Carroll, Green, and Schaffer 1986, 1987) and Levine's (1979) pick-any procedure. Although correspondence analysis has the best claim to being the appropriate method for summarizing such data, these methods differ only in their scaling of the dimensions used to represent households and brands.

All these models analyze residuals from a main effects model, which assumes that each purchase frequency is proportional to its row (household) and column (brand) marginals. Because the column marginals are market shares, these methods do not explain market shares but instead remove them from the analysis, and brands with very different market shares may have identical positions in the maps.

Probabilistic models. SCULPTRE (Ramaswamy and DeSarbo 1990) and Choice Map (Elrod 1988a, b) are probabilistic models for purchase frequencies that differ in how they represent market structure and consumer heterogeneity. Choice Map, in both its vector (Elrod 1988a) and ideal point (Elrod 1988b) versions, infers a spatial structure by assuming that consumer preferences (whether vectors or ideal points) are multinormally distributed. Because numerical integration is utilized to estimate the models, the dimensionality of the space is kept to one or two dimensions.

Elrod and Winer (1991) compare spatial models and bench mark models (such as the Dirichlet-multinomial) for purchase frequencies in terms of goodness of fit, predictive validity, and ease of use. Choice Map performed the best of the market structure models considered, but the structureless Dirichlet-multinomial stochastic brand choice model predicted best of all. Elrod and Winer therefore conclude that none of the models successfully recovers the structures thought to underlie either of the two product categories they analyzed.

[1] For vector models (DeSarbo and Cho 1989; Jedidi and DeSarbo 1991), the likelihood function is maximized by a vector of infinite length whenever all brands chosen by that consumer project higher onto that vector than all brands not chosen. Furthermore, any such vector maximizes the likelihood function. Thus, the consumer's vector is of infinite length and indeterminate direction. This problem will arise for all households that choose only one brand, unless that brand lies interior to the other brands. It will also arise for households that choose two or more brands if those brands lie to one side of the brand map. The same households that have nonexistent vectors in the vector models will also have nonexistent ideal points—their likelihood-maximizing ideal points will lie at infinity in an indeterminate direction.

An alternative to spatial representations is a tree structure. Holman (1972) shows that simple tree structures can capture what requires many dimensions to capture using a spatial model. Ramaswamy and DeSarbo (1990) fit a probabilistic ultrametric tree model called SCULPTRE to one of the Elrod and Winer data sets and found it outperformed both Choice Map and the Dirichlet-multinomial models, according to the least squares measure of fit and predictive validity used by Elrod and Winer in the first (1988) version of their working paper. They attribute the apparent superiority of SCULPTRE to the absence of parametric assumptions about the distribution of household preferences for brands. However, it may be that SCULPTRE's success is more the result of Choice Map's inability to capture household heterogeneity in more than two dimensions.

A Model for (Disaggregate) Panel Data

Finally, we consider a model designed for panel data. Aggregation, whether over households or over time, is avoided, so all information about marketing variables as well as interhousehold heterogeneity is retained. Models designed for such data can also be estimated using data on purchase frequencies or brand switching by recoding the data and arbitrarily assigning brand choices to purchase occasions.

Most recently, Chintagunta (1994) has developed a latent class model for inferring market structure from such data. The model is a special case of the latent class model developed by Kamakura and Russell (1989). Segment-specific brand intercepts are constrained to lie within a subspace of few dimensions, with the inferred brand and segment locations in that space constituting a product-market map. It partakes of the virtues of latent class models, because (1) it is relatively easy to program, (2) it can be estimated using a fast personal computer, (3) it readily incorporates marketing mix effects, and (4) its nonparametric treatment of consumer heterogeneity is thought to be robust.

Although latent class models represent an important addition to methods for handling unobserved consumer heterogeneity, they are hampered by the coarseness of their approximation of consumer heterogeneity, particularly in the usual case in which within-segment heterogeneity is precluded. For example, Chintagunta (1994) approximates one-, two-, and three-dimensional distributions using two, three, and five points, respectively. Kamakura and Russell (1989) use five points to approximate a four-dimensional distribution.[2]

The limited ability of latent class models to adequately represent heterogeneity distributions has been better recognized in the econometrics literature in which these models originated. The seminal paper by Laird (1978) notes that using the latent class approach "gives a very limited picture of what the underlying distribution looks like," even for a univariate distribution. Heckman and Singer (1984), in an investigation of latent class models for duration data, "find that the NPMLE [nonparametric maximum likelihood estimator] recovers the structural parameters of the underlying models very well but does not accurately estimate the distribution of unobservables even in very large samples." In other words, although latent class models substantially improve estimates of the mean vector of household coefficients in the presence of heterogeneity, their ability to adequately characterize the heterogeneity itself is much less certain.

A NEW, FACTOR-ANALYTIC PROBIT MODEL FOR PANEL DATA

Examination of Table 1 shows that existing market structure models for real world choice data cannot explain differences in brand market shares, yield inconsistent estimates of model parameters (those shown in the table as having idiosyncratic consumer heterogeneity), and/or preclude households from having plausible coefficient vectors. We believe these shortcomings are consequential.

Models that explain consumer heterogeneity force household parameters to be a linear function of observed household characteristics. Latent class models (except for those for brand switching matrices) force every household to have one of a handful of parameter vectors. Spatial models use reduced-rank characterizations of consumer heterogeneity. They constrain brand and household parameters to lie within a subspace of few dimensions, a constraint that often is not satisfied in practice (Carroll 1972; Elrod 1988b, 1991).

A vector or ideal point model of market structure can represent every purchase probability vector only if the dimensionality of the space is at least one fewer than the number of brands.[3] This means that a market for only eight brands requires a seven-dimensional space, which is hard both to estimate and portray.

Fortunately, it is possible to portray in a few dimensions spatial market structures that nonetheless have the capability of representing any vector of purchase probabilities. The method was first proposed and implemented by J. Douglas Carroll and coauthors (Carroll, De Soete, and DeSarbo 1990; Winsberg and Carroll 1988). They represent the variability in consumer vectors in a $(J - 1)$-dimensional space, where J is the number of brands, as arises from variability on common factors and brand-specific factors, as in factor analysis. The market structure is captured by the common factors, which usually number no more than one or two.

We apply the approach of Carroll and associates, while extending their work in two ways. First, we infer factor-analytic structure from real world choice data, specifically panel data, in which households make repeated purchases from sets of J alternatives. Previously developed methods required pairwise dissimilarity judgments from subjects. Second, the model presented here is part of a probit modeling system, which allows the simultaneous investigation of brand differentiation, marketing mix effects, consumer heterogeneity (both observed and unobserved), purchase event feedback, and nonstationarity over time, all within a random utility framework. A demonstration of many of these properties of the system is provided by Keane (1994b).

Our purpose is to assess our model's ability, using a constrained form of factor analytic structure, to correctly recov-

[2]They first remove households that are observed to buy only one brand. Their application of latent class analysis to the remaining households yielded five segments.

[3]Allowing household-specific rescalings of a space reduces the requisite number of dimensions for that space, but it adds an equal number of dimensions for the scaling parameters, so the dimensionality of household-level parameters remains the same.

er the market structure in a product category that is well understood, using a data set that can be analyzed by competing market structure models for comparison. The alternative models fit to the data include Chintagunta's (1994) latent class model, Choice Map in its vector and ideal-point versions (Elrod 1988a, b), SCULPTRE (Ramaswamy and DeSarbo 1991), and the Dirichlet-multinomial stochastic brand choice model (Bass, Jeuland, and Wright 1976). The latter model is structureless, in that it assumes consumers adhere to the property of independence of irrelevant alternatives (IIA) both at the individual level and in aggregate. This property is inconsistent with the existence of market structure, where some brands compete more closely than others (Elrod 1988a). Market structure models that describe the data less adequately than the Dirichlet-multinomial model cannot be said to have recovered the structure for the product category.

FITTING PROBIT MODELS TO PANEL DATA

The method proposed here is made feasible by a fundamental advance in the estimation of probit models from panel data. Probit models have much appeal for the representation of choice processes. Like all random utility models, they assume consumers purchase brands with the highest utility on each shopping occasion, a conceptualization that aids in the formulation and interpretation of models of brand choice behavior. The assumption of normally distributed errors for utilities is readily justified and familiar to researchers from other contexts. Furthermore, probit models can readily represent departures from IIA by allowing the covariance matrix for the errors to depart from proportionality to the identity matrix. Finally, probit models can incorporate, in a theoretic manner, the effects of marketing mix variables and observed household characteristics.

The implementation of probit models has been hampered until recently by the lack of sufficiently fast and accurate estimation methods. Numerical methods have been used to evaluate the probability that each brand is chosen, which entails approximating an integral over $J - 1$ dimensions, where J is the number of brands. However, this method is too time-consuming for parameter estimation if J is larger than three or four. An important breakthrough in the estimation of probit models, termed the method of simulated moments, was developed by McFadden (1989) for cross-sectional data. This method uses simulation, rather than numerical integration, in such a way that few simulates are needed per individual. McFadden shows how his method can be used to characterize two consecutive purchases per individual from a set of J brands by treating each of the J^2 possible pairs as a different alternative, the approach taken by Chintagunta and Honore (1990) and by Chintagunta (1992).

Unfortunately, the method of simulated moments, as originally developed by McFadden, is not practical for panel data with more than four or five brands and two time periods. (For a discussion of this issue, see the Technical Appendix.) It is also not possible to distinguish among some of the different sources of error variance, such as between consumer heterogeneity in preferences and autocorrelation in utilities over time, using only two periods of data.

Keane (1990) describes a practical method for extending the method of simulated moments to the panel data case. It is based on recursive simulation of transition probabilities—the probabilities that each of the J brands are purchased at time t, given which brands were purchased by a household on all previous occasions. The method is illustrated here using choices from a set of eight brands on up to 24 different choice occasions. Keane (1994a) proves that the estimates are consistent and asymptotically normal as the number of simulates increases, but he also shows using simulated data that the method works well with few simulates per choice occasion. We provide a more complete discussion in the Technical Appendix. The reader should consult Keane (1993) for a thorough exposition of the estimation technology; we extend his work by adding a special form of factor-analytic structure to the model.[4]

In the next section we develop and justify two factor-analytic specifications that are useful for market structure analysis. One of these is designed to infer market structure, whereas the other is designed to represent structureless markets. We then report on the fits of these specifications to a data set, which is contained in Elrod and Winer (1991) and consists of purchase frequencies for eight different brands of household detergent by 100 different households. We find that a factor-analytic probit model outperforms, in terms of goodness of fit, predictive validity, and face validity, alternative market structure and brand choice models fit to the same data.

MODEL SPECIFICATION

Probit models are instances of random utility models. As such, they are consistent with the notion of utility maximization on the part of consumers. Letting d_{ijt} equal one if brand j is purchased by the ith household on the tth purchase occasion and zero otherwise, our probit model assumes that:

$$(1) \qquad d_{ijt} = 1 \text{ if and only if } \tilde{u}_{ijt} = \max\{\tilde{u}_{i1t}, \ldots, \tilde{u}_{iJt}\},$$

where $\tilde{u}_{ijt}$ is the utility of the jth brand (of J brands) to the ith household on the tth choice occasion. Discrepancies between what appears to the researcher to be the highest utility brand and the brand chosen by a consumer are accounted for by a random (to the researcher) component to utility, which captures unobserved and often transient determinants of utility. Thus, $\tilde{u}_{ijt}$ is random, as signified by the tilde (~). The probability that household i chooses brand j on purchase occasion t is equal to the probability that u_{ijt} is the largest element in the J-element column vector $\tilde{u}_{it} = [\tilde{u}_{i1t}, \ldots, \tilde{u}_{iJt}]'$.

The Factor-Analytic Probit Model

Our factor-analytic probit model can be expressed as

[4]We should also mention an even more recent advance in inference for probit models, that of Gibbs sampling with data augmentation (Albert and Chib 1993; Geweke, Keane, and Runkle 1994; McCulloch and Rossi 1994). In this method, prior distributions are specified for all model parameters, and the simulations are used to calculate properties of posterior distributions. Gibbs sampling provides a simple method for generating simulates from posterior distributions. A comparison of the method of simulated moments and Gibbs sampling lies outside the scope of this article. Interested readers are referred to Geweke, Keane, and Runkle (1994).

$$\tilde{u}_{it} = \alpha + \tilde{\mu}_i + \tilde{\varepsilon}_{it}, \tag{2}$$

where $\tilde{\varepsilon}_{it}$ is independently and identically distributed multinormally with zero mean vector and a covariance matrix proportional to the identity matrix, that is,

$$\mathrm{var}(\tilde{\varepsilon}_{it}) = kI \tag{3}$$

for some positive k. The vector α of brand intercepts in this case is also the average utilities for the brands (averaged over all households and time periods), and $\tilde{\mu}_i$ is the difference between the ith household's mean brand utilities and the population mean.

We avoid estimating the household-specific parameters $\tilde{\mu}_i$, and the accompanying problems of nonuniqueness or inconsistency of the estimates by assuming a prior distribution for these parameters and estimating only the parameters of the prior distribution. Specifically, we assume that the $\tilde{\mu}_i$ are multinormally distributed with covariance matrix Σ and (without loss of generality) mean vector 0. Multinormality is a natural choice of prior distribution, because the $\tilde{\mu}$ are logically unbounded, and the multinormal distribution is uniquely qualified to represent linear structures (Johnson and Kotz 1972).

We impose a particular factor structure on the covariance matrix Σ. We let

$$\Sigma = LL' + \kappa I, \tag{4}$$

where L is a J $\times$ M lower triangular matrix of loadings of the J brands on M common factors, L′ is its transpose, κ is a positive scaler, and I is a J $\times$ J identity matrix. Thus, κI is a diagonal matrix of brand-specific variances that are constrained to all be equal. This constraint makes the brand-specific variances a property of the product class as a whole, and therefore it is applicable to new or repositioned brands once they become established in the marketplace. The constraint, if it is consistent with the data, improves parsimony, interpretability, and usefulness for forecasting for new or repositioned brands.

It should be noted that the model given by (2)–(4) is underidentified, because the same positive affine transformation can be applied to all elements of $\tilde{u}_{it}$ without changing the choice probabilities. By convention, the estimation algorithm identifies the model by constraining, for all i and t, $\tilde{u}_{iJt} = 0$ *and var* $(\tilde{u}_{i1t}) = 1$, which implies that $l_{11}^2 + \kappa + k = 1$. These constraints are satisfied by setting k in (4) equal to $1 - l_{11}^2 - \kappa$, and by setting $\alpha_J = 0$, $\sigma^2_{\tilde{u}_{iJt}} = 0$, $\xi^2_{i1t} = 0$ and the Jth row of L equal to zero. Although these constraints serve to identify the model, they hinder exposition by destroying the symmetry in (4). We therefore discuss the factor-analytic probit model in its symmetric underidentified form given by (2)–(4), with the understanding that these constraints are subsequently imposed for model identification. All references to the number of parameters being estimated take the identifiability constraints into account.

To make (2)–(4) a market structure model, the brand intercepts (α) must also be explained in terms of the common brand attributes. This is accomplished by adding the constraint:

$$\alpha = Lc, \tag{5}$$

where c is an M-element column vector. Then the model given by (2)–(5) allows the estimation of the market share for a brand without requiring any brand-specific information other than its values for the M common factors. This is seen by reexpressing the factor-analytic probit model as

$$\tilde{u}_{ijt} = l_j \tilde{w}_i + \tilde{\eta}_{ij} + \tilde{\varepsilon}_{ijt}, \tag{6}$$

where l_j is the vector of values that the jth brand has for the M common attributes. None of the other parameters pertaining to (6) are brand specific. The vector of coefficients that the ith household has for the M common factors ($\tilde{w}_i$) is multinormally distributed over households with mean parameter vector c and an identity covariance matrix; $\tilde{\eta}_{ij}$, which is independently and identically distributed (iid) normally over brands and households with mean 0 and variance parameter κ; and $\tilde{\varepsilon}_{ijt}$ is also iid normal over brands, households, and shopping occasions with zero mean and variance k. Thus, the model can be used to predict shares for new or repositioned brands once their positions on the common factors have been specified. This would not be the case if there were brand-specific brand intercepts or brand-specific variances for the unique factors.

Equisimilar Probit: The Case of No Market Structure

When inferring market structure from real world choice data it is useful to also fit a model that is structureless. Then, substantially superior fit by a market structure model assures the researcher that structure in a market has in fact been detected.

A prime candidate for a structureless model is the Dirichlet-multinomial model (Bass, Jeuland, and Wright 1976). It is structureless because knowledge of a household's purchase history or purchase probability for a brand provides no information about the household's *relative* probabilities of purchase for the other brands. Specifically, for any household i, brands j $\neq$ k $\neq$ k′, purchase probability p, and purchase history $d_{ij} = (d_{ij1}, d_{ij2}, \ldots, d_{ijT})'$:

$$\frac{E(p_k \mid p_j)}{E(p_{k'} \mid p_j)} = \frac{E(p_k \mid d_{ij})}{E(p_{k'} \mid d_{ij})} = \frac{E(p_k)}{E(p_{k'})}. \tag{7}$$

Intuitively, (7) implies that information about a household's preference for any brand (as signified by its past purchase history or its purchase probability) provides no information about its relative preferences (as signified by probabilities of purchase) for the other brands. In a structured market, we would expect, for example, that knowing a household purchases a brand j less often than average means it probably shies away from brands similar to brand j. With models adhering to (7), the only effect that purchase probabilities have on each other is through the constraint that the probabilities must sum to one over all brands. Such models are said to possess the previously mentioned property of *independence of irrelevant alternatives* (IIA).

Elrod and Winer (1991) found that the structureless Dirichlet-multinomial model predicted better than any of the market structure models in the two product categories they investigated. On the basis of this evidence, they conclude that none of the market structure methods can claim to have recovered market structure for these data sets.

Table 2

PARAMETER ESTIMATES AND STANDARD ERRORS FOR THE FACTOR-ANALYTIC AND EQUISIMILAR PROBIT MODELS

Probit Models:	*Factor-Analytic Probit (M=2)*		*Equisimilar Probit*
	l_1	l_2	α
L1	−.134 (.133)	**.000**	.015 (.143)
P2	.052 (.168)	−.259 (.199)	−.474 (.176)
L3	−.082 (.247)	−.091 (.195)	−.516 (.197)
L4	−.106 (.136)	−.062 (.053)	−.226 (.156)
P5	.142 (.129)	.035 (.079)	.426 (.136)
L6	−.158 (.137)	−.043 (.058)	−.142 (.151)
P7	.053 (.108)	−.043 (.048)	−.081 (.148)
P8	**.000**	**.000**	**.000**
c	.062 (.122)	.316 (.116)	
κ		2.220 (.701)	4.884 (1.060)

Note: The standard errors are shown in parentheses. Parameter estimates that are constrained to equal zero for model identification are shown in bold.

We assess the fit and predictive validity of the Dirichlet-multinomial model, but it would also be useful to have a structureless model within the same family as the factor-analytic probit model. Then, any differences in fit or predictive validity could be attributed to the presence or absence of structure, because the difference in performance could not be accounted for by other differences in model assumptions. In addition, a structureless probit model could be fit to data with time-varying exogenous variables, which is not possible with the Dirichlet-multinomial model.

A structureless market can be modeled using the probit model as given in (2)–(4) with the additional constraint

$$L = 0. \tag{8}$$

Then there are no common factors. This is the closest the probit model can come to the case of IIA—knowledge of a consumer's *utility* for any subset of brands will provide no additional information about the consumer's utilities for the other brands.

Because there are no common factors in the equisimilar probit model, it is no longer appropriate to constrain the brand intercepts to be a linear function of them (as in (5)). This would force all brands to have an equal share. Therefore, (5) is not imposed and the equisimilar probit model is not an instance of market structure analysis nor even a special case of such a model, but it is within the same modeling family.

The equisimilar probit model requires the estimation of only J parameters: κ and the J − 1 unconstrained elements of α. The parameter κ captures the degree of heterogeneity (or brand loyalty) in the product category. As κ approaches zero, so does consumer heterogeneity, and no household will be loyal to any of the brands. At the other extreme, as κ approaches plus infinity, every household becomes absolutely loyal to one of the brands. The Dirichlet-multinomial model also estimates J parameters. As with the equisimilar probit model, J − 1 parameters suffice to fit the market shares, leaving one parameter to capture the degree of brand loyalty in the product category.

The equisimilar probit model is also very similar to the equisimilar version of Choice Map investigated by Elrod (1988a). The latter is identical to the probit version except that the errors $\tilde{\epsilon}_{ijt}$ were iid double exponential, as in the logit model, rather than iid standard normal. Elrod found that the Dirichlet-multinomial model performed almost identically to the equisimilar version of Choice Map in terms of predictive validity and goodness of fit. Therefore, we can expect that the equisimilar probit model will yield fits similar to the Dirichlet-multinomial, with the advantage that it is within the probit family.

EMPIRICAL ASSESSMENT

We now evaluate the factor-analytic probit model against several alternative models using a data set reproduced in Table 2 of Elrod and Winer (1991). The data set consists of a count of the number of times each of 100 households purchased each of eight brands of laundry detergent in each of two 6-month time periods. The panel purchased a total of 607 and 592 units in the first and second periods, respectively. The order of the brand purchases is not recorded, nor are the prices or other attributes of the brands.

Elrod and Winer compare several methods for market structure analysis. They use knowledge of which brands were powdered detergents and which were liquid to assess the face validity of the maps. Predictive validity was assessed by fitting models to the purchases made in the first time period and predicting purchase frequencies for the second. The performance of the factor-analytic probit model in terms of these criteria of predictive and face validity, as well as on goodness of fit, is reported here.

We include several models in our comparison. First, we include the vector and ideal point versions of Choice Map. Elrod and Winer (1991) found that the ideal point version of Choice Map performed the best of the mapping methods included in their study. They also found the vector version performed nearly as well, and it is more similar to the factor-analytic probit model introduced here. In fact, there are only two differences between the vector version of Choice Map and the factor-analytic probit model considered here. The smaller difference is in the choice of distribution for the error term ($\tilde{\epsilon}_{ijt}$); Choice Map assumes it is iid double exponential rather than iid normal. The more important difference is that Choice Map assumes that there are no brand-specific factors. Specifically, it assumes that $\kappa = 0$.

We include the Dirichlet-multinomial model, because it predicted best of the models studied by Elrod and Winer, even though it assumes that there is no structure in the market. We also include SCULPTRE, by Ramaswamy and DeSarbo (1990). SCULPTRE performed better than the Dirichlet-multinomial model on a data set contained in Elrod and Winer's (1991) study, according to least squares criteria they used in the 1988 version of their working paper. Here, no deterministic models are included, so it is possible to compare all models using more satisfactory likelihood-based criteria. We also compare the models on a somewhat improved least squares criterion used in the 1991 version of the Elrod and Winer paper.

Finally, we include the latent class market structure model by Chintagunta (1994), which is similar to the vector version of Choice Map except that (1) consumer heterogeneity is represented using a number of homogeneous segments instead of the bivariate normal distribution, (2) it may incorporate the effects of exogenous variables, and (3) it is not restricted to two dimensions, although in practice it is limited to perhaps three. However, it shares with Choice Map the assumption that every household's vector of expected utilities for the J brands (corresponding for the ith household to $\alpha + \tilde{\mu}_i$ in (2)) is confined to an M-dimensional subspace, where M is the number of dimensions in the map and typically is far less than J. The factor-analytic model can mirror this assumption by assuming that there are no brand-specific factors; that is, by assuming $\kappa = 0$.

Chintagunta (1994) describes his model as employing a factor structure. However, although his model allows for common factors, it does not contain brand-specific factors. Because ordinary factor analysis requires the existence of unique factors, we would describe his method as using principal components analysis. The distinction is more than semantic, because models that constrain expected utilities to lie within an M-dimensional subspace have asserted that there is zero probability of households possessing expected utilities in any of the other $J - M - 1$ feasible dimensions. By including unique (in this case, brand specific) factors and employing a continuous distribution for utilities, we assign a positive probability to every possible vector of expected brand utilities. We seek to demonstrate that assigning zero probabilities to an infinite number of expected utility vectors is a mispecification that is substantial enough to prevent all models that share this assumption from recovering market structure.

Finally, we include in our comparison the factor-analytic and equisimilar probit models. The former is estimated with two dimensions. Experience with Choice Map has shown that one dimension does not suffice to capture household heterogeneity and differences in market shares. The equisimilar probit model, if it performs similarly to the Dirichlet-multinomial model, may serve as a bench mark for evaluating the appropriateness of the factor-analytic probit model for a given data set.

Notes on Estimation Procedures

The factor-analytic models were estimated using a mainframe computer that ran a specially written FORTRAN program. Because Ramaswamy and DeSarbo (1990) fit their model to the same data set used here, we had no need to reestimate their SCULPTRE model, but we use the estimates reported in their paper to assess their model on the performance criteria reported here.

The latent class model (Chintagunta 1994) was estimated using modifications of the program supplied by the author. We generalized the program to allow estimation of a two-dimensional map for any number of segments and brands in the case of data without marketing mix covariates. We also investigated the matter of local optima in the maximum likelihood estimation routine by estimating the model 50 different times for each number of segments using random starting values.

The results showed sensitivity to starting values that increases with the number of segments. For instance, only the four best-fitting maps of 50 yielded essentially identical maps for the three-segment solution. We chose three segments as a starting number because it was the number used by the author to generate two-dimensional solutions.

We also estimated maps using five, seven, eight, nine, ten, and eleven segments. The best-fitting solutions for all these cases failed to terminate normally, because some of the parameters being estimated were iterating toward infinite values. We therefore introduced a constraint to yield finite estimates for these cases.

Estimation for the latent class model can fail when either a brand location or a segment's importance weights iterate toward infinity. The least arbitrary constraint appeared to be to require that every segment have a probability of buying each brand that is no less than 10^{-12}, a constraint that was binding for all maps involving five or more segments.

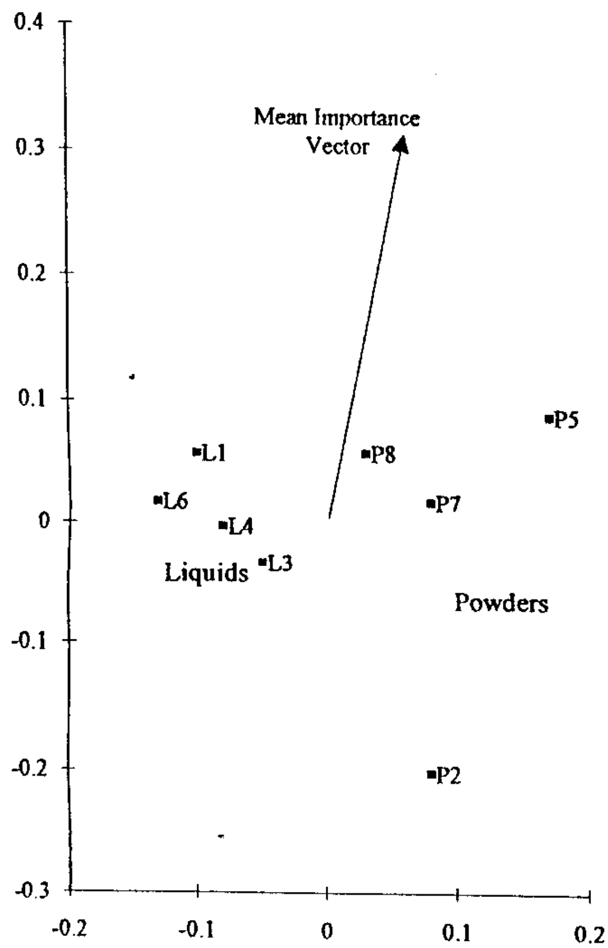
Beginning with the ten-segment solution, the 50 random starting values yielded maps with worse fit than the best-fitting nine-segment solution, indicating that none of the 50 solutions was globally optimal. Therefore, we used another variation of the program that utilized a well-fitting map of fewer segments as a starting solution. This yielded a ten-segment map with slightly better fit but with worse BIC and AIC scores. Using the same procedure for an 11-segment solution yielded no improvement in fit. Examination of the estimates for that case showed it had simply split one of the ten segments in two.

The two Choice Map models were also fit 50 times using random starting values, and the best solutions were retained for evaluation. Local optima were as problematic as they were for the latent class models, but no constraints were necessary to yield maximum likelihood estimates. Chintagunta (1994), in his comparison of Choice Map and his latent class model, does not specify how he handled the matter of starting values. His program for the Choice Map model also used a different numerical integration routine than that given in Elrod (1988a).

The Probit Model Estimates

The parameter estimates and standard errors for our two probit models are provided in Table 2. As in factor analysis, no test for significance of single parameters is meaningful with the exception of the parameter κ, which is highly statistically significant for both models. Its estimate is smaller for the factor-analytic probit model than for the equisimilar probit model. That is to be expected, because as is shown in

Figure 1
DISPLAY OF THE FACTOR-ANALYTIC PROBIT MAP



(4), when L is constrained to equal a zero matrix, κ alone must approximate the covariance matrix Σ.

Information about market structure is provided by the loadings of the brands on the common factors, that is, by the estimates of the matrix L of the two-factor model. The meaning of the estimates is best conveyed by plotting them in Figure 1. (The brand locations in the figure have been mean centered.) Elrod and Winer (1991) do not provide brand names by agreement with the provider of the data, but liquid detergents are distinguished from powders by the letters L and P in their pseudonyms.

It appears from Figure 1 that the model recovers a liquid-powder distinction that roughly corresponds to the horizontal axis. The near orthogonality of the mean vector to the liquid-powder dimensions confirms the need for two dimensions. We have no basis for interpreting the vertical axis, but its strong relationship to the mean importance vector suggests that it has something to do with how established the brands are in the market. We turn now to more formal assessments of this model and its competitors.

Predictive Validity

Elrod and Winer (1991) divided each household's purchasing data into two 6-month time periods. We estimated all models on the first time period, and the estimates were then used to assess the fit of each model to the holdout data.

Derivation of the loglikelihood predictive criterion. The Dirichlet-multinomial model, SCULPTRE, and Choice Map all assume that every household has a constant probability of choosing each brand on each choice occasion and that these choices are made independently across households and choice occasions. The factor-analytic probit algorithm is designed to model the effects of time-varying exogenous variables, past purchases, past error terms, and household characteristics, but because the Elrod and Winer data consist simply of a count of the number of times each brand was bought by each household, its reduced form also assumes that each household has a constant probability of purchasing each brand.[5] Therefore, all the models are consistent with the multinomial distribution for the number of purchases of each brand. The multinomial probability of observing purchase frequencies $y_i = [y_{i1}, y_{i2}, \ldots, y_{iJ}]'$ (where $y_{ij} = \Sigma_{t=1}^{T_i} d_{ijt}$) for the ith household possessing purchase probabilities $p_i = [p_{i1}, p_{i2}, \ldots, p_{iJ}]'$, is given by:

$$\text{(9)} \qquad Pr(y_i \mid p_i) = \left(\frac{T_i!}{\prod_{j=1}^{J} y_{ij}!} \right) \prod_{j=1}^{J} p_{ij}^{y_{ij}}.$$

Using the symbol # to denote the holdout data, for any model the likelihood for the holdout data of the parameters estimated from the calibration data is:

$$\text{(10)} \qquad \exp\,[L^{\#}(\hat{\theta})] = \prod_{i=1}^{N} \int Pr(y_i^{\#} \mid p) f(p;\hat{\theta}) dp,$$

where $L^{\#}(\hat{\theta})$ is the loglikelihood for the holdout data, $Pr(\cdot)$ is the multinomial probability of the holdout frequencies for the ith household having purchase probabilities p given by (9), and $f(\cdot)$ is the density of the purchase probabilities inferred by fitting the model to the calibration data. All the models included in our comparison differ only in their choice of $f(\cdot)$, the heterogeneity distribution. Equation (10) is also simply the probability of observing holdout frequencies $Y^{\#} = [y_1^{\#}, \ldots, y_i^{\#}, \ldots, y_N^{\#}]$, assuming that the choice of heterogeneity distribution $f(\cdot)$ and of the estimates $\hat{\theta}$ is correct.

Table 3 summarizes the performance of the models on their predictive validity for the holdout data, on their goodness of fit to the estimation data, and on face validity. The models are ranked in the table from best to worst, according to their values for $-L^{\#}(\hat{\theta})$. (The measures AIC, HQ, BIC, and CAIC are all based on $-L^{\#}(\hat{\theta})$.) Comparisons across measures for the holdout and estimation data sets are simplified by reporting standardized values in the table. Raw values for the measures may be reconstructed using the means and standard deviations (with the divisor = 12) for the raw val-

[5]Because the factor-analytic probit algorithm is designed to analyze cross-sectional time series data, it was applied to a random ordering of the detergent data purchases. The resultant likelihood was then multiplied by the multinomial coefficients to make the likelihood comparable to the likelihoods of the other models.

Table 3

AGREEMENT OF MODELS WITH THE ESTIMATION AND HOLDOUT DETERGENT DATA SETS[a]

Model	Number of parameters	Holdout data		Estimation data					Face validity errors
		$-L^{\#}(\theta)$	MSE_d	$-L(\hat{\theta})$	AIC[b]	HQ[b]	BIC[b]	CAIC[b]	
FAP[c]	16	**<–1.48**	<–0.64	**<–0.70**	**<–0.89**	**<–0.96**	<–1.02	<–1.03	
DM	8	–1.09	**–.77**	–.34	–.75	–.94	**–1.16**	**–1.26**	NA
EP[c]	8	<–0.90	<–0.66	<–0.28	<–0.69	<–0.88	<–1.11	<–1.21	NA
SCULPTRE	>17	–.34	–.14	.24	>0.09	>0.01	>–0.10	>–0.16	2
LC10[d]	40	–.29	–.41	–.65	–.16	.09	.43	.62	1
LC9[d]	37	–.28	–.36	–.63	–.23	–.02	.27	.43	2
LC8[d]	34	–.13	–.28	–.55	–.24	–.06	.17	.30	2
LC7[d]	31	.29	–.15	–.34	–.10	.02	.19	.28	1
CM-IP	18	.42	–.51	–.41	–.51	–.55	–.58	–.58	
CM-V	15	.45	–.12	–.07	–.28	–.38	–.50	–.56	
LC5[d]	25	.83	1.08	.67	.76	.77	.77	.76	3
LC3	19	2.52	2.95	3.07	3.02	2.89	2.62	2.42	3
Mean		564.26	131.85	520.35	1085.54	1109.17	1143.94	1166.35	
Std. Dev.		42.37	25.10	36.15	71.19	73.04	78.51	83.51	

[a]Designations of preferred measures are shown in bold, as is the best score on each measure. The symbol > signifies that the number shown is a lower bound, whereas < signifies an upper bound. Measures of agreement with the estimation and holdout data sets have all been standardized. The raw means and standard deviations are shown at the foot of the table. The model abbreviations are: FAP (factor-analytic probit), DM (Dirichlet-multinomial), EP (equisimilar probit), LCs (Chintagunta's latent class model with s segments), CM-IP (Choice Map, ideal point version), and CM-V (Choice Map, vector version).

[b]This measure is designed to choose among nested models. Because none of the models here is a special case of any of the others, it must be regarded as an approximate indicator only.

[c]Because the factor-analytic and equisimilar probit models are not estimated by maximum likelihood, all likelihood-based measures of agreement with the holdout and estimation data sets (all measures except MSE_d) understate the performance of these models.

[d]This latent class model required imposition of a constraint to yield finite maximum likelihood estimates of its parameters. The constraint imposed was that every segment have a probability of buying every brand of at least 10^{-12}.

ues shown at the bottom of the table. Horizontal broken lines in the table divide the models into groups so that those higher in the table dominate the lower groups on all (or nearly all) measures. The best measure for each criterion is shown in bold.

The table shows that the factor-analytic probit model (FAP) with two common factors outpredicted all other models on the holdout loglikelihood criterion, including the two structureless models: the Dirichlet-multinomial (DM) and the equisimilar probit (EP). Therefore, we can conclude that the loglikelihood criterion of predictive validity supports the ability of the factor-analytic probit model to recover market structure for these data, a conclusion that is not supported for any of the other market structure models. This occurred despite the loglikelihood measure understating the performance of the factor-analytic probit model. (Unlike the other models, the probit models are not fit by maximum likelihood.)

We also see by this criterion that SCULPTRE outperformed all the Choice Map and latent class models. Latent class models with only three or five segments performed worst, but maps with seven or more segments performed better than the two Choice Map models, which performed similarly. The equisimilar probit model performed similarly to the Dirichlet-multinomial model, particularly when one considers that its fit is understated by this measure.

A weighted least squares criterion for predictive validity. At the suggestion of a reviewer, we also show a weighted least squares measure of predictive validity, the one used by Elrod and Winer (1991). Except for a scaling constant their criterion was,

$$(11) \qquad MSE_d = \sum_{i=1}^{N} \sum_{j=1}^{J} T^{\#}_i (p^{\#}_{ij} - \hat{p}_{ij})^2,$$

where $p^{\#}_{ij}$ is the observed proportion of purchases of brand j made by the ith household during the holdout period $(= y^{\#}_{ij}/T^{\#}_i)$, and $\hat{p}_{ij}$ is a model's prediction of the household's probability of buying brand j.[6] For each of the models included here we use for p_{ij} the mean of the posterior distribution of each household's purchase probability; that is,

$$(12) \qquad \hat{p}_{ij} = E[p_j \mid y_i; \hat{\theta}] = \int p_j \Pr(y_i \mid p) f(p; \hat{\theta}) dp.$$

We do not favor MSE_d for three reasons that are made apparent by comparing it to the multinomial distribution. First, it ignores that the variance of an observed proportion is a function of the true proportion. Second, it ignores that the distribution of an observed proportion about its true value is skewed. And finally, it ignores the substantial variability in the true proportion p_{ij}. Elrod and Winer had no choice but to use a measure such as MSE_d, because most of the models included in their study were deterministic and yielded many predicted probabilities of zero, so a measure such as (10) could not be used. Despite the measure's shortcomings, we see from Table 3 that a number of findings are preserved. Although the factor-analytic model is marginally outperformed by the two structureless models on the MSE_d criterion, it still outperforms all the other market structure models.

[6]Elrod and Winer (1991) also used a criterion that assessed each model's ability to fit the brand market shares.

We should point out that (11) differs from the criterion used in the 1988 version of the Elrod and Winer working paper cited by Ramaswamy and DeSarbo (1990), in which they computed the squared differences between observed and predicted *frequencies*; that is, their measure was (except for a scaling constant)

$$(13) \qquad \sum_{i=1}^{N} \sum_{j=1}^{J} (y_{ij}^{\#} - T_i^{\#} \hat{p}_{ij})^2 = \sum_{i=1}^{N} \sum_{j=1}^{J} T_i^{\#2} (p_{ij}^{\#} - \hat{p}_{ij})^2.$$

Under the multinomial distribution, observed proportions have a variance that is inversely proportional to the number of purchase occasions, so squared errors should be weighted by $T_i^{\#}$, as in (11), and not by $(T_i^{\#})^2$. Unlike (11), (13) assigns too much weight to households that purchase many units during the holdout period. This accounts for the apparent discrepancy between our results and those of Ramaswamy and DeSarbo (1990), who found that SCULPTRE outperformed the Dirichlet-multinomial model when evaluating predictive fit using (13).

Goodness of Fit

The fourth column of the table shows minus the loglikelihood that results when each model's estimates are applied to the *estimation* data $(L(\hat{\theta}))$. This criterion does not take into account the number of parameters estimated, so models that involve more parameters are favored. Furthermore, we cannot use likelihood ratio tests for the significance of additional parameters, because our models are not nested. However, the likelihood function still allows a partial ranking of methods. Some evidence for a model's superiority for these data is obtained whenever it fits better than a competing model that estimates the same number or more parameters.

Thus, on the loglikelihood criterion, the factor-analytic probit model with 16 parameters fit better than all other models, including those that estimated as many at 40 additional parameters. The structureless models also performed well, considering they estimate only eight parameters, but some models that estimate many more parameters fit the data better. What remains in doubt, however, is whether the better fit to the estimation data obtained by the factor-analytic probit model warrants the estimation of eight more parameters than are estimated by the structureless models.

To attempt to answer to this question, we use a number of measures that attempt to correct for the number of parameters estimated by a model. All these measures were developed to choose among nested models for which the errors are independently and identically distributed across observations. None of these conditions apply to any of the models considered here. Therefore, they are offered here tentatively and in lieu of better measures for such models, so the fits to the holdout data must be considered primary. Nonetheless, such measures are of value because researchers will often be reluctant to set aside a substantial part of available data to use for holdout criteria. Furthermore, these measures aid in our comparison of the different models, because the measures are based on additional data (the estimation rather than the holdout data). The measures are shown in Table 3 in order of the increasing penalty imposed for estimating additional parameters. Perhaps the best known of such measures is the Akaike Information Criterion:

$$(14) \qquad AIC = -2L(\hat{\theta}) + 2k,$$

where $L(\hat{\theta})$ is the loglikelihood for the estimation data and k is the number of parameters estimated by the model.

We see that this criterion identifies the FAP model as the best, followed by the DM and EP models. A lower bound is shown for SCULPTRE, which requires a brief explanation. SCULPTRE seeks an ultrametric tree representation of a market. Ultrametric trees must satisfy ultrametric inequalities,[7] some of which are binding and some not. Ramaswamy and DeSarbo (1990) provide an upper bound on the number of free parameters, which for these data is 47. It is also possible to obtain a lower bound.

Given a choice of tree structure, there are still free parameters that must be estimated. Specifically, the model must determine the height of each node in the tree (except the lowest, which is arbitrary) and the size of each customer segment (except one, because segment sizes must sum to one). For these data, this implies 17 free parameters. Because SCULPTRE does not require prespecification of a tree structure, but instead seeks that tree that best represents the data, the true number of free parameters lies somewhere between these two bounds.

We have found that using 17 as the number of free parameters in SCULPTRE makes all the measures of fit involving sample size agree better with the holdout loglikelihood results than any other choice of k in the interval [17,44]. It also overstates the performance of the model on these measures.

Although the AIC criterion is information-theoretic, it has been criticized on the grounds that it will always lead to the choice of the most general of a set of nested models given a sample of infinite size, a phenomenon that is termed dimension inconsistency in the model selection literature (Teräsvirta and Mellin 1989). Dimension refers to the number of free parameters estimated by a model. A model selection criterion is termed dimension consistent if, in a comparison of a series of nested models of which the most parsimonious is the correct one, the measure correctly identifies the true model, with probability one as the sample size goes to infinity. Neither AIC nor the sequential application of loglikelihood ratio tests possesses this property.

This criticism has led to the development of three other measures shown in the table that are dimension consistent. All these measures impose a penalty for additional parameters that increases with sample size, and for these data the penalties are greater than those imposed by AIC. The first of these measures is offered by Hannan and Quinn (1979) and is given by

$$(15) \qquad HQ = -2L(\hat{\theta}) + 2k\ln[\ln(N)],$$

where N is the number of iid observations in the data. We see from Table 3 that on this criterion the factor-analytic

[7]Ultrametric inequalities are defined as follows. Let d_{sj} be the distance between segment s and brand j in an ultrametric tree. (This distance is defined as the height of their least common ancestor.) Then for all segments $r \neq s$ and brands $j \neq k$, it must be true that $d_{sj} \leq \max\{d_{rj}, d_{rk}, d_{sk}\}$.

probit model marginally outperforms the structureless models and substantially outperforms all others.

Perhaps best known to marketers is the Bayes Information Criterion (Schwarz 1978)

$$\text{BIC} = -2L(\hat{\theta}) + k\ln N. \tag{16}$$

Another criterion that imposes an even greater penalty for additional parameters than BIC is the consistent Akaike information criterion (CAIC) developed by Bozdogan (1987)

$$\text{CAIC} = -2L(\hat{\theta}) + k(\ln N + 1). \tag{17}$$

We see that by the BIC and CAIC criteria the structureless models are to be preferred to the factor-analytic probit model, with the poorer performance of the other models unchanged.

Of the five measures of fit shown in Table 3, the factor-analytic probit model is preferred on three and the Dirichlet-multinomial model on two. However, the measures are not of equal value for choosing among models. The unadjusted loglikelihood is of the least value and, in our opinion, the HQ measure is of the greatest value. The HQ measure was derived by Hannan and Quinn (1979) in the context of autoregressive models to minimize the penalty for extra parameters while preserving the property of dimension consistency. An overemphasis on dimension consistency is unwise, because for finite samples it increases the risk of incorrectly accepting a model that contains too few parameters. It can be argued that it is better to run the risk of selecting a model with too many parameters, because at least it is not misspecified. Criteria such as BIC and CAIC impose a greater penalty for extra parameters than the property of dimension consistency requires, whereas HQ does not.

Our preference for HQ on theoretical grounds is supported by its agreement with the loglikelihood fits to the holdout data. The HQ measures in Table 3 correlate .90 with the holdout loglikelihoods; AIC is almost as highly correlated at .89. The BIC and CAIC measures are correlated .87 and .84, respectively. Finally, the loglikelihood fit to the estimation data is correlated .83.

We should note that the dimension-consistent measures were computed using a value for N equal to the number of households, because all the measures that use N assume that the observations are independently distributed, and all of the models evaluated in Table 3 regard households as being sampled independently of the population. This choice of N is supported by the data: Any choice of N greater than 100 reduces the correlation of all measures involving N with the holdout loglikelihood.

Face Validity

Elrod and Winer (1991) hypothesized a priori that the detergent category would show some degree of loyalty to product form. They expected mapping methods to distinguish between liquid and powdered detergents. Therefore, we evaluate the models on their ability to recover the liquid-powder distinction.

In hierarchical tree structures such as those produced by SCULPTRE, it is trivial to determine the success of this distinction because the uppermost branch in the tree divides the brands into two groups. With spatial models, it is less obvious how best to determine a map's consistency with the liquid-powder distinction.

We wrote a computer program to find for each spatial model the partitioning of the eight brands into two groups that minimized total within-group variance. For all types of models, the number of incorrect assignments of brands is then defined as the minimum number of brand reassignments required to attain the liquid-powder distinction. We see from the last column of Table 3 that the factor-analytic probit model correctly attained the distinction. Of the 162 different ways of partitioning eight brands into two groups, only two are consistent with the hypothesis, so the probability of the correct partitioning occurring by chance is 2/162 = .0123.

Every other market structure method except the Choice Map models misassigned at least one brand. Of course, it is possible that the liquid-powder partitioning hypothesized by Elrod and Winer is wrong, but the models that disagree with this partitioning do not agree among themselves on an alternative partitioning of the market.

Conclusions Concerning the Empirical Assessment

Table 3 provides eight measures designed to assess the ability of the factor-analytic and competing models to explain the detergent data. Before focusing on the individual measures, we want to emphasize conclusions arrived at by the measures as a whole.

First, the first group of three models listed in the table outperforms all of the other models on all criteria except their likelihoods for the estimation data. The second group of seven models outperforms Chintagunta's five-segment model on all measures, which in turn outperforms the three-segment model on all measures. The measures disagree only on the ordering of the models within those groups. Because the best performing group includes both structureless models, the measures agree that the competing market structure models fail to recover market structure for these data. The estimation loglikelihood does not contradict this conclusion, because the models that fit the estimation data better than the first three models require many more parameters.

Second, we find that within the first group of models every measure places the factor-analytic probit model either first or last. Thus, regardless of the choice of measure, the same conclusion about whether the factor-analytic probit model has or has not detected structure within a market would be reached by comparing that model's performance to the equisimilar probit model or to the Dirichlet-multinomial model.

Therefore, it appears that tests for market structure may remain within the probit modeling family. This is important because the equisimilar probit model may, like the factor-analytic probit model, include many other effects on choice, such as those of previous brand choices, exogenous variables, autocorrelated errors, and household characteristics. The Dirichlet-multinomial model, which is a stochastic brand choice model, cannot assess any of these other effects. We now have a basis for believing that the equisimilar probit model may be relied upon to test for market structure even in the presence of these other effects.

Finally, we see that the Dirichlet-multinomial model slightly outperformed the equisimilar probit model on all criteria. This is to be expected, because the two models have a very similar structure, yet the Dirichlet-multinomial model was estimated by the more efficient method of maximum likelihood. Because the method of simulated moments is asymptotically as efficient (in the number of households) as maximum likelihood, the apparent difference in performance of the two models will be reduced for panel data sets of a more usual size. The relative performance of the factor-analytic probit model will also be improved.

The eight measures disagree on whether the factor-analytic probit model has recovered market structure for these data; five indicate that it has and three that it has not. The criterion of face validity indicates that it has, as does the more theoretic of the two criteria for predictive validity. Three of the five measures of fit to the estimation data also support the factor-analytic probit model, including what we argue is the most defensible measure, HQ (Hannan and Quinn 1979). Thus, it appears that the factor-analytic probit model has in fact recovered market structure for these data. Certainly the Choice Map, latent class, and SCULPTRE models can make no such claim.

SUMMARY AND CONCLUSION

We have proposed and evaluated a new factor-analytic probit model for inferring market structure from panel data. It explains and predicts choice market shares for brands on the basis of knowledge regarding their location on a few common dimensions. All other parameters of the model are properties of the product class as a whole.

The ability of the model to recover the market structure for a data set is assessed on three criteria: predictive validity, goodness of fit, and face validity. Our model performs best among all models that are considered on all three criteria. Furthermore, becaue it is the only model to outperform the structureless models on either predictive validity or goodness of fit, we conclude that it is the only successful market structure model among those considered for these data.

We now offer some speculations about why the different models performed as they did. First, we note that the three best performing models (the Dirichlet-multinomial and the two probit models) ascribe a positive probability to every possible purchase probability vector. We believe the other models perform substantially worse in part because they necessarily preclude many purchase probability vectors. SCULPTRE assumes that there are only as many distinct purchase probability vectors as there are segments (six in this case). The Choice Map models require that the distribution of household ideal points or vectors, which underlie the purchase probabilities, vary in at most a two-dimensional subspace of the full $J - 1$ ($= 7$) dimensions. Finally, Chintagunta's latent class model imposes both types of restrictions. Neither set of constraints is supported by the data.

Second, we obtain some information about the relative performance of parametric and nonparametric treatments of heterogeneity. Clearly, reliance on parametric distributions for heterogeneity is not a critical shortcoming of a model, because the three best models are all parametric. However, by comparing the Choice Map and latent class models we can obtain some indication of how the two approaches compare while controlling for the dimensionality of the map. Because we rank-order the models by their predictive log-likelihoods, the Choice Map models appear to perform worse than latent class models with seven or more segments. However, the ideal point version of Choice Map performs the best of all previously existing market structure models on six of the eight criteria, and it is surpassed on the likelihood criterion only by models that estimate many more parameters.

The vector version of Choice Map performs better than these other models on five of the criteria. Thus, it appears that a parametric approach to market structure analysis is not a hindrance to model performance. This is not surprising, given the degree of latency of the distribution of consumer heterogeneity. Remember, only choices are observed. The distributions for heterogeneity apply only to the expected values of the random variables that underlie these choices.

Finally, we should point out what the models included in our comparison have in common. First, all are probabilistic. The two Choice Map models were found by Elrod and Winer's (1991) study to outperform the three other mapping methods, which were deterministic. Those methods were correspondence analysis (Hoffman and Franke 1986), which was assessed only on face validity, CGS scaling (Carroll, Green, and Schaffer 1986), and a variant of singular value decomposition. Thus, it appears that probabilistic models offer the best prospects for understanding real world choice data.

Second, all the models considered are examples of empirical Bayes methods (Casella 1985; Morris 1983). That is, all of them avoid estimating parameters at the household level but instead recognize that households differ in their preferences and that the models estimate the parameters of the *distribution* of these preferences. (The SCULPTRE and latent class models accomplish this nonparametrically.)

Current research with the factor-analytic probit model exploits its ability to incorporate as well the effects of observed brand characteristics, observed household characteristics, and their interactions, along with purchase event feedback and nonstationarity in consumer preferences over time (Keane 1994b). The breakthrough in the estimation of probit models from panel data means that many effects which are thought to be important determinants of buying behavior may now be incorporated simultaneously in a single modeling system.

TECHNICAL APPENDIX

This appendix describes a method developed by Keane (1990) for estimating probit models on panel data. We begin with a description of McFadden's (1989) method of simulated moments (MSM), which was designed to estimate probit models from cross-sectional data.

Estimating Probit Models from Cross-Sectional Data

Following McFadden (1989), suppose we observe for a sample of households a single choice from a choice set containing J alternatives. Let i index the household and j index the alternative, let $d_{ij} = 1$ if household i chose brand j (and zero otherwise), let p_{ij} be the probability that the ith household would choose the jth alternative, and let θ be a vector of unknown parameters (none of which are household specific). Then the *score* associated with the loglikelihood function L (θ) is

$$\text{(A1)} \qquad \partial L(\theta)/\partial\theta = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij}[\partial \ln(p_{ij})/\partial\theta].$$

A maximum likelihood (ML) estimator can be obtained by finding that value of θ, which makes the score equal to the zero vector. For the probit model in general, choice probabilities (and functions of these probabilities) are hard to compute, making ML estimation infeasible for all but the simplest cases. However, a method of moments estimator can be obtained from (A1) by exploiting that choice probabilities over mutually exclusive alternatives necessarily sum to one; that is, $\sum_{j=1}^{J} p_{ij} = 1$, which implies

$$\text{(A2)} \qquad \sum_{j=1}^{J} (\partial p_{ij}/\partial\theta) = 0.$$

Because $\partial \ln p_{ij}/\partial\theta = (\partial p_{ij}/\partial\theta)/p_{ij}$, we can substitute $p_{ij}\partial \ln p_{ij}/\partial\theta$ for $\partial p_{ij}/\partial\theta$ in (A2) and then subtract the result from both sides of (A1) to obtain

$$\text{(A3)} \qquad \partial L(\theta)/\partial\theta = \sum_{i=1}^{N} \sum_{j=1}^{J} w_{ij}(d_{ij} - p_{ij}),$$

where

$$\text{(A4)} \qquad w_{ij} = \partial \ln p_{ij}/\partial\theta = (\partial p_{ij}/\partial\theta)/p_{ij}.$$

An estimator of θ that sets the expressions in (A3) equal to zero is known as a method of moments (MOM) estimator, because it sets to zero the (weighted) difference between the observed responses (d_{ij}) and their expected values (p_{ij}). Given optimal weights (A4), MOM estimators are asymptotically equivalent to ML estimators. Given weights that are correlated with those in (A4) and uncorrelated with the residuals in (A3) yields MOM estimates that are consistent and asymptotically normal but not efficient.

The method of *simulated* moments (MSM) simulates the p's in (A3). This is very useful for probit models, because no closed-form expression for the choice probabilities exists for more than two alternatives. For given weights, the equations in (A3) are linear in the choice probabilities, so errors in the simulation of those probabilities tend to cancel over households. No such cancellation occurs when using simulates in ML estimation, so MSM estimation allows much less intensive simulation of the purchase probabilities. The MSM estimates are consistent for a fixed number of simulates per household, which is not the case for simulated ML estimation.

The optimal weights (A4) are also functions of the choice probabilities, but the weights need not be exact nor do they need to be updated on every iteration. Instead, an initial method of simulated moments (MSM) estimator is found, which sets (A3) approximately equal to zero for an initial choice of weights, after which the weights are updated and a more efficient MOM estimator is found. The primary limitation to this approach is that small probabilities are hard to estimate accurately by simulation, and weights involving small probabilities in their denominator can be far from optimal. But the method is feasible for 20 or so alternatives, which is a great improvement over previous methods.

Estimating Probit Models from Cross-Sectional Time Series (Panel) Data

McFadden's MSM is designed for cross-sectional data. When households are observed to make more than one choice, the problem becomes more complicated because purchase probabilities are not independent over time. That is, households observed to have chosen a brand in the past are more likely than other households to choose it again.

McFadden (1989) suggested a method for using MSM to fit probit models to panel data of minimal dimension. His idea is to reexpress panel data in cross-sectional form. For instance, if households are observed to choose from five different brands on two occasions, these data can be reexpressed as a single choice from 25 alternatives. In general, with J brands and T choice occasions, there are J^T different choice sequences. Application of McFadden's procedure requires that the probability of each of these sequences be simulated. For large J and/or T, this will involve an unfeasibly large number of calculations. Thus, this approach becomes unfeasible unless both J and T are small.

Keane (1990) developed a method for estimating probit models on panel data that is feasible for large J and T. The approach is based on a rewriting of (A3) and (A4) in terms of transition probabilities. In essence, the probability that a household makes a particular sequence of choices can be decomposed as the product of conditional probabilities. Letting $C_{it} = j$ signify that the ith household selected alternative j on the tth choice occasion, then (A3) and (A4) can be reexpressed for panel data as

$$\text{(A5)} \qquad \partial L(\theta)/\partial\theta = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{j=1}^{J} w_{ijt}(d_{ijt} - p^*_{ijt})$$

and

$$\text{(A6)} \qquad w_{ijt} = (\partial p^*_{ijt}/\partial\theta)/p^*_{ijt},$$

where

$$\text{(A7)} \qquad p^*_{ijt} = \text{Prob}(d_{ijt} = 1 \mid D_{i,t-1})$$

and $D_{i,t-1}$ denotes all choices made by the household before the tth choice occasion. Note that now the number of probabilities that must be simulated per household is JT rather than J^T. Also note that none of the probabilities on the right-hand side of (A5)–(A7) involves choices from sets larger than J, so the problem of simulating very small probabilities from very large choice sets is avoided.

This new approach generates a new problem: Specifically, simulation of the transition probabilities in (A7) is much more difficult than simulation of the unconditional probabilities in (A3). However, Keane (1990) developed a highly efficient algorithm for the simulation of transition probabilities. Algorithms of the same mathematical form were also developed independently by Geweke (1991) and Hajivassiliou and McFadden (1990). This method has been termed the Geweke-Hajivassiliou-Keane (GHK) simulator by Hajivassiliou, McFadden, and Rund (1992).

The GHK simulator provides accurate simulations of transition probabilities on the basis of simulates generated from truncated normal distributions. Details are provided by Keane (1993, 1994a). The GHK simulator is exploited in this paper to simulate the transition probabilities in (A7). This enabled us to fit the factor-analytic probit model to the detergent data, which includes households that made as many as 24 repeated choices from the set of eight alternatives.

REFERENCES

Albert, James H. and Siddhartha Chib (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Society*, 88 (June), 669–79.

Allenby, Greg M. (1989), "A Unified Approach to Identifying, Estimating and Testing Demand Structures With Aggregate Scanner Data," *Marketing Science*, 8 (Summer), 265–80.

Bass, Frank M., Abel Jeuland, and Gordon P. Wright (1976), "Equilibrium Stochastic Choice and Market Penetration Theories: Derivations and Comparisons," *Management Science*, 22 (June), 1051–1063.

Bozdogan, Hamparsun (1987), "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions," *Psychometrika*, 52 (September), 345–70.

Carroll, J. D. (1972), "Individual Differences and Multidimensional Scaling," in *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, R. N. Shepard, A. K. Romney, and S. Nerlove, eds. New York: Academic Press, 105–55.

———, Geert De Soete, and Wayne S. DeSarbo (1990), "Two Stochastic Multidimensional Choice Models for Marketing Research," *Decision Sciences*, 21 (Spring), 337–56.

———, Paul E. Green, and Catherine M. Schaffer (1986), "Interpoint Distance Comparisons in Correspondence Analysis," *Journal of Marketing Research*, 23 (August), 271–80.

———, ———, and ——— (1987), "Comparing Interpoint Distances in Correspondence Analysis: A Clarification," *Journal of Marketing Research*, 24 (November), 445–50.

Casella, George (1985), "Introduction to Empirical Bayes Data Analysis," *The American Statistician*, 39 (May), 83–7.

Chintagunta, Pradeep K. (1992), "Estimating a Multinomial Probit Model of Brand Choice Using the Method of Simulated Moments," *Marketing Science*, 11 (May), 386–407.

——— (1994), "Heterogeneous Logit Model Implications for Brand Positioning," *Journal of Marketing Research*, 31 (May), 304–11.

——— and B. E. Honore (1990), "Investigating the Effects of Marketing Variables and Unobserved Heterogeneity in a Multinomial Probit Model," working paper, S. C. Johnson Graduate School of Management, Cornell University.

Cooper, Lee G. (1988), "Competitive Maps: The Structure Underlying Asymmetric Cross Elasticities," *Management Science*, 34 (June), 707–23.

DeSarbo, Wayne S. and Jaewun Cho (1989), "A Stochastic Multidimensional Scaling Vector Threshold Model for the Spatial Representation of 'Pick Any/N' Data," *Psychometrika*, 54 (March), 105–30.

———, G. De Soete, J. D. Carroll, and V. Ramaswamy (1988), "A New Stochastic Ultrametric Tree Unfolding Methodology for Assessing Competitive Market Structure and Deriving Market Segments," *Applied Stochastic Models and Data Analysis*, 4, 185–204.

———, ———, and Jehoshua Eliashberg (1987), "A New Stochastic Multidimensional Unfolding Model for the Investigation of Paired Comparison Consumer Preference/Choice Data," *Journal of Economic Psychology*, 8, 357–84.

——— and Donna L. Hoffman (1986), "Simple and Unweighted Unfolding Threshold Models for the Spatial Representation of Binary Choice Data," *Applied Psychological Measurement*, 10 (September), 247–64.

——— and ——— (1987), "Constructing MDS Joint Spaces from Binary Choice Data: A Multidimensional Unfolding Threshold Model for Marketing Research," *Journal of Marketing Research*, 24 (February), 40–54.

——— and Vithala R. Rao (1986), "A Constrained Unfolding Methodology for Product Positioning," *Marketing Science*, 5 (Winter), 1–19.

Elrod, Terry (1988a), "Choice Map: Inferring a Product-Market Map From Panel Data," *Marketing Science*, 7 (Winter), 21–40.

——— (1988b), "Inferring an Ideal-Point Product-Market Map From Consumer Panel Data," in *Knowledge, Data and Computer-Assisted Decisions*, Martin Schader and Wolfgang Gaul, eds. Heidelberg: Springer-Verlag, 240–49.

——— (1991), "Internal Analysis of Market Structure: Recent Developments and Future Prospects," *Marketing Letters*, 2, 253–66.

——— and Russell S. Winer (1991), "An Empirical Comparison of Mapping Methods Based on Panel Data," working paper, University of Alberta.

Geweke, J. (1991), "Efficient Simulation From the Multivariate Normal and Student-t Distributions Subject to Linear Constraints," in *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*. Alexandria, VA: American Statistical Association, 571–78.

———, Michael Keane, and David Runkle (1994), "Alternative Computational Approaches to Inference in the Multinomial Probit Model," *Review of Economics and Statistics*, forthcoming.

Grover, Rajiv and V. Srinivasan (1987), "Simultaneous Approach to Market Segmentation and Market Structuring," *Journal of Marketing Research*, 24 (May), 139–53.

Hajivassiliou, V. and D. McFadden (1990), "The Method for Simulated Scores for the Estimation of LDV Models With an Application to External Debt Crises," Cowles Foundation Discussion Paper 967.

———, ———, and P. Ruud (1992), "Simulation of Multivariate Normal Orthant Probabilities: Theoretical and Computational Results," Cowles Foundation Discussion Paper 1021.

Hannan, E. J. and B. G. Quinn (1979), "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society, Series B*, 41, 190–95.

Hauser, John R. (1978), "Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information-Theoretic Approach," *Operations Research*, 26 (May-June), 406–21.

——— and Steven M. Shugan (1983), "Defensive Marketing Strategies," *Marketing Science*, 2 (Fall), 319–60.

Heckman, J. and B. Singer (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52 (March), 271–320.

Hoffman, Donna L. and George R. Franke (1985), "Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research," *Journal of Marketing Research*, 23 (August), 213–27.

Holman, Eric W. (1972), "The Relation Between Hierarchical and Euclidean Models for Psychological Distances," *Psychometrika*, 37 (December), 417–23.

Jain, Dipak, Frank M. Bass, and Yu-Min Chen (1990), "Estimation of Latent Class Models With Heterogeneous Choice Probabilities: An Application to Market Structuring," *Journal of Market-*

ing Research, 27 (February), 94–101.

Jedidi, Kamel and Wayne S. DeSarbo (1991), "A Stochastic Multidimensional Scaling Procedure for the Spatial Representation of Three-Mode, Three-Way Pick Any/J Data," *Psychometrika*, 56 (September), 471–94.

Johnson, Norman L. and Samuel Kotz (1972), *Distributions in Statistics: Continuous Multivariate Distributions*. New York: John Wiley & Sons, Inc.

Kamakura, Wagner A. and Gary J.Russell, (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (November), 379–90.

Keane, Michael P. (1990), *Four Essays in Empirical Macro and Labor Economics*, unpublished doctoral dissertation, Brown University.

——— (1993), "Simulation Estimation for Panel Data Models With Limited Dependent Variables," in *Handbook of Statistics, 11: Econometrics*, G. S. Madalla, C. R. Rao, and H. D. Vinod, eds. Amsterdam: Elsevier Science Publishers, 545–72.

——— (1994a), "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62 (January), 95–116.

——— (1994b), "Modelling Heterogeneity and State Dependence in Consumer Choice Behavior," working paper, Industrial Relations Center and Department of Economics, University of Minnesota.

Laird, Nan (1978), "Nonparametric Maximum Likelihood Estimation of a Mixing Distribution," *Journal of the American Statistical Association*, 73 (December), 805–11.

Levine, Joel H. (1979), "Joint-Space Analysis of 'Pick-Any' Data: Analysis of Choice from an Unconstrained Set of Alternatives," *Psychometrika*, 44, 85–92.

MacKay, David B. and Cornelia Dröge (1990), "Extensions of Probabilistic Perceptual Maps With Implications for Competitive Positioning and Choice," *International Journal of Research in Marketing*, 7, 265–82.

——— and Joseph L. Zinnes (1986), "Probabilistic Model for the Multidimensional Scaling of Proximity and Preference Data," *Marketing Science*, 5 (Fall), 325–44.

McCulloch, R. and Peter E. Rossi (1994), "An Exact Likelihood Analysis of the Multinomial Probit Model," *Journal of Econometrics*, 64, 207–40.

McFadden, Daniel (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 57 (September), 995–1026.

Morris, Carl N. (1983), "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 78 (March), 47–54.

Ramaswamy, Venkatram and Wayne S. DeSarbo (1990), "SCULPTRE: A New Methodology for Deriving and Analyzing Hierarchical Product-Market Structures From Panel Data," *Journal of Marketing Research*, 27 (November), 418–27.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–64.

Shugan, Steven M. (1987), "Estimating Brand Positioning Maps Using Supermarket Scanning Data," *Journal of Marketing Research*, 24 (February), 1–19.

Teräsvirta, T. and I. Mellin (1989), "Linear Model Selection, Criteria and Tests," in *Encyclopedia of the Statistical Sciences*, Suppl. Vol., Samuel Kotz and Norman L. Johnson, eds. New York: John Wiley & Sons, Inc, 83–7.

Waarts, Eric, Martin Carree, and Berend Wierenga (1991), "Full-Information Maximum Likelihood Estimation of Brand Positioning Maps Using Supermarket Scanning Data," *Journal of Marketing Research*, 28 (Winter), 483–90.

Winsberg, Suzanne and J. Douglas Carroll (1988), "A Quasi-Nonmetric Method for Multidimensional Scaling via an Extended Euclidean Model," *Psychometrika*, 53 (June), 217–29.