



ELSEVIER

Journal of Econometrics 89 (1999) 57–78

JOURNAL OF
Econometrics

Marketing models of consumer heterogeneity

Greg M. Allenby^{a,*}, Peter E. Rossi^b

^aMax M. Fisher College of Business, Ohio State University, 1775 College Road, Columbus,
OH 43210, USA

^bGraduate School of Business, University of Chicago, Chicago, IL 60637, USA

Abstract

The distribution of consumer preferences plays a central role in many marketing activities. Pricing and product design decisions, for example, are based on an understanding of the differences among consumers in price sensitivity and valuation of product attributes. In addition, marketing activities which target specific households require household level parameter estimates. Thus, the modeling of consumer heterogeneity is the central focus of many statistical marketing applications. In contrast, heterogeneity is often regarded as an ancillary nuisance problem in much of the applied econometrics literature which must be dealt with but is not the focus of the investigation. The focus is instead on estimating average effects of policy variables. In this paper, we discuss various approaches to modeling consumer heterogeneity and evaluate the utility of these approaches for marketing applications. © 1999 Elsevier Science S.A. All rights reserved.

JEL classification: C11; C33; C35

Keywords: Random effects; Heterogeneity; Probit models

The purpose of marketing is to understand consumer preferences and to help design and deliver appropriate goods and services. Marketers are interested in determining what products to offer, what prices to charge, how the products should be promoted and how to best deliver the products to the consumer. One of the greatest challenges in marketing is to understand the diversity of preferences and sensitivities that exists in the market. Heterogeneity in preferences gives rise to differentiated product offerings, market segments and market

* Corresponding author. E-mail: gm + @osu.edu, peter.rossi@gsb.uchicago.edu

niches. Differing sensitivities are the basis for targeted communication programs and promotions. As consumer preferences and sensitivities become more diverse, it becomes less and less efficient to consider the market in the aggregate.

In contrast to this emphasis on individual differences, economists are often more interested in aggregate effects and regard heterogeneity as a statistical nuisance parameter problem which must be addressed but not emphasized. Econometricians frequently employ methods which do not allow for the estimation of individual-level parameters. For example, random coefficient models are often implemented through an unconditional likelihood approach in which only hyper-parameters are estimated. Furthermore, the models of heterogeneity considered in the econometrics literature often restrict heterogeneity to subsets of parameters such as model intercepts. In the marketing context, there is no reason to believe that differences should be confined to the intercepts and, as indicated above, differences in slope coefficients are critically important. Finally, economic policy evaluation is frequently based on estimated hyper-parameters which are measured with much greater certainty than individual-level parameters. This is in contrast to marketing policies which often attempt to respond to individual differences that are measured less precisely. The determination of optimal marketing decisions must account for the substantial uncertainty that exists in individual-level estimates, and the impact of this uncertainty on decision criteria which often involve non-linear functions of model parameters.

Marketing practices which are designed to exploit consumer differences require flexible models of heterogeneity coupled with an inference method which adequately describes uncertainty in consumer level estimates. The goal of this paper is to outline an approach to this problem which is the culmination of developments from a series of papers (Rossi and Allenby, 1993; McCulloch and Rossi, 1994; Rossi, McCulloch and Allenby, 1996). We advocate a continuous model of heterogeneity and employ a Bayesian approach to inference. We contrast this approach with discrete models of heterogeneity popular in the marketing literature. We also comment on non-Bayesian methods of inference which dominate the econometrics literature. An example of price promotions targeted to the household is used to illustrate the differences between different models of heterogeneity and methods of inference.

The remainder of the paper is organized as follows: Section 1 examines issues in modeling heterogeneity, contrasting various approaches that appear in the econometric and marketing literature. Section 2 introduces the discrete choice model used to explore the heterogeneity issues, and Section 3 provides an empirical application using a scanner panel dataset of household ketchup purchases. Section 4 then discusses various marketing policies related to the analysis of this type of data, and illustrates the benefit of a detailed understanding of the extent and nature of heterogeneity. Section 5 offers concluding remarks.

1. Issues in modeling heterogeneity

Data describing consumer preferences and sensitivities to variables such as price are typically obtained through surveys or household purchase histories which yield very limited individual-level information. For example, household purchases in most product categories often total less than 12 per year. Similarly, survey respondents become fatigued and irritable when questioned for more than 20 or 30 min. As a result, the amount of data available for drawing inferences about any specific consumer is very small, although there may exist many consumers in a particular study.

The fixed-effects approach to heterogeneity has much appeal since it delivers the individual household level parameter estimates and does not require the specification of any particular probability distribution of heterogeneity. However, the sparseness of individual-level data renders this approach impractical. In many situations, incomplete household-level data causes a lack of identification at the household level. In other cases, the parameters are identified in the household level likelihood but the fixed effects estimates are measured with huge uncertainty which is difficult to quantify using standard asymptotic methods.

1.1. Inference in random coefficient models

An alternative approach is to specify a random-effects model which stochastically pools data across consumers (Heckman, 1982). This approach is frequently employed in the analysis of panel data and has been extensively applied in both marketing and economic studies. Examples include Chamberlain (1984), Heckman and Singer (1984), Kamakura and Russell (1989), Chintagunta et al. (1991), and Gonul and Srinivasan (1993). In the random effects model, individual household level parameters are viewed as draws from a super-population distribution which is termed the random effects or mixing distribution. Traditionally, the random effects distribution is viewed as being a part of the likelihood. To make this discussion more concrete, we now introduce a generic notation for this problem. The likelihood of the household-specific parameters $\{\alpha_i\}$ and the common parameters of the mixing distribution can be written as

$$\ell(\{\alpha_i\}, \beta) \equiv p(\text{data}|\{\alpha_i\}, \beta) = \prod_{i=1}^N p(\text{data}|\alpha_i)\pi(\alpha_i|\beta). \quad (1)$$

Here ‘ i ’ denotes the i th of N consumers, ℓ denotes the likelihood, α_i is the vector of individual parameters, and $\pi(\alpha_i|\beta)$ is the random effects distribution indexed by the parameter β . In most econometric applications, interest focuses on the common parameters, β , of the super-population distribution. For example in an economic analysis, we might be interested in the average effects of a policy

intervention. Inference about the super-population parameters is conducted with the average conditional likelihood or the marginal likelihood which is obtained by integrating out the $\{\alpha_i\}$.

$$\ell(\beta) \equiv p(\text{data}|\beta) = \prod_{i=1}^N \int p(\text{data}_i|\alpha_i)\pi(\alpha_i|\beta) d\alpha_i. \quad (2)$$

Outside of normal linear models with a normal random coefficient distribution performing the integral in Eq. (2) can be computationally challenging. Recent advances in simulation methods have considerably expanded the set of models for which it is now possible to conduct marginal likelihood inference using Eq. (2) (see Hajivassiliou and Ruud, 1994; Mariano et al., 1997).

In the econometric literature, little attention has been paid to the problem of obtaining household level estimates. Unfortunately, many marketing problems require knowledge of the parameters of particular customers. For example, marketers are often interested in targeting their communications and offers to subpopulations who they expect will react most favorably. The identification and characterization of these consumers requires knowledge of α_i as well as β . In this and other cases, knowledge of the distribution of heterogeneity is insufficient to identify optimal marketing actions. An approximate Bayesian approach to this problem would be to estimate β by maximizing Eq. (2) and then using $\pi(\alpha_i|\beta = \hat{\beta})$ as a prior in the analysis of an individual household conditional likelihood.

$$p(\alpha_i|\text{data}) \propto p(\text{data}_i|\alpha_i)\pi(\alpha_i|\beta = \hat{\beta}). \quad (3)$$

This approximation is justified when there are enough consumers in the study such that β is precisely estimated and any one consumer does not have a large influence on its estimate. In many problems such as the choice model application considered in Section 2 below, this analysis of household level likelihoods would require numerical integration methods as the prior and likelihood will not ‘mix’ to form a simple posterior (see Rossi and Allenby, 1993 for an illustration of this approach). In this paper, we outline a comprehensive approach to Bayesian analysis of this problem which yields both household level and common parameter inferences without resort to approximate methods.

In theory, information about individual parameters can be obtained from the random-effects model which does not integrate the likelihood as in Eq. (2), and instead considers the joint distribution of $\{\alpha_i\}$ and β . This can be accomplished by recasting Eq. (1) as a hierarchical Bayes model:

$$\pi(\{\alpha_i\}, \beta|\text{data}) \propto \prod_{i=1}^N l(\text{data}_i|\alpha_i)\pi(\alpha_i|\beta)\pi(\beta), \quad (4)$$

where $\pi(\beta)$ is a prior distribution placed on β so that the joint (posterior) distribution is defined. In this Bayesian formulation, the mixing distribution is considered as part of the prior distribution. Inferences about the preferences and sensitivities of a specific individual are obtained by appropriately marginalizing the joint posterior distribution:

$$\pi(\alpha_j|\text{data}) = \int \pi(\{\alpha_i\}, \beta|\text{data}) d\alpha_{-j} d\beta, \quad (5)$$

where ‘ $-j$ ’ reads ‘all units except j .’ The advantage of this approach is that all the data in the study are used to derive individual-level estimates. This helps insure that the parameters are identified and results in more accurate household parameter estimates via parameter shrinkage in which the random effects distribution is used a prior and the household parameter estimates are ‘shrunk’ toward the mean.

In addition, Eq. (5) provides information about the entire distribution of α_j and not just some measure of central tendency. A common practice in many studies in marketing and economics is to employ point estimates of model parameters when considering the implications of the estimated parameters. For example, the effect of a general change in price can often be accurately evaluated in linear models by considering the average price sensitivity as reflected in β . However in nonlinear models, and in situations where the average affects are not of primary interest then the impact of parameter uncertainty needs to be considered. Individual-level estimates are much less accurate than aggregate-level estimates, and this uncertainty needs to be explicitly considered when designing marketing activities which are tailored to specific individuals.

1.2. *Models of heterogeneity: Continuous vs. discrete distributions*

Until this point, we have not specified the form of the random coefficient or mixing distribution, $\pi(\cdot)$. It has become popular in the marketing literature to specify a discrete distribution for $\pi(\alpha|\beta) = \sum_{j=1}^J \phi_j I(\alpha = \beta_j)$, often termed a finite mixture model (c.f. Titterton et al., 1985; Kamakura and Russell, 1989; Chintagunta et al., 1991; Heckman and Singer, 1984). The finite mixture approach is popular in part due to the fact that the marginal likelihood is easily evaluated as a sum over the J mass points. Heckman and Singer (1984) have emphasized the flexibility of the discrete distribution in that, for sufficient mass points, any distribution can be approximated to a high degree of accuracy.

In practice, however, researchers find it difficult to estimate finite mixture models with more than a half dozen or so mass points. While the finite mixture model may approximate the central tendencies of the mixing distribution, there is a growing body of evidence which suggests that finite mixture models with

a small number of mass points may inadequately capture the full extent of heterogeneity in the data. A disadvantage of the finite mixture model is that the posterior distribution of individual-level parameters is constrained to lie with the convex hull of β_j . As a result, these finite mixture models lead to posterior estimates of individual effects which are much less heterogeneous than those obtained from continuous mixing distributions (see Allenby and Ginter, 1995b; Lenk et al., 1996; Rossi et al., 1996).

Recent advances in Markov Chain Monte Carlo (MCMC) estimation now permit the evaluation of continuous random effect models and the exact posterior distribution of individual effects Eq. (5). In addition, MCMC methods provide a new tool for analysis of certain discrete choice models for which the likelihood itself is computationally challenging to evaluate. To illustrate these points, we introduce a discrete choice model of household purchases in the next section and contrast it to alternative formulations found in the economics and marketing literature.

2. Approaches to incorporating heterogeneity in choice models

In this section, we discuss approaches to modeling heterogeneity in the context of discrete choice models. The basic model we will employ in this discussion is an independence Probit in a household panel data setting. We observe $I_{h,t}$, the choices of H households, each for T_h periods. $I_{h,t}$ is a multinomial outcome conditional on a set of covariates, $X_{h,t}$. In the standard random utility framework, these multinomial outcomes are thought to result from an underlying latent regression model.

$$y_{h,t} = X_{h,t}\beta_h + \varepsilon_{h,t} \quad \varepsilon_{h,t} \sim \text{iid } N(0, A). \quad (6)$$

We assume there are m possible brand choices and that households choose the brand with the highest utility. $I_{h,t}$ is the index of the maximum of the elements of $y_{h,t}$. $X_{h,t}$ is a matrix of choice characteristics which includes an intercept term for each of the m brands and price, display, and feature variables, $X_{h,t} = [I_m, p_{h,t}, d_{h,t}, f_{h,t}]$ where I_m is the $m \times m$ identity matrix, $p_{h,t}$ is an m vector of log prices, $d_{h,t}$ is an indicator vector of length m such that the i th element is 1 if the i th brand is on display and 0 otherwise, and $f_{h,t}$ is an indicator vector for feature. β_h is a vector representing the household h 's preferences and sensitivity to marketing mix variables and $\varepsilon_{h,t}$ is an error term. Given the short length of our panel data (the typical household is observed for at a maximum of 1.5 yr), it seems reasonable to assume that the β_h parameters do not vary over time; that is, we assume that brand preferences and marketing mix sensitivities are not time-varying.

Different specifications of the error structure result in various probit (normal error) and logit (extreme value error) models. We use a diagonal covariance

structure $\varepsilon_{h,t} \sim \text{iid } N(0, A)$, where A is a $m \times m$ diagonal matrix. A scalar covariance structure (equal variances for each choice alternative random utility errors) yields a Probit model which displays the restrictive IIA property (Hausman and Ruud, 1987; Allenby and Ginter, 1995a). Finally, we should note that if we specify a random effects model for the β_h parameters, then we have specified a complicated correlated probit model with heteroskedastic errors (see Section 2.2 for further elaboration of this point). Below we consider alternative approaches to modeling heterogeneity in household preferences and sensitivities.

2.1. Bayesian approaches

Consider a continuous random coefficient model in which the $\{\beta_h\}$ are drawn from a multivariate normal distribution:

$$\beta_h \sim \text{iid } N(\bar{\beta}, V_\beta). \quad (7)$$

While the normal distribution is a reasonable starting point for a continuous distribution of heterogeneity, more flexible distributions are possible. For example, Allenby et al. (1995b) and Rossi et al. (1996) introduce observable household characteristics into the specification of $\bar{\beta}$. McCulloch and Rossi (1996) outline a strategy for using a mixture of normals to provide additional flexibility (see also, Allenby et al., 1998). In Section 3 below, we consider some diagnostic evidence regarding the appropriateness of the normal random coefficient distribution in our data context.

As discussed in Section 1, we introduce priors for the parameters to insure that the posterior distribution is defined. For convenience, we use natural conjugate priors in which the prior on $\bar{\beta}$ is normal and the prior on V_β is inverted Wishart:

$$\bar{\beta} \sim N(\bar{\bar{\beta}}, aV_\beta) \quad (8)$$

and

$$V_\beta^{-1} \sim W(v_0, V_0). \quad (9)$$

In our empirical analysis, we employ relatively diffuse priors for $\bar{\beta}$ and V_β since we do not have direct information on these quantities. More specifically, we specify these priors with the following parameters: $\bar{\bar{\beta}} = 0$, $a = 10$, $v_0 = k + 4$, $V_0 = v_0 I$. In our analysis below, we found our results to be insensitive to doubling a and v_0 , indicating that we have indeed specified a prior which is diffuse relative to the likelihood.

To complete the Bayesian model, we must formulate a prior for the elements of Λ . We use the standard natural conjugate Inverted Gamma priors where:

$$A = \text{diag}(1, \lambda_2, \dots, \lambda_m) \sqrt{\lambda_i} \sim \text{IG}(v, s_i). \quad (10)$$

Again, we employ very diffuse settings for these IG priors, letting $v = 3$ and $s_i = 1.0$. Note that we have set $\lambda_1 = 1$ for identification purposes.

Our complete hierarchical model is specified by a series of conditional distributions. At the bottom of the hierarchy is the latent utility regression model *conditional* on β_h . The random utility regression is followed by successively higher levels of priors which incorporate views about the distribution of the β_h coefficients. Here we use the notation that $y|x$ is the conditional distribution of y given x .

$$I_{h,t}|y_{h,t}, \quad (11)$$

$$y_{h,t}|X_{h,t}, \beta_h, A, \quad (12)$$

$$\beta_h|\bar{\beta}, V_\beta, \quad (13)$$

$$\bar{\beta}|\bar{\beta}, aV_\beta, \quad (14)$$

$$V_\beta|v_0, V_0, \quad (15)$$

$$A|n, s, \quad (16)$$

where $y'_h = (y'_{h,1}, \dots, y'_{h,T_h})$ and $X'_h = [X'_{h,1}, \dots, X'_{h,T_h}]$. Eqs. (11)–(16) give the conditional likelihood for β_h and the random coefficient model. The hierarchy is constructed by these sets of conditional distributions which combine to specify the model. From a Bayesian perspective, the random coefficient model in Eq. (4) is part of a prior in which the household coefficients are viewed as having some commonality through the mixing or random coefficient distribution.

Our goal in the analysis of household choice data is not only to describe the extent and nature of household heterogeneity but also to make inferences about specific households for the purpose of customizing various marketing actions. Fortunately, Bayesian analysis of hierarchical models with Markov Chain Monte Carlo methods has made it feasible to conduct inference about both household level and common random coefficient parameters.

Bayesian analysis of hierarchical models has been made feasible by the development of Markov chain simulation methods which directly exploit the hierarchical structure (see Gelfand and Smith, 1990; Gelfand et al., 1990; Tierney, 1991 for general discussion of these methods). The basic idea behind

these methods is to construct a Markov chain which has the posterior as its stationary or invariant distribution and then simulate the chain to obtain a sequence of draws which can be used to approximate the posterior to any desired degree of accuracy. In this paper, we use the Gibbs sampler constructed for the hierarchical MNP model by McCulloch and Rossi (1994) (Allenby and Lenk (1994) consider a logistic normal regression model applied to scanner panel data). The Gibbs sampler is implemented by drawing successively from the following set of posterior distributions which are based on the data consisting of the X_{ht} explanatory variables and $I_{h,t}$ (the index of observed choices). The exact forms for these conditional distributions are given in McCulloch et al. (1995). Our Gibbs sampler proceeds by drawing successively from each of the distributions above and iterating this procedure to obtain a long sequence of draws. These draws are then used to compute the marginal posterior distribution of various quantities of interest. There are a number of technical issues which arise in using these sorts of procedures (see McCulloch and Rossi, 1994; Gelman and Rubin, 1992 for a thorough discussion of these issues).

2.2. Classical approaches

Parametric random coefficient models have been widely used in econometrics but usually from the classical point of view. This means that the model of heterogeneity laid out in Section 2.1 is used to average the conditional likelihood given β_h (see Eq. (1)). In the probit model with a normal random coefficient distribution, substitution of the random coefficient distribution into the latent variable regression results in a correlated probit with a heteroskedastic error structure (as in Hausman and Wise, 1978)

$$y_{h,t} = X_{h,t}(\bar{\beta} + v_h) + \varepsilon_{h,t}, \quad (17)$$

$$y_{h,t} = X_{h,t}\bar{\beta} + u_{h,t}. \quad (18)$$

Here $u_{h,t}$ has the following covariance structure:

$$\text{Var}(u_{h,t}) = \begin{bmatrix} \Lambda + X_{h,1}V_{\beta}X'_{h,1} & \dots & X_{h,1}V_{\beta}X'_{h,T_h} \\ & \ddots & \vdots \\ & & \Lambda + X_{T_h,T_h}V_{\beta}X'_{T_h,T_h} \end{bmatrix}. \quad (19)$$

This results in a very high dimensional correlated probit problem with a $(m-1)T_h$ dimensional integrals required to evaluate the unconditional probit likelihood. Advances in evaluation of these integrals via simulation methods have made the method of simulated maximum likelihood feasible for problems with small to moderate values of T_h . It is important to insure that adequate

numbers of draws are used in evaluating the likelihood for the method of simulated maximum likelihood and the method of simulated scores (see Geweke et al., 1996). Without sampling experiments such as those pursued in Geweke et al., it is hard to establish, a priori, adequate number of simulations in order to assure good performance of maximum likelihood methods.

In many marketing contexts, we desire estimates of the household level parameter vector in order to tailor marketing actions to specific households. The standard classical approach to random effects models does not provide these estimates automatically. An approximate Bayes procedure can be developed to make inferences at the household level. As shown in Eq. (3) we can use the random coefficient model conditional on the maximum likelihood estimates of the super-population or common parameters as a prior in the analysis of each households likelihood function.

$$p(\beta_h | X_h, y_h) \propto p(y_h | X_h, \beta_h, \hat{A}) p(\beta_h | \hat{\beta}, \hat{V}_\beta). \quad (20)$$

This is an approximate Bayesian analysis since it is conditional on the estimated parameters and the ‘prior’ on β_h is based on all of the data including the data for this household. It should be noted that numerical integration methods will have to be used to find the posterior mean of β_h using the approximate posterior above. Thus, the classical approach for inferring about household parameters is substantially more computationally demanding than the full Bayesian approach and offers only approximate answers.

3. Empirical application

The hierarchical Bayes and finite mixture models described above were fit to a scanner panel dataset of ketchup purchases of households residing in Springfield, MO. Four brands of ketchup in 32 oz containers were considered in our analysis: Heinz, Hunt’s, Del Monte and a House Brand. The data is comprised of 8191 purchases by 1401 households during an 18 month period beginning in 1986. Explanatory variables include price and dummy variables for the presence of display and feature advertising activity. Summary statistics are provided in Table 1. Identification of household level parameters is informationally demanding and it is encouraging to see the large standard deviations of the price variables which is produced by frequent price promotions in which prices are reduced by 25% or more.

3.1. Comparison with finite mixture approaches

Parameter estimates from the finite mixture model are provided in Table 2. The number of mixing components were determined by examining the Bayesian

Table 1
Descriptive statistics

Brand	Choice share	Avg price (SD) (\$)	% of Time	
			Displayed	Featured
Heinz	0.429	1.239 (0.28)	0.082	0.132
Hunt's	0.251	1.263 (0.25)	0.110	0.057
Del Monte	0.103	1.287 (0.33)	0.056	0.010
House	0.217	0.775 (0.10)	0.086	0.047

Table 2
Finite mixture model parameter estimates(standard errors)

	Components					
	1	2	3	4	5	6
Mass	0.13	0.29	0.06	0.25	0.20	0.07
<i>Parameter</i>						
Hunt's	− 3.19 (0.43)	− 0.66 (0.08)	0.94 (0.12)	0.17 (0.07)	0.26 (0.09)	− 0.46 (0.14)
Del Monte	− 3.68 (0.51)	− 1.48 (0.10)	− 1.05 (0.29)	− 0.20 (0.08)	− 0.48 (0.13)	0.34 (0.10)
House	− 7.61 (0.97)	− 3.76 (0.20)	− 3.01 (0.38)	− 3.53 (0.21)	0.02 (0.14)	− 1.53 (0.26)
ln Price	− 3.89 (0.95)	− 3.11 (0.19)	− 2.02 (0.32)	− 4.63 (0.28)	− 3.87 (0.12)	− 3.33 (0.33)
Display	0.57 (0.32)	1.03 (0.11)	0.55 (0.25)	1.59 (0.13)	1.37 (0.12)	0.46 (0.19)
Feature	0.19 (0.23)	1.51 (0.15)	0.65 (0.30)	0.76 (0.12)	1.26 (0.14)	0.16 (0.22)
<i>Lambda (exp[λ])</i>						
Hunt's	− 0.62 (0.12)					
Del Monte	− 0.48 (0.09)					
House	0.48 (0.06)					

Information Criteria (Schwarz, 1978). In general, the parameter estimates differ greatly across components, indicating the existence of consumer heterogeneity. For example, preferences for Hunt's range from − 3.19 to 0.94, while sensitivity to (log) price ranges from − 4.63 to − 2.02. Components one, three, five and six

describe consumers who prefer each of the four brands, Heinz, Hunt's, House and Del Monte, respectively. Component two describes consumers who prefer either Heinz or Hunt's and are display and feature sensitive. Component four describes consumers with approximately equal intercept estimates for the national brands and large sensitivity to price, in-store displays and, to a lesser extent, feature advertising. Consumers described by this latter component frequently switch between the national brands.

Parameter estimates for the common parameters of the hierarchical Bayes model are reported in Table 3. The mean of the normal random effects distribution is reported on the left of the table, and the covariance matrix of random effects is reported on the right. The upper right portion of the matrix displays the corresponding correlations. The negative estimates for the brand names indicate that Heinz is the preferred brand, consistent with the choice shares reported in Table 1. Estimates of the covariance matrix of the random effects distribution reveal many large diagonal elements, indicating the extent of heterogeneity in brand preferences and sensitivities to price, display and feature. When compared to the range of estimates from the finite mixture model, the hierarchical Bayes estimates indicate a much larger degree of heterogeneity in the market. For example, a model-based estimate of the 95% HPD interval for price sensitivity ranges from -9.22 to -0.22 , which is more than two times larger than that obtained with the finite mixture model. Unless the actual distribution of heterogeneity has a pronounced multi-modal shape, the low dispersion of the finite mixture model distribution is an indication of an inadequate approximation. Below, we present diagnostic evidence that suggests that our assumption of a unimodal heterogeneity distribution is justified.

In a Bayesian setting, we can formally compare the models by computing the posterior probability of each of a set of models. The posterior probability of

Table 3
Hierarchical Bayes model parameter estimates (posterior standard deviation)

Parameter	Mean	Covariance\correlation matrix					
Hunt's	- 0.50 (0.07)	1.82 (0.21)	0.67	0.47	- 0.32	- 0.08	0.02
Del Monte	- 1.33 (0.11)	1.37 (0.18)	2.25 (0.27)	0.57	- 0.39	0.13	- 0.07
House Brand	- 4.52 (0.28)	2.12 (0.38)	2.81 (0.47)	10.94 (1.49)	- 0.07	0.20	0.20
ln Price	- 4.72 (0.25)	- 0.98 (0.30)	- 1.32 (0.35)	- 0.49 (0.83)	5.05 (0.99)	- 0.01	0.08
Display	1.69 (0.13)	0.14 (0.19)	0.25 (0.21)	0.87 (0.43)	- 0.04 (0.41)	1.76 (0.39)	0.33
Feature	1.52 (0.14)	0.03 (0.18)	- 0.14 (0.21)	0.87 (0.48)	0.23 (0.42)	0.57 (0.25)	1.69 (0.43)

a model is proportional to the marginal density of the data (we use θ to denote the parameters of the model).

$$p(M | \text{Data}) \propto p(\text{Data}) = \int p(\theta | \text{Data}) p(\theta) d\theta. \quad (21)$$

We compute the marginal density of the data for the mixture model by using the popular Schwarz asymptotic approximation, $\ln(p(\text{Data})) \doteq \ln(p(\text{Data} | \theta = \hat{\theta}_{MLE})) - k/2 \ln(N)$. K is the number of parameters in the model and N is the number of observations. For the Hierarchical Bayes model, we can employ a more accurate approximation by using Importance Sampling to compute the integral in Eq. (21) as suggested by Newton and Raftery, 1994, p. 21). When we compute the marginal densities of the data for the finite mixture and continuous mixture HB models, it is clear that the HB model fits the data much better and has a much higher posterior probability:

$$p(\text{Data} | M = \text{finite mixture}) = -5998.3,$$

$$p(\text{Data} | M = \text{continuous mixture}) = -3928.3.$$

Disaggregate estimates of brand preference for Hunt's and price sensitivity are displayed in Fig. 1 for a selected subset of households. The MCMC estimate of

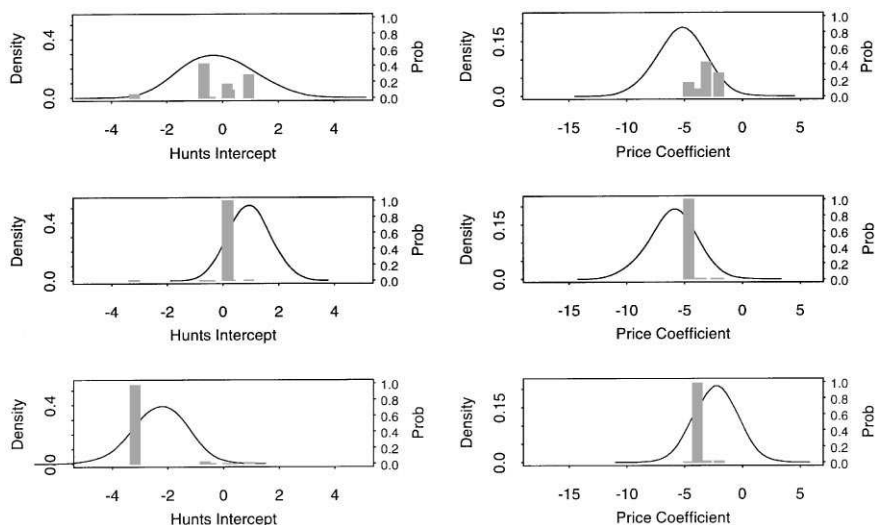


Fig. 1. Comparison of household parameter posterior distributions: continuous vs. finite mixture models.

the posterior marginal density of the parameter from the hierarchical Bayes model is shown by the solid line. By conditioning on the MLE's of the finite mixture model parameters, we are able to compute an approximation to the posterior distribution of household parameters by applying Bayes theorem conditional on the parameter estimates (see also Kamakura and Russell, 1989). This discrete approximate posterior distribution is depicted by the bar charts superimposed over the density plots in Fig. 1.

What stands out in Fig. 1 is the large discrepancies between the marginal distributions obtained from a continuous mixture approach and those obtained using the finite mixture approach. The continuous mixture marginals show dramatically greater dispersion in contrast to the finite mixture distributions which, in many cases, put virtually all posterior mass on one value. It is our view that these finite mixture approximations are unrealistically tight, conveying a misleading impression about the extent of sample information regarding household level parameters. This discrepancy becomes problematic when considering marketing policies (e.g. temporary price reductions) which are primarily driven by specific parameters (e.g. the price coefficients).

Fig. 2 compares disaggregate estimates of model parameters for the hierarchical Bayes and finite mixture models. Plotted are the posterior means of model parameters for 200 randomly selected households in the panel. The plot illustrates that the finite mixture model is very restrictive in that posterior estimates are constrained to lie within the convex hull of the estimated mass points. The

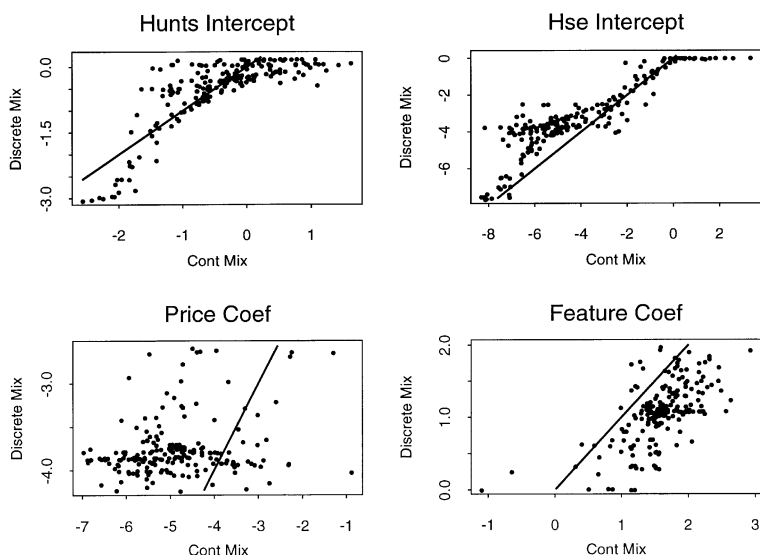


Fig. 2. HB vs. finite mixture household estimates (posterior means).

finite mixture model therefore does a poor job of describing tail behavior of the distribution of heterogeneity, which is often of considerable interest in marketing problems. This issue is discussed in more detail in Section 4 below.

3.2. *Diagnostic checking of the prior*

In our Bayesian approach, priors play a central role in producing household level inferences. In the hierarchical model, the prior is specified in a two stage process:

$$\beta_h \sim N(\bar{\beta}, V_\beta), \quad p(\bar{\beta}); p(V_\beta). \quad (22)$$

In the classical literature, the normal distribution of β_h called the random effects model would be considered part of the likelihood rather than part of the prior. As discussed below, the assumption of a normal distribution could be problematic here. The second stage of the prior is also of potential concern. At issue is the tightness of the second stage priors which determines the amount of shrinkage. We use extremely diffuse (barely proper) priors here. In order to check prior sensitivity, we have considerably tightened (doubled the hyperparameters) in the second stage and found that the posterior distribution of the common parameters (displayed in Table 3) does not change at all (up to simulation error in the Gibbs draws). The huge number of households (1401) used here is undoubtedly the reason why large changes in the second stage prior make no material difference. Thus, it is the first stage prior which is important and will always remain important as long as there are only a few observations available per household. Since the parameters of the first stage prior are inferred from the data, the main focus of concern should be on the form of this distribution.

In the econometric literature, the use of parametric distributions of heterogeneity (e.g. normal distributions) are often criticized on the grounds that their mis-specification leads to inconsistent estimates of the common model parameters (cf. Heckman and Singer, 1984). For example, if the true distribution of household parameters were skewed or bimodal, our inferences based on the symmetric, unimodal normal prior could be misleading. One simple approach would be to plot the distribution of the posterior household means and compare this to the implied normal distribution evaluated at the Bayes estimates of the hyperparameters, $N(E[\bar{\beta}|\text{data}], E[\text{Var}(\beta_h)])$. The posterior means are not constrained to follow the normal distribution since the normal distribution is only part of the prior and the posterior is influenced by household data. This simple approach is in the right spirit but could be misleading since we do not properly account for uncertainty in the household parameter estimates, $\{E[\beta_h|\text{data}]\}$, $E[\bar{\beta}|\text{data}], E[\text{Var}(\beta_h)|\text{data}]$.

Fig. 3 provides a diagnostic check of the assumption of normality in the first stage of the prior distribution which properly accounts for parameter uncertainty. To handle uncertainty in our knowledge of the common parameters of

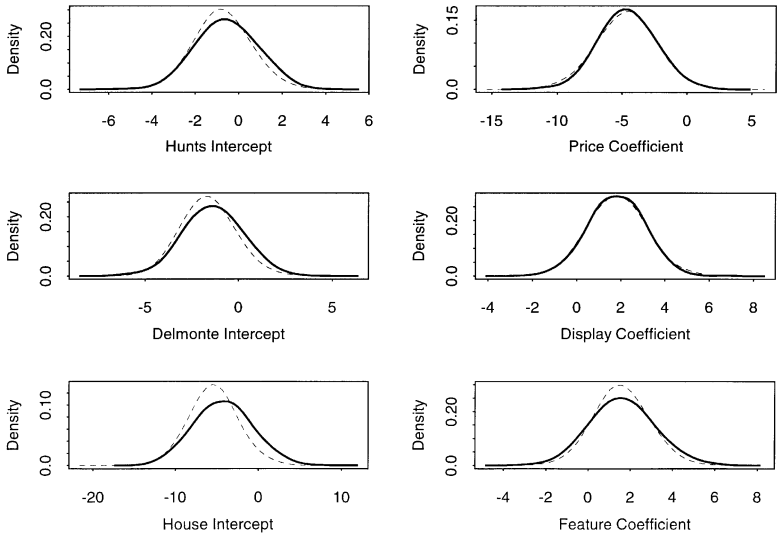


Fig. 3. Comparison of predictive and empirical marginal parameter distributions.

the normal distribution, we compute the predictive distribution of $\beta_{h'}$ for household h' selected at random from the population of households with the random effects distribution. Using our data and model, we can define the predictive distribution of $\beta_{h'}$ as follows:

$$\beta_{h'}|\text{data} = \int \phi(\beta_{h'}|\bar{\beta}, V_{\beta})p(\bar{\beta}, V_{\beta}|\text{data})d\bar{\beta}dV_{\beta}. \quad (23)$$

Here $\phi(\beta_{h'}|\bar{\beta}, V_{\beta})$ is the normal prior distribution. We can use our Gibbs draws of $\bar{\beta}$, V_{β} , coupled with draws from the normal prior to construct an estimate of this distribution which is plotted as a solid line in Fig. 3.

As an informal diagnostic tool, we can pool together all of the Gibbs draws of β_h for all H households in our dataset and compare this distribution with draws from the predictive distribution given in Eq. (23). If there is evidence of large departures from normality in this data, we should expect to see discrepancies between the predictive distribution from the model and what we call the ‘empirical’ marginal distribution of the household parameter vector. This ‘empirical’ distribution is shown in Fig. 3 as a dashed line and shows no significant departure from the predictive distribution from the model. This lends support to our assumption of normality.

Our overall conclusion from the comparison of household level inferences using the finite mixture and continuous mixture approach is that the finite

mixture approximations give a misleading impression of the level of information available to infer about household parameters. In addition, the posterior probability calculations dramatically favor the continuous mixture approach. While we will never know for sure whether actual distribution of household parameters follows a normal continuous mixture or a finite mixture, all available evidence seems to favor the normal continuous mixture approach.

4. Marketing decisions

In this section, we consider various decisions of interest to marketers and demonstrate the value of considering disaggregate estimates of model parameters. It is rarely the case that a marketing activity is uniformly directed at an entire market. Even mass media decisions, such as the design and execution of television advertising, involves the identification of a target audience and the selection of an appropriate method of message delivery (e.g. a specific network and television show). This requires knowledge of the preferences and the viewing habits of those individuals with whom the firm wants to communicate.

Product design decisions also requires detailed knowledge of the distribution of consumer preferences. In general, there are two general strategies available to firms wishing to maintain a profitable existence in competitive markets. The first is to be a low cost producer and design products to have the greatest possible demand. Because of competitive pressures this strategy typically results in lower margins, but profits can be obtained through higher volume. A second strategy of more interest to marketing is to design products for a portion of the market who place high value on particular features. For example, a manufacturer of outboard marine engines may decide to improve the durability of its product which is of particular interest to fishermen who come in frequent contact with submerged logs and weeds. Alternatively, the manufacturer might decide to improve the acceleration property of its engines which is of interest to water skiers. These product differentiation strategies are oriented toward a much smaller customer base who is willing to pay for additional performance. Classical random effects models, which only yield aggregate estimates of the distribution of demand ($\bar{\beta}$, V_{β}), do not provide sufficient information for the identification and characterization of these key customers.

At a more tactical level, there are many problems in marketing which require knowledge of the distribution of response and estimates of where particular 'units' fall within the distribution. For example, marketers are often concerned with the response of individual units (e.g. stores, sales regions, customers) to marketing expenditures. Units which display the strongest response are often used to illustrate 'best practices' while those with the lowest response are candidates for 'corrective action.' Similarly, the efficient allocation of resources

requires knowledge of the expected response of individual units of analysis and not just the distribution of responses.

An important characteristic of these marketing decisions is that actions frequently take place at a very disaggregate level, with different units of analysis often aggressively targeted with different programs. Furthermore, the identification of optimal actions often requires knowledge of nonlinear functions of model parameters. Because of the severe data limitation at the disaggregate level, substantial uncertainty exists about the true value of model parameters. This causes great difficulty in identifying optimal decision whenever nonlinear rules are encountered.

To illustrate, consider the following stylized example which involves identifying optimal temporary price reductions for households in the scanner panel dataset described above. Disaggregate price reductions are now very common in many grocery stores with the advent of point-of-purchase coupons issued at the checkout counter. The current triggering mechanism for issuing most of these coupons is related to the current purchases of the household. For example, a coupon for Hunt's ketchup is issued to buyers of Heinz. While current coupons are for a fixed amount (e.g. 50 cents), there is nothing to prevent the use of coupons of differing amount. There is also no reason that the triggering mechanism should not be based on the entire purchase history of the household rather than just the current purchase. These enhancements are currently being developed by firms such as Catalina Marketing Inc. and frequent shopper programs which issue coupons electronically (see Rossi et al., 1996 for a comprehensive treatment of target couponing strategies).

Optimal price reductions can be identified by considering the profit function:

$$\pi = \Pr(i)(p_i - c_i), \quad (24)$$

where $\Pr(i)$ is the probability of purchasing alternative i with price p_i and cost c_i . The profit function in Eq. (24) is a stylized one which is meant to reflect the profits of a manufacturer with a rather passive retailer. Retailers now frequently capture information on purchase histories. Specialized firms such as Catalina Marketing have products which issue customized incentives based on purchase history data; Catalina rents *exclusive* access to their system. We do not consider the effects of competing firms, each with its one customized price promotions problem. We leave this and other questions for further research.

An optimal price (p_i^*) to charge the household is the one for which expected profits are maximized:

$$\partial E[\pi] / \partial p_i |_{p_i^*} = 0, \quad (25)$$

where the expectation is taken with respect to the posterior distribution of model parameters.

Fig. 4 plots the optimal price reductions for Hunt's ketchup for the first 200 households in our panel. The price reductions were determined assuming a regular price of \$1.50 and a manufacturer cost of \$1.00. The plot reflects the wide dispersion in optimal price reductions, ranging from zero to \$0.34 with a mean of \$0.15. There is a large group of households for which the optimal price reduction is zero with a wide dispersion in non-zero price reductions. Profits increase from \$0.106 to \$0.117 per household when these household-specific price reductions are used. This suggests that there are substantial profit opportunities for the use of customized pricing if the costs of acquiring and using household purchase history information are sufficiently low.

As discussed in Section 2, optimal decision making requires not only individual-level parameter estimates but also must acknowledge the substantial uncertainty in those estimates. In other words, we must use the entire posterior distribution of the household parameters rather than just point estimates to solve the marketing problem. To illustrate the problems with a 'plug-in' approach in which parameter estimates are simply inserted into the profit function, we solved the optimal price reduction problem conditioning on the posterior means of each household parameter. Conditioning on the posterior means introduces over-confidence into the targeting of price reductions. We are much more certain about the benefits of large price reductions, so we tend to give

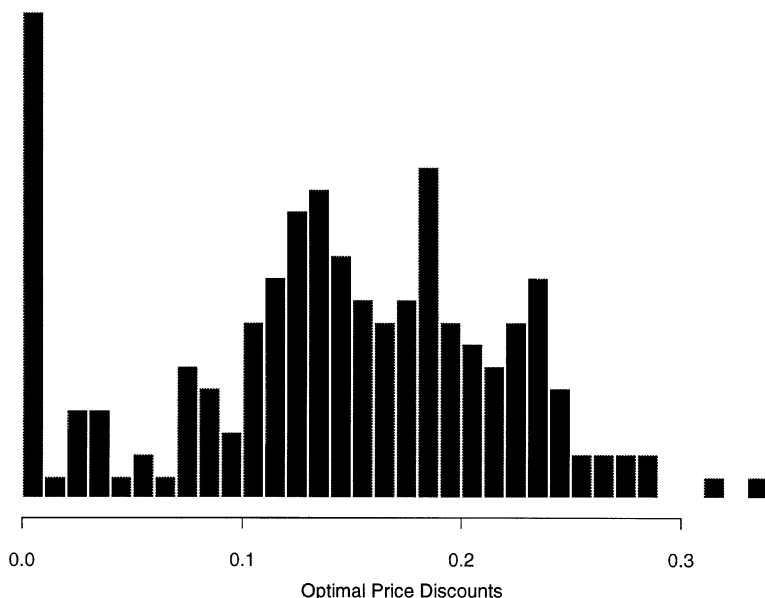


Fig. 4. Distribution of optimal price reductions.

larger price reductions to many households. The average value of the price reductions in the decisions problem which conditions on parameter estimates is 0.18 (contrast with 0.15). This results in higher profit numbers as well, \$0.129 vs. \$0.117. In our view, conditioning on the posterior means will overestimate the value of targeting households since it does not incorporate uncertainty properly.

5. Concluding remarks

A fundamental assumption of marketing is that people are different. People differ in the products they prefer, where they shop, how they communicate and in their sensitivity to variables such as price. This diversity provides firms with the economic incentive to offer different products to different groups of people, which are promoted, priced and placed in the market in a manner appropriate for the targeted groups.

The successful execution of these marketing activities requires a detailed understanding of distribution of consumer heterogeneity and identification of preferences at the customer level. For example, a firm wishing to increase sales may consider mailing coupons to price sensitive individuals who are not currently customers. Of critical interest are the preferences and sensitivities of these individuals who, by definition, are not well represented by summary statistics which describe the distribution of heterogeneity for an entire market. In the extreme, the determination of an optimal coupon value which is unique for each individual requires knowledge of model parameters for each individual.

The advantage of hierarchical Bayes models of heterogeneity is that they yield disaggregate estimates of model parameters. These estimates are of particular interest to marketers pursuing product differentiation strategies in which products are designed and offered to specific groups of individuals with specific needs. In contrast, classical approaches to modeling heterogeneity yield only aggregate summaries of heterogeneity and do not provide actionable information about specific groups. The classical approach is therefore of limited value to marketers.

The disaggregate nature of many marketing decisions, coupled with the assumption of individual differences and the limited data available about any specific individual, has created the need for models of consumer heterogeneity which successfully pool data across individuals while allowing for the (posterior) analysis of individual model parameters. The development of Markov chain methods of inference has made such an analysis possible by exploiting the hierarchical structure typically present in models which incorporate distributional assumptions of heterogeneity. Individual estimates are available as a by-product of this estimation procedure, which provide a rich source of information for a wide variety of marketing decisions.

References

- Allenby, G., Arora, N., Ginter, J., 1998. On the heterogeneity of demand. *Journal of Marketing Research*, forthcoming.
- Allenby, G., Ginter, J.L., 1995a. The effects of in-store displays and feature advertising on consideration sets. *International Journal of Research in Marketing* 12, 76–80.
- Allenby, G., Ginter, J., 1995b. Using extremes to design products and segment markets. *Journal of Marketing Research* 32, 392–403.
- Allenby, G., Lenk, P., 1994. Modeling household purchase behavior with logistic normal regression. *JASA* 89, 1218–1231.
- Chamberlain, G., 1984. Panel data. In: Griliches, Intriligator (Eds.), *Handbook of Econometrics*. North-Holland, Amsterdam.
- Chintagunta, P., Jain, D.C., Vilcassim, N.J., 1991. Investigating heterogeneity in brand preferences in logit models for panel data. *Journal of Marketing Research* 28, 417–428.
- Gelfand, A. et al., 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. *JASA* 85, 972–985.
- Gelfand, A., Smith, A., 1990. Sampling-based approaches to calculating marginal densities. *JASA* 85, 398–409.
- Gelman, A., Rubin, D., 1992. A single series from the Gibbs sampler provides a false sense of security. In: Bernardo, J.M. et al. (Eds.), *Bayesian Statistics 4*. Oxford University Press, Oxford, pp. 625–632.
- Geweke, J., Keane, M., Runkle, D., 1996. Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics*, forthcoming.
- Gonul, F., Srinivasan, K., 1993. Modeling multiple sources of heterogeneity in multinomial logits: methodological and managerial issues. *Marketing Science* 12, 213–229.
- Hajivassiliou, V., Ruud, P., 1994. Classical estimation methods for LDV models using simulation. In: Engle, McFadden (Eds.), *Handbook of Econometrics*. North-Holland, Amsterdam, pp. 2384–2438.
- Hausman, J., Ruud, P., 1987. Specifying and testing econometric models for rank-ordered data. *Journal of Econometrics* 34, 83–104.
- Hausman, J., Wise D. A., 1978. A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. *econometrica* 46, 393–408.
- Heckman, J., 1982. Statistical models for analysis of discrete panel data. In: Manski, McFadden (Eds.), *Structural Analysis of Discrete Data*. MIT Press, Cambridge, pp. 114–178.
- Heckman, J., Singer, B., 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52, 271–320.
- Kamakura, W., Russell, G., 1989. A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research* 26, 379–390.
- Lenk, P., DeSarbo, W., Green, P., Young, M., 1996. Hierarchical Bayes conjoint analysis: recovery of part worth heterogeneity from reduced experimental designs. *Marketing Science* 15, 173–191.
- Mariano, R., Weeks, M., Schuermann, T. (Eds.), 1997. *Simulation-Based Inference in Econometrics*. Cambridge University Press, Cambridge, forthcoming.
- McCulloch, R., Rossi, P., 1994. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* 64, 207–240.
- McCulloch, R., Rossi, P., 1996. Bayesian analysis of multinomial probit model. In: Mariano, Weeks, Schuermann (Eds.), *Simulation-Based Inference in Econometrics*. Cambridge University, Cambridge.
- McCulloch, R., Rossi, P., Allenby, G., 1995. Hierarchical modeling of consumer heterogeneity: an application to target marketing. In: Kass, Singpurwalla (Eds.), *Case Studies in Bayesian Statistics* (1995). Springer, New York, pp. 323–350.

- Newton, M., Raftery, A., 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. *JRSS B* 56, 3–48.
- Rossi, P., Allenby, G., 1993. A Bayesian approach to estimating household parameters. *Journal of Marketing Research* 30, 171–182.
- Rossi, P., McCulloch, R., Allenby, G., 1996. On the value of household purchase history information in target marketing. *Marketing Science* 15, 321–340.
- Schwarz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics* 6, 461–464.
- Tierney, L., 1991. Markov chains for exploring posterior distributions. Technical Report no. 560, School of Statistics, University of Minnesota.
- Titterton, D., Smith, A.F.M., Makov, U.E., 1985. *Statistical Analysis of Finite Mixture Distributions*. Wiley, New York.