Model-Based Segmentation Featuring Simultaneous Segment-Level Variable Selection

Author(s): SUNGHOON KIM, DUNCAN K.H. FONG and WAYNE S. DESARBO
Source: *Journal of Marketing Research*, Vol. 49, No. 5 (October 2012), pp. 725-736
Published by: American Marketing Association
Stable URL: http://www.jstor.org/stable/41714461
Accessed: 16-04-2018 18:05 UTC

SUNGHOON KIM, DUNCAN K.H. FONG, and WAYNE S. DESARBO*

The authors propose a new Bayesian latent structure regression model with variable selection to solve various commonly encountered marketing problems related to market segmentation and heterogeneity. The proposed procedure *simultaneously* performs segmentation and regression analysis within the derived segments, in addition to determining the optimal subset of independent variables per derived segment. The authors present comparative analyses contrasting the performance of the proposed methodology against standard latent class regression and traditional Bayesian finite mixture regression. They demonstrate that their proposed Bayesian model compares favorably with these traditional benchmark models. They then present an actual commercial customer satisfaction study performed for an electric utility company in the southeastern United States, in which they examine the heterogeneous drivers of perceived quality. Finally, they discuss limitations of the research and provide several directions for further research.

*Keywords*: Bayesian analysis, finite mixtures, perceived quality, multiple regression, customer satisfaction

# Model-Based Segmentation Featuring Simultaneous Segment-Level Variable Selection

Market segmentation involves viewing a heterogeneous aggregate market as decomposable into several smaller homogeneous markets in response to the heterogeneous preferences attributable to the desires of consumers for more precise satisfaction of their varying wants and needs (Smith 1956). Since the early pioneering work of Wendell Smith (1956), market segmentation has become one of the most pervasive activities in both the marketing academic literature and actual marketing practice. According to a recent survey reported in Roberts, Kayande, and Stremersch (2007), both marketing academicians and practitioners ranked marketing segmentation as having the *strongest* impact in the field of marketing among all marketing tools evaluated. As DeSarbo and Grisaffe (1998) state, in addition to being one of the major ways of operationalizing the marketing concept, market segmentation provides guidelines for a firm's marketing strategy and resource allocation among products and markets. Facing heterogeneous markets and customers, a firm employing a market segmentation strategy can typically increase expected profitability as suggested by the classic price discrimination model, which provides the major theoretical rationale for market segmentation (Frank, Massey, and Wind 1972).

As Wedel and Kamakura (2000) note, the major methodological methods for implementing market segmentation and representing market heterogeneity have evolved over the past several decades. Initially, researchers used classical multivariate statistical methods such as regression analysis, cluster analysis, and discriminant analysis, depending on whether they desired a priori or post hoc segmentation schemes. More recently, emphasis has shifted to model-based segmentation methodologies involving more complex optimization and numerical methods, finite mixtures, and/or Bayesian approaches given the various criteria established for effective market segmentation (see DeSarbo and DeSarbo 2003). When such segmentation applications involve an

*Sunghoon Kim is a doctoral student in Marketing (e-mail: suk220@psu.edu), Duncan K.H. Fong is Professor of Marketing and Statistics (e-mail: i2v@psu.edu), and Wayne S. DeSarbo is the Smeal Distinguished Professor of Marketing (e-mail: wsd6@psu.edu), Smeal College of Business, Pennsylvania State University. Fred Feinberg served as associate editor for this article.


ISSN: 0022-2437 (print), 1547-7193 (electronic)

obvious dependent variable and multiple independent variables, researchers have used two major classes of procedures extensively published in marketing for segmentation: clusterwise regression and latent structure regression (involving both finite mixtures and Bayesian mixtures) approaches (Wedel and Kamakura 2000). Both classes of procedures simultaneously estimate the number of market segments (K), the membership of each individual observation into the K market segments, and the parameters of the specified regression model per derived market segment. Clusterwise regression procedures (DeSarbo et al. 2008; DeSarbo, Grewal, and Scott 2008; DeSarbo and Mahajan 1984; DeSarbo, Oliver, and Rangaswamy 1989; Späth 1982; Wedel and Kistemaker 1989; Wedel and Steenkamp 1991) are deterministic least squares–based methods involving mixed discrete and continuous solutions to complex nonlinear optimization problems. As Wedel and Kamakura (2000) mention, some of the methods in this class of procedures suffer from the limitation that users must subjectively specify various tuning parameters that influence the degree of separation of the clusters. In addition, the statistical properties of the estimators are not well established.

These disadvantages are largely alleviated by more recent advances in latent structure/class regression methods based on finite mixture distribution theory in which the conditional means of the underlying support distributions of these finite mixtures are typically reparameterized by linear regression functions of prespecified independent variables. The dependent variable is typically specified to be distributed as a finite mixture of a member of the exponential family of distributions (e.g., normal, binomial, Poisson). Numerous information heuristics are typically employed for model selection (see Wedel and Kamakura 2000). DeSarbo and Cron (1988) first generalized the stochastic switching regression models considered in Quandt (1972) and Quandt and Ramsay (1978) to more than two regimes. Wedel and DeSarbo (1994), Wedel and Kamakura (2000), and DeSarbo, Kamakura, and Wedel (2006) provide a summary of the numerous marketing publications relating to the development of latent structure regression models. Wedel and DeSarbo (1995) provide a finite mixture approach for mixtures of general linear models that encompass a framework for the majority of such endeavors. A variety of software packages has become available for implementing these general linear model mixtures, including Latent Gold (Statistical Innovations Inc.), GLIMMIX, and, more recently, Flexmix (Grün and Leisch 2008).

In recent years, there has been substantial interest in the Bayesian analysis of finite mixture models, some of which have been used to address heterogeneity issues in marketing. Diebolt and Robert (1994) provide a Gibbs sampling algorithm to obtain Bayes estimators for finite mixture distributions. Allenby, Arora, and Ginter (1998) introduce a Bayesian normal component mixture model to investigate the heterogeneity of demand. Lenk and DeSarbo (2000) adopt a Bayesian approach to model parameter heterogeneity in generalized linear models with random effects. Hurn, Justel, and Robert (2003) show how Bayesian inference for finite mixture regression can be achieved by specifying loss functions and adopting the birth-and-death technique to determine the unknown number of components. Several studies (Andrews, Ansari, and Currim 2002; Otter, Tüchler, and Frühwirth-Schnatter 2004) also compare the performance of the latent class model and the (Bayesian) random coefficients model, concluding that no one method completely dominates the other. Recently, Büschken, Otter, and Allenby (2010) developed a Bayesian mixture model to analyze haloed responses in customer satisfaction data.

Akin to ordinary multiple regression analyses, traditional procedures for finite mixture regressions (i.e., latent class regression and Bayesian finite mixture regression) also have several limitations. First, there is the problem of *masking variables*, as described in Fowlkes and Mallows (1983), which, in this context, involves the problem associated with including irrelevant independent variables that can produce biased clustering/segmentation results. This potential problem becomes more prominent in situations in which strong a priori theory is lacking and/or when a plethora of plausible independent variables are available. Second, there is the issue of limited sample sizes. Recall that latent structure regression methodologies assume (like all latent class–related procedures) an underlying partitioning of the sample space into K distinct clusters or segments. As such, the use of such procedures tacitly assumes an allocation of the aggregate sample into smaller segments, thereby reducing the sample size within each derived segment. Retaining a full specification of independent variables in each derived segment thereby places heavy demands on degrees of freedom for estimation, because typically $K(P + 3) - 1$ parameters are estimated (rather than $P + 1$ in the aggregate multiple regression case, where P is the number of independent variables). While sequential selection procedures exist in the case of *aggregate* multiple regression (e.g., forward selection, backward deletion, stepwise selection) for the selection of an optimal set of independent variables, no such procedures exist for the latent structure regression approach. We are also not aware of any such work on the Bayesian finite mixture regression model with variable selection.

Previous research has thoroughly examined the issue of masking effects and variable selection and weighting in the context of ordinary cluster analysis (see Brusco and Cradit 2001; DeSarbo et al. 1984; De Soete, DeSarbo, and Carroll 1985; Fowlkes, Gnanadesikan, and Kettenring 1988; Gnanadesikan, Kettenring, and Tsao 1995). Recent clustering models have begun to use stochastic variable selection in conducting both clustering and variable selection simultaneously. In the stochastic variable selection methods, researchers have posited two approaches: (1) the Bayesian variable selection approach (George and McCulloch 1993, 1997; Gilbride, Allenby, and Brazell 2006) and (2) the frequentists' penalized likelihood approach, called the Lasso (least absolute shrinkage and selection operator) or SCAD (smoothly clipped absolute deviation) (see Fan and Li 2001; Tibshirani 1996). Tadesse, Sha, and Vannucci (2005) propose a Bayesian variable selection model in clustering high-dimensional data focusing on data that have smaller sample sizes than the number of covariates. Their model handles the problem of selecting a few variables among a large number in the multivariate normal mixture model by adding a binary vector updated by a stochastic technique. Kim, Tadesse, and Vannucci (2006) extend the Tadesse, Sha, and Vannucci model by formulating the clustering with Dirichlet process mixture models. In addition, Raftery and Dean (2006) propose an alternative variable selection approach in clustering that provides a search algorithm and adopt an approximate Bayes factor for finding a local optimum (see also Dean and Raftery 2010). For sequential or longitudinal observed data, Fong and DeSarbo (2007) present a Bayesian model to simultaneously estimate the location of change points and regimes, the corresponding subset of significant independent variables per regime, and the associated regimes' regression parameters. To analyze genomic data, Gupta and Ibrahim (2007) assume a multivariate regression setup and demonstrate a Monte Carlo method for variable selection (motifs) and genes clustering by using ridge-regression-type priors but find that the selection of tuning parameter in ridge regression can severely affect the analysis results.

Several somewhat related penalized likelihood approaches have been recently developed explicitly for determining the number of components and variables in latent structure regression models. Naik, Shi, and Tsai (2007) propose a mixture regression criterion for the simultaneous determination of the number of market segments and the number of independent variables. This criterion is composed of three terms: the first measures a lack of fit, the second imposes a penalty for regression parameters, and the third term penalizes the number of components to be retained. Khalili and Chen (2007) propose another such penalized likelihood approach developed explicitly for latent structure regression by employing the Lasso or SCAD technique for variable selection. More recently, Städler, Bühlmann, and Van de Geer (2010) have proposed alternative penalty functions in this likelihood framework. One of the main problems associated with such penalty function approaches involves the stipulation that the user must set tuning parameter values somewhat arbitrarily, and those values can dramatically affect the cluster/segmentation and variable selection results. In addition, the Kahlili and Chen software is limited to only two market segments.

The current research contributes to the area of model-(regression-) based segmentation, in which we devise a new Bayesian methodology with variable selection to simultaneously derive unknown market segments (e.g., the number of market segments), the memberships of the derived segments (e.g., their composition and size), and each specific segment-level regression model with variable selection. We next present a small illustration that highlights the need for such an approach.

## *AN ILLUSTRATION*

As a continuing illustration of these issues in model-based segmentation, we assume a hypothetical customer satisfaction (CSM) scenario in which the dependent variable is an overall satisfaction, value, or perceived quality measure and the independent variables are some 19 performance attribute ratings (plus an intercept) that are all positively correlated with the dependent variable. We created such hypothetical data for 200 consumers with $K = 3$ market segments.

Table 1 presents the true parameter values used to create this synthetic data (we also added error to the data) and the analysis results using two existent methods: (1) traditional latent class regression (Flexmix in R package) and (2) Bayesian finite mixture regression (RegmixMH in the Mixtools R package; see Benaglia et al. 2009).[1] Note that the true parameter values are all positive, as would be expected in most CSM applications in marketing.

First, if we take the raw data and submit them to a traditional latent class regression analysis (Flexmix), Table 2 shows the summary results for model selection. Here, Akaike information criterion (AIC) and the modified AIC measure (AIC3) heuristics suggest continuing the analysis beyond five market segments. The more conservative Bayesian information criterion (BIC) and consistent AIC (CAIC) measures indicate no significant heterogeneity and point to the aggregate $K = 1$ market segment solution. No model selection heuristic points to $K = 3$ (the true number of market segments) as the most parsimonious solution.

Second, if we force the $K = 3$ segment solution for Flexmix, we obtain the Flexmix results in Table 1 and witness some of the problems that can arise. First, nearly all the coefficients are significant at .05 in all three derived segments. The procedure thus does a poor job of discriminating the true drivers of customer satisfaction by market segment. Second, some of the estimated and significant coefficients are negative, which lacks face validity in such CSM marketing studies. Finally, the membership hit rate is poor, at 51.5%.

Next, we consider Bayesian finite mixture regression and examine the RegimixMH results in Table 1. It is clear that the Bayesian finite mixture regression analysis also fails to recover the true coefficients, and this can lead to distortions involving the true underlying drivers (e.g., see derived Segment 2). In addition, the procedure shows problematic results in estimating the mixture proportions when the proportion of membership for the second segment ($k = 2$) is much smaller (15%) than the true proportion (33%), while the first ($k = 1$) and third segments ($k = 3$) are oversized (i.e., 42% vs. true 32% and 43% vs. true 35%). Thus, the results that RegmixMH yields reveal poor performance in both true parameter recovery and true mixture proportion recovery.

In summary, by using these traditional methods, the marketing managerial implications and resulting resource allocations drawn from the analyses are misleading. It appears that problems can arise in such classic model-based segmentation studies in which there are a large number of independent variables with small to moderately sized samples and many of the independent variables have selected impact across the derived market segments. As has been shown extensively in the classification literature, excess variables can mask the true underlying cluster structure. We return to this same illustration subsequently.

We organize the rest of the article as follows: The next section presents the technical description of the proposed Bayesian finite mixture model with variable selection (for more details, see Web Appendixes A–D at www.marketingpower.com/jmr_webappendix). After providing a description of

[1]We ran Flexmix multiple times for each value of K and selected the best-fitting solution given its propensity for local optimum solutions in all analyses. For RegmixMH, we performed multiple runs, with 10,000 burn-in iterations and an additional 10,000 iterations for estimation, and picked the best solution closest to the true parameter values.

## Table 1
COMPARISON OF RESULTS FROM VARIOUS MODELS FOR SYNTHETIC DATA

| | True K = 3 Results | | | Flexmix Results[a] | | | RegmixMH Results[b] | | | Bayes Variable Selection Results[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | k = 1 | k = 2 | k = 3 | k = 1 | k = 2 | k = 3 | k = 1 | k = 2 | k = 3 | k = 1 | k = 2 | k = 3 |
| Intercept | −.01 | 0 | 1.22 | .55 | 1.05 | .55 | .09 | .17 | .74 | 0 | 0 | 1.18 |
| Attribute1 | .01 | 0 | 1.2 | −1.39 | .62 | −1.39 | .06 | .2 | 1.48 | 0 | 0 | 1.19 |
| Attribute2 | 0 | .02 | 1.2 | .7 | 1.21 | .7 | −.02 | .75 | .94 | 0 | 0 | 1.23 |
| Attribute3 | −.01 | .01 | 1.17 | −.91 | .89 | −.91 | .11 | 1.92 | 1.25 | 0 | 0 | 1.22 |
| Attribute4 | 0 | .01 | .79 | .73 | .15 | .73 | −.32 | .19 | .4 | 0 | 0 | .74 |
| Attribute5 | .01 | −.01 | .8 | .3 | .97 | .3 | .06 | 1.17 | .81 | 0 | 0 | .79 |
| Attribute6 | −.02 | .98 | .81 | 1.08 | .23 | 1.08 | −.05 | .75 | 1.15 | .06 | .93 | .83 |
| Attribute7 | 0 | 1 | −.01 | .82 | .39 | .82 | .08 | −.27 | .5 | .06 | .96 | 0 |
| Attribute8 | 0 | .99 | −.01 | −.09 | −.47 | −.09 | .00 | 1.49 | .02 | .06 | .97 | 0 |
| Attribute9 | .03 | .98 | −.02 | −.5 | .69 | −.5 | −.06 | .18 | .01 | .06 | .98 | 0 |
| Attribute10 | −.01 | 1 | 0 | .6 | .44 | .6 | .20 | .67 | .2 | .06 | .91 | 0 |
| Attribute11 | −.02 | 1 | .02 | 1.21 | .18 | 1.21 | −.07 | 1.85 | .2 | .06 | .9 | 0 |
| Attribute12 | .68 | 0 | −.01 | 1.09 | .09 | 1.09 | .74 | .14 | −.17 | .68 | .04 | 0 |
| Attribute13 | .71 | .02 | 0 | 1.4 | .32 | 1.4 | .47 | .28 | −.01 | .66 | .04 | 0 |
| Attribute14 | .7 | −.02 | .01 | .32 | 1.6 | .32 | .9 | −.10 | −.05 | .7 | .05 | 0 |
| Attribute15 | 1.51 | .02 | 0 | 2.08 | .09 | 2.08 | 1.39 | .62 | −.01 | 1.39 | .09 | 0 |
| Attribute16 | 1.49 | .02 | .02 | .88 | −.25 | .88 | 1.69 | .43 | −.21 | 1.4 | .08 | 0 |
| Attribute17 | 1.51 | 0 | −.01 | .92 | −.01 | .92 | .91 | .26 | .10 | 1.37 | .08 | 0 |
| Attribute18 | 1.52 | −.02 | −.01 | 1.17 | −.86 | 1.17 | 1.14 | .25 | .21 | 1.4 | .09 | 0 |
| Attribute19 | 1.49 | 0 | .01 | .36 | .81 | .36 | 1.39 | 1.43 | −.05 | 1.43 | .09 | 0 |
| Mixing Prop | .32 | .33 | .35 | .2 | .38 | .42 | .42 | .15 | .43 | .33 | .335 | .335 |
| Memb HIT | | 100% | | | 51.50% | | | Not available | | | 93% | |

Variable selection criteria:

[a] $p < .05$.

[b] No zero inclusion in 95% credible intervals.

[c] Odds ratio > 20.

Notes: Shaded boxes denote selected coefficients.

Table 2
MODEL SELECTION CRITERIA FOR THE SYNTHETIC DATA

| | *K = 1* | *K = 2* | *K = 3 (True K)* | *K = 4* | *K = 5* |
|---|---|---|---|---|---|
| AIC | 933.44 | 883.75 | 814.96 | 734.01 | 644.93[a] |
| AIC3 | 954.44 | 926.75 | 879.96 | 821.01 | 753.93[a] |
| BIC | 1002.71[a] | 1025.57 | 1029.35 | 1020.96 | 1004.45 |
| CAIC | 1023.70[a] | 1068.57 | 1094.35 | 1107.96 | 1113.45 |
| Log marginal likelihood | –471.32 | –429.63 | –331.8[a] | –1063.82 | –1298.01 |

[a]Indicates the most parsimonious solution selected by each criterion.

the proposed model, we return to the preceding illustration to demonstrate the comparative performance of the proposed methodology with the traditional procedures discussed previously (latent class regression and Bayesian finite mixture regression analysis). Following this, we introduce an actual commercial application involving a customer satisfaction and perceived quality study performed for a major electric utility company in the southeastern United States. We conclude with a discussion of the results and list some limitations of the research as well as several potential future research directions in this area.

## *THE PROPOSED BAYESIAN FINITE MIXTURE REGRESSION MODEL WITH VARIABLE SELECTION*

### *The Proposed Model*

Let $Y_i$ be the observation on the dependent variable for respondent i. We assume a multiple regression setting:

$$(1)\qquad Y_i = \underline{X}_i^T \underline{\beta}_{H_i} + \varepsilon_i, i = 1, ..., n,$$

where $\underline{X}_i^T$ is a row vector of dimension P containing values of the independent variables for respondent i, $\underline{\beta}_{H_i}$ is a (column) vector of segment level regression coefficients, respondent i is in market segment k if $H_i = k$, and the error terms $\varepsilon_i$ are independently and normally distributed as N(0, $\sigma^2$). We obtain a finite mixture regression model when we assume the segment indicator variable $H_i \in \{1, \ldots, K\}$ to follow a discrete distribution with positive parameters $d_1, \ldots, d_K$:

$$(2)\qquad \pi\left(H_i \middle| \underline{d}\right) = \text{discrete}\left(d_1, ..., d_k\right),$$

where discrete(*) denotes a specific discrete distribution—that is, $\pi(H_i = k|\underline{d}) = d_k$ and $\Sigma_{k=1}^K d_k = 1$. In a Bayesian approach, these parameters are usually assumed to follow a Dirichlet distribution, and we adopt this assumption here:

$$(3)\qquad \pi(d_1, \ldots, d_K) = \text{Dirichlet}(\alpha_1, \ldots, \alpha_K).$$

To allow for variable selection, we employ a spike and slab–type prior on $\underline{\beta}_k$, k= 1, ..., K, (see Ishwaran and Rao 2005; Mitchell and Beauchamp 1988). For each component $\beta_{kp}$, p = 1, ..., P, we assume the following:

$$(4)\qquad \begin{cases} \pi\left(\beta_{kp} \middle| Z_{kp} = 1\right) = N\left(0, \tau_p^2\right) \\ \pi\left(\beta_{kp} = 0 \middle| Z_{kp} = 0\right) \neq 1, \end{cases}$$

where $Z_{kp}$ is a Bernoulli random variable with a parameter w. In other words, when $Z_{kp} = 1$ (with a probability of w), we select the pth variable in the kth segment and assume that its coefficient $\beta_{kp}$ follows a normal distribution with zero mean, which is a common assumption in the variable selection literature. In the other case in which $Z_{kp} = 0$ (with a probability of 1 – w), we do *not* select the pth variable in the kth segment, and we set its coefficient ($\beta_{kp}$) as zero. Thus, we introduce the binary latent variables $Z_{kp}$ to indicate whether the variables have impact in a segment. In addition, we assume an exchangeable prior for $Z_{kp}$ with parameter w following a beta distribution:

$$(5)\qquad \pi(Z_{kp}|w) = \text{Bernoulli}(w), \text{ and}$$

$$(6)\qquad \pi(w) = \text{Beta}(a, b).$$

Finally, for the variances, we assume the commonly used inverse gamma distributions:

$$(7)\qquad \pi(\sigma^2) = \text{InvGamma}(r_1, r_2), \text{ and}$$

$$(8)\qquad \pi(\tau_p^2) = \text{InvGamma}(s_{p1}, s_{p2}).$$

### *The Full Conditional Distributions*

In our Bayesian computation, we generate an approximate sample from the joint posterior distribution by drawing random deviates iteratively and recursively from the following full conditional distributions (for their formal derivations, see Web Appendix A at www.marketingpower.com/jmr_webappendix), which we summarize subsequently:

(9) $\pi(H_i|\text{all others})$ is a discrete distribution,

(10) $\pi(\underline{d}|\text{all others})$ is a Dirichlet distribution,

(11) $\pi(\beta_{kp} = 0|Z_{kp} = 0 \text{ and others}) = 1$,

and $\pi(\beta_{kp}|Z_{kp} = 1 \text{ and all others})$ is a normal distribution,

(12) $\pi(Z_{kp}|\text{all others except } \beta_{kp})$ is a Bernoulli distribution,

(13) $\pi(w|\text{all others})$ is a beta distribution,

(14) $\pi(\sigma^2|\text{all others})$ is an inverse gamma distribution, and

(15) $\pi(\tau_p^2|\text{all others})$ is an inverse gamma distribution.

Note that all these full conditional distributions are standard probability distributions, and therefore we can use a modified Gibbs sampling algorithm here. Web Appendix B (www.marketingpower.com/jmr_webappendix) delineates the specific steps of our proposed Markov chain Monte Carlo (MCMC) estimation procedure. To use this model, it is necessary to input the data and specify the number of segments (K), the number of sampling chains, and the number of burn-ins. Then, this MCMC procedure will compute and save all the posterior results throughout all the iterations. In particular, researchers often focus their interest on the segment coefficients ($\underline{\beta}$), variable selection information ($\underline{Z}$), and segment membership ($\underline{H}$); this procedure will report posterior mean values of $\underline{\beta}$ and $\underline{Z}$, as well as the posterior membership ($\underline{H}$) sampled from the membership allocation probabilities. It is also easy to construct various plots from the saved posterior outputs (e.g., trace plot, density plot, cumulative density plot).

*A Relabeling Algorithm*

An important issue in any Bayesian mixture model using MCMC is the label-switching identification problem, in which the order of components can be arbitrarily changed multiple times between iterations (see Jasra, Holmes, and Stephens 2005). The label-switching problem produces an identical likelihood for all permutations of the indexes of the components. In other words, when exchangeable priors with no segment-specific information are employed, the sampler cannot distinguish the parameters of one segment from those of another segment according to the likelihood. This is problematic for such mixture models, because an MCMC sampler with such a label-switching problem can jump around across segments, producing an inconsistent and meaningless ergodic average estimate. If we do not solve this problem, it will be difficult to infer parameters specific to each derived market segment.

To address the label-switching problem, we follow Marin and Robert (2007) to first simulate from the unconstrained posterior distribution and then impose identifiability constraints on the generated MCMC sample. Note that when we have simulated an MCMC sample from an unconstrained posterior distribution, any ordering constraint can be imposed on this sample *after* the simulations have been completed for estimation purposes, as Stephens (1997) stresses (see also Lee et al. 2009). This postprocessing approach alleviates any concern of an adverse effect on simulation by imposing a constraint on the support of the posterior as simulations are made from the unconstrained posterior distribution. After the simulation is completed, we relabel the $\underline{\beta}_k$ for each MCMC iteration according to a constraint—e.g., $\beta_{1p} < \beta_{2p} < \cdots < \beta_{Kp}$—for a given component p. Then, we reorder the other associated parameters $\underline{Z}$, $\underline{H}$, and $\underline{d}$ accordingly to match that of $\underline{\beta}_k$. This method is easy to implement, and it works well for all simulation data sets we have analyzed. Indeed, it compares favorably with more complicated techniques in terms of programming effort and computational cost (for a brief review of existing relabeling algorithms, see Sperrin, Jaki, and Wit 2010).

*Model Selection*

To compare models $M_1$ and $M_{1*}$, we can use the Bayes factor, which is given by the ratio of the two corresponding marginal likelihoods. To find the marginal likelihoods $\pi(\underline{Y}|M_k)$, k = 1, ..., K, we use the basic marginal likelihood identity Chib (1995) suggests:

$$\pi(\underline{Y}|M_k) = \frac{f(\underline{Y}|M_k, \theta^*)\pi(\theta^*|M_k)}{\pi(\theta^*|M_k, \underline{Y})}, \tag{16}$$

where we set $\theta^* = \{\underline{H}^*, \underline{\beta}^*, \sigma^{2*}\}$ equal to the posterior modes here (for the technical details involved in the computation of this marginal likelihood, see Web Appendix C at www.marketingpower.com/jmr_webappendix). Note that Berkhof, Van Mechelen, and Gelman (2003) propose a variant of Chib's estimator to account for the nonidentifiability of the mixture components (see also Lee et al. 2009). From a computational viewpoint, in general, it is more efficient to compute Equation 16 in logarithmic form:

$$\ln\left[\pi(\underline{Y}|M_k)\right] = \ln\left[f(\underline{Y}|M_k, \theta^*)\right] + \ln\left[\pi(\theta^*|M_k)\right] - \ln\left[\pi(\theta^*|M_k, \underline{Y})\right], \tag{17}$$

Here, we can compute $\ln[f(\underline{Y}|M_k, \theta^*)]$ and $\ln[\pi(\theta^*|M_k)]$ analytically. For $\ln[\pi(\theta^*|M_k, \underline{Y})]$, we employ a data augmentation scheme to estimate the quantity (Chen, Shao, and Ibrahim 2000, p. 239). Because the logarithm of the Bayes factor is equal to the difference of the log marginal likelihoods, we can similarly employ the log marginal likelihood as a model selection heuristic and select the model with the largest value.

*The Illustrative Example*

We return now to the small synthetic/hypothetical customer satisfaction segmentation example presented previously and consider the results of our proposed model incorporating the variable selection contained in Table 1. For the proposed Bayes model, we ran the model with 10,000 burn-in and then an additional 10,000 iterations and inspected the corresponding trace plots for convergence, which showed steady results after the burn-in period. Because we assume Bernoulli distributions for variable selection, we adopted the ratio of the posterior odds and prior odds to decide on variable selection. Web Appendix D (www.marketingpower.com/jmr_webappendix) describes the calculation of odds ratio for our Bayesian variable selection procedure using a cutoff value of 20 (see Jeffreys 1961). As we show, our proposed methodology performs much better in recovering the significant drivers of customer satisfaction by derived segment. In addition, our model selection criterion (the log-marginal likelihood at the bottom of Table 2) correctly points to K = 3 market segments. The correct membership recovery (membership hit rate) is also substantially higher in using our proposed model. Thus, for this hypothetical model-based segmentation illustration, the use of the traditional methods (Flexmix and RegmixMH) leads to inaccurate managerial implications, as well as to inaccurate market segment membership predictions. Web Appendix E presents a formal comparison of these three procedures using a more complete Monte Carlo analysis.

## *A COMMERCIAL APPLICATION: CUSTOMER SATISFACTION AND PERCEIVED QUALITY MEASUREMENT*

Note that CSM studies often involve multiple research objectives. Managers of a company often want to measure and track performance on key business dimensions in an attempt to integrate customer requirements throughout their organization. In addition, there is a desire to identify aspects of the business that have the most impact on overall measures or indexes of customer opinion or perceptions on such constructs as customer satisfaction, value, perceived quality, recommendations/positive word of mouth, repurchase intentions, loyalty, and so on. The impact of these elements often are identified through the application of regression-based procedures that model some overall performance (dependent) measure as a function of more specific attribute evaluation measurements. For example, in many CSM studies, overall satisfaction or quality perceptions are typically modeled as a function of more specific product and service attribute ratings.

*The DeSarbo and Grisaffe (1998) CSM Study*

To explore such analysis issues, we analyzed CSM data from commercial and industrial customers of a large electric utility company as presented in DeSarbo and Grisaffe (1998). Although some of the names of measures are disguised in the 1998 study, it uses the actual numerical data, and we analyzed them as originally collected. Customers of the client received a notification letter one week before the interview to alert them that they would receive a telephone phone shortly regarding their customer satisfaction opinions. Following this mailing, the associated fieldwork was implemented. The survey took approximately 20 minutes per respondent and was conducted by telephone interview. The survey and data collection occurred for 15 months. We include 1509 cases in our analysis. All cases all had valid and complete values on the dependent variable of interest: a rating of overall perceived quality. For reference, we denote overall quality with the traditional dependent variable symbol Y. The independent variables used to predict overall quality included ratings of power reliability, preventative maintenance, repair services, account representative, technical support, customer service, record keeping, billing, and a measure of price perceptions. We symbolize these independent variables with traditional labeling as X variables. A variety of firm-specific and demographic variables were also collected (referred to as Z variables). They included measures of region, account type, business type, respondent job type, presence of relationships with other suppliers, number of employees, number of years as a customer, and a standardized measure of revenue coming from the particular account.

*Traditional Analyses (Aggregate Multiple Regression, Flexmix, and RegmixMH)*

Table 3 presents the unconstrained aggregate sample multiple regression results performed across all 1509 sample respondents. Note that all coefficients (except the intercept term) are positive, as expected, indicating that increased performance on each attribute will raise perceptions of overall quality. Although all independent variables are significant at $p < .05$ (except the intercept and repair), reliability of the product, technical support, and the account representative appear to be the most important determinants or drivers of perceived quality for the N = 1509 industrial customers of this electric utility company. The current issue is whether this aggregate analysis masks potential segment-level differences (e.g., heterogeneity) in such coefficients.

Table 4 presents the various information model selection heuristics for the Flexmix latent structure regression analysis for K = 1–6 market segments. We ran Flexmix multiple times with different random starts given the preponderance of local optimum solutions and report the best solution for each value of K. As Table 4 indicates, depending on which information criteria heuristic we focus on, different model solutions are suggested, which is a potential problem associated with such traditional finite mixture–based methods. For the more conservative BIC and CAIC measures, the K = 3 solution is to be selected; the AIC3 points to the K = 4 solution. In addition, the AIC measure keeps decreasing as K increases and tends to prefer larger K ≥ 6 solutions. Tables 5 and 6 present both the K = 3 and K = 4 summary results. (The K = 1 solution is identical to the aggregate multiple regression results presented in Table 3.) For the K = 3 solution (Table 5), Segment 1 consists of 31.7% of membership proportion, and this segment encompasses technical support, account representative, maintenance, product reliability, repair, and customer service as important drivers for perceived quality. For Segment 2 (19.5%), records keeping, price, billing accuracy, customer service, and product reliability are significant drivers of quality perceptions. However, it is strange to observe that repair, maintenance, and account representative have negative relationships with perceived quality in this Segment 2. For Segment 3 (the largest segment, encompassing 48.8% of the sample), prod-

Table 3

AGGREGATE MULTIPLE REGRESSION RESULTS ACROSS THE ENTIRE CUSTOMER SATISFACTION SAMPLE

| | *Estimate* | *SE* | *t-Value* | *Pr(>\|t\|)* | *Significance* |
|---|---|---|---|---|---|
| Intercept | .012 | .084 | .147 | .884 | |
| Price | .055 | .021 | 2.563 | .010 | * |
| Product | .319 | .018 | 17.594 | < 2e – 16 | *** |
| Maint | .086 | .019 | 4.524 | .000 | *** |
| Repair | .040 | .023 | 1.784 | .075 | † |
| Acctrep | .135 | .022 | 6.017 | .000 | *** |
| Techsup | .140 | .022 | 6.420 | .000 | *** |
| Custsvc | .081 | .021 | 3.939 | .000 | *** |
| Records | .055 | .019 | 2.876 | .004 | ** |
| Billing | .048 | .021 | 2.298 | .022 | * |

†Significant at .05.
*Significant at .01.
**Significant at .001.
***Significant at .000.

Table 4

INFORMATION CRITERIA FOR MODEL SELECTION FOR THE FLEXMIX LATENT STRUCTURE REGRESSION ANALYSIS OF THE CSM DATA

| | *K = 1* | *K = 2* | *K = 3* | *K = 4* | *K = 5* | *K = 6* |
|---|---|---|---|---|---|---|
| AIC | 2683.28 | 2660.08 | 1696.68 | 1678.85 | 1677.42 | 1675.66[a] |
| AIC3 | 2694.28 | 2683.08 | 1731.68 | 1725.85[a] | 1736.42 | 1746.66 |
| BIC | 2741.79 | 2782.42 | 1882.85[a] | 1928.85 | 1991.26 | 2053.32 |
| CAIC | 2752.793 | 2805.42 | 1917.85[a] | 1975.85 | 2050.26 | 2124.32 |

[a]Denotes minimum information heuristic value.

Table 5

LATENT STRUCTURE REGRESSION RESULTS FOR THE CSM EXAMPLE USING *FLEXMIX* (K = 3)

| | *k = 1* | *k = 2* | *k = 3* |
|---|---|---|---|
| Intercept | .174 | .277 | .010 |
| Price | .045 | .466* | –.003 |
| Product | .109* | .192* | 1.010* |
| Maint | .160* | –.272* | –.010* |
| Repair | .101* | –.483* | .006 |
| Acctrep | .168* | –.217* | .005 |
| Techsup | .236* | –.021 | .001 |
| Custsvc | .072* | .211* | –.007 |
| Records | .019 | .453* | .004 |
| Billing | .007 | .221* | –.004 |
| Membership | 478 | 294 | 737 |
| Memb Prop | 31.7% | 19.5% | 48.8% |

*$p \leq .05$.

Table 6
LATENT STRUCTURE REGRESSION RESULTS FOR THE CSM EXAMPLE USING *FLEXMIX* (K = 4)

| | k = 1 | k = 2 | k = 3 | k = 4 |
|---|---|---|---|---|
| Intercept | .083 | .125 | .007 | .965* |
| Price | .003 | .464* | –.003 | .081 |
| Product | .153* | .195 | 1.010* | –.048 |
| Maint | .078 | –.249* | –.011* | .579* |
| Repair | .053 | –.498* | .007 | .312* |
| Acctrep | .212* | –.156 | .006 | –.139 |
| Techsup | .290* | .009 | .001 | –.053 |
| Custsvc | .153* | .218* | –.006 | –.293* |
| Records | .038 | .445* | .004 | –.064 |
| Billing | .031 | .196 | –.004 | –.041 |
| Membership | 646 | 49 | 729 | 85 |
| Memb Prop | 42.8% | 3.2% | 48.3% | 5.6% |

*$p \leq .05$.

uct reliability is the most important driver of perceived quality, followed by maintenance (which, surprisingly, also has a negative regression coefficient in this segment).

For the K = 4 solution (Table 6), the largest segment is Segment 3 (48.3% of the sample), in which quality perceptions appear to be primarily driven by ratings of product reliability and there are surprisingly negative significant effects due to maintenance. In Segment 1 (the second-largest segment, with 42.8% of the sample), quality perceptions relate significantly to experiences with technical support, account representative, product reliability, and customer services. Quality perceptions in Segment 2 (3.2% of the sample) are driven most strongly by price, records keeping, and the customer service experience. However, in the case of Segment 2, we also found that perceived quality has two problematic negative coefficients: for maintenance and repair. Finally, for Segment 4 (5.6% of the sample), quality perceptions are driven mostly by maintenance, repair, and customer service, and here, customer service has a negative relationship with perceived quality as well. Thus, Flexmix (for both K = 3 and K = 4 solutions) derives negative coefficients in each solution, which is most difficult to justify in this particular CSM commercial setting. It is also notable that for the K = 4 solution, the two smallest derived segments (i.e., Segments 2 and 4) also possess the two largest mean quality perception rating of the four market segments.[2] There is an issue of how real such small segments are given their tiny mixing proportions.

Next, we submitted the CSM data to RegmixMH (Bayesian finite mixture regression; Hurn, Justel, and Robert 2003). Table 7 presents the various model selection heuristics (differences in various information criteria) computed in the software for comparing K > 1 components against the aggregate K = 1 solution. As shown, most of the model selection heuristics point to the K = 4 solution. For RegmixMH, we performed multiple runs with 10,000 burn-in and then an additional 10,000 iterations for estimation.

[2]We also ran Latent Gold as a comparison for the application, assuming that the dependent variable is ordinal. Whereas BIC and CAIC point to the aggregate model (K = 1), AIC and MAIC point to K = 4. In the K = 4 result, there are also several negative coefficients. Moreover, we ran the Khalili and Chen (2007) method for K = 2 market segments. Here, we obtained one large segment with 99.5% of the sample and the other with only .5%.

Table 7
INFORMATION CRITERIA FOR MODEL SELECTION FROM MIXTOOLS FOR BAYESIAN FINITE MIXTURE REGRESSION ANALYSIS OF THE CSM DATA

| | K = 2 | K = 3 | K = 4 | K = 5 | K = 6 |
|---|---|---|---|---|---|
| ΔAIC | –845.1 | –1196 | –445.7 | –678 | –403.8[a] |
| ΔBIC | –903.6 | –1286.4 | –568.2[a] | –832.3 | –590 |
| ΔCAIC | –914.6 | –1303.4 | –591.1[a] | –861.3 | –624 |
| ΔICL | –902.9 | –1285.5 | –566.7[a] | –831.1 | –588.5 |

[a]Denotes selected maximum information heuristic value.
Notes: We used reported values to compare models with K components versus the base model with 1 component and the highest value indicates the optimal model. (All K models are better than the base model here.)

We inspected the corresponding trace plots for convergence, which showed steady results after the burn-in period. Table 8 presents the estimated coefficients and mixture proportions by using RegmixMH for four segments. The largest segment is Segment 3 (55.5% of the sample), and all factors except price, repair, and technical support are significantly related to perceived quality. Segment 2 is the second-largest segment, with 34.6% of sample, and we find that all the independent variables are significant drivers of perceived quality. Segment 1 is a small segment, with 5.2% of sample, and all factors in Segment 1 are tagged as significant drivers to perceived quality. Last, Segment 4, the smallest segment (4.8% of the sample), is plagued by many significant negative coefficients that lack face validity for this particular application. Thus, this solution is also of questionable value for actual managerial practice given the extreme dispersion in the relative sizes of the four derived market segments and the lack of face validity.

### *The Proposed Bayesian Finite Mixture Regression Model with Variable Selection*

We estimated our newly proposed Bayesian model using the MCMC algorithm (described fully in Web Appendix B [www.marketingpower.com/jmr_webappendix]), in which we discarded the first 10,000 iterations as a burn-in period and saved the following 10,000 iterations to estimate the posterior distributions for our parameters of interest. To check convergence of the MCMC chains, we inspected the

Table 8
RESULTS FROM A BAYESIAN FINITE MIXTURE REGRESSION MODEL (K = 4) FOR THE CSM EXAMPLE

| | *Coefficients Estimates* | | | |
|---|---|---|---|---|
| *Variables* | k = 1 | k = 2 | k = 3 | k = 4 |
| Intercept | 1.61[a] | .98[a] | .07 | –.80[a] |
| Price | 1.02[a] | 1.06[a] | .01 | –.97[a] |
| Product | .99[a] | 1.04[a] | .25[a] | –.91[a] |
| Maint | 1.01[a] | 1.05[a] | .10[a] | –1.09[a] |
| Repair | 1.05[a] | 1.08[a] | .04 | –.86[a] |
| Acctrep | .85[a] | 1.05[a] | .14[a] | –.77[a] |
| Techsup | 1.28[a] | 1.02[a] | .09 | –.90[a] |
| Custsvc | 1.04[a] | .99[a] | .08[a] | –1.27[a] |
| Records | .87[a] | 1.05[a] | .06[a] | –.03 |
| Billing | 1.04[a] | 1.00[a] | .17[a] | –.96[a] |
| Memb Prop | 5.2% | 34.6% | 55.5% | 4.8% |

[a]Zero is not included in a 95% HPD credible set.

corresponding trace plots, which showed steady results after the burn-in period. Table 9 presents the log marginal likelihood values for K = 1–5 for the proposed Bayesian mixture with variable selection approach introduced in the article. As shown, the criteria clearly points to the K = 4 segment solution. Table 10 presents the summary results for this K = 4 solution. Here, we utilize the computed odds ratio to determine the dominant drivers of perceived quality per derived market segment. In Segment 1 (19% of the sample), maintenance and technical support are the dominant drivers of perceived quality; product reliability is barely not significant (odds ratio 15.87). Product reliability is a dominant factor for Segment 2 (28.7% of the sample), followed in importance by technical support and the account representative. Product is a dominant factor for Segment 3 as well, and customer service and records keeping also drive perceived quality. Price is the most important driver for Segment 4 (21.4% of the sample), followed by billing. Outside the intercept, virtually all the estimated coefficients across the four derived segments are positive. In addition, the sizes of all four derived market segments are better balanced than those obtained in both Flexmix and RegmixMH.

Thus, we provide an explicit comparison of the three methods in this CSM study. The Flexmix latent structure regression results suggest that there is uncertainty as to which solution (i.e., value of K) is appropriate—K = 3, 4, or 6—given the different recommendations provided by the many information heuristics used. The associated K = 4 solution contains a market segment that has little face validity in terms of the many significant, negative coefficients estimated with respect to the various service attributes. In addition, two of the four derived market segments are exceedingly small. Another method, Bayesian finite mixture regression (RegmixMH), also shows problematic results, with several negative coefficients of the drivers and extremely disproportional segmentation mixtures. Of the solutions obtained, the K = 4 solution provided by the proposed Bayesian variable selection-based methodology appears to have the highest face validity and practicality from a marketing perspective.[3]

The proposed Bayes model can also provide information about credible intervals and posterior density distributions for the parameters of interest in a straight forward manner. Figure 1 shows six illustrative cumulative distribution plots for several posterior mean coefficients. Panels A and B represent the posterior mean coefficient plots of a strongly selected variable (odds ratio > 100), Panels C and D represent the posterior mean coefficient plots of moderately strongly selected variables (20 < odds ratio ≤ 100), and Panels E and F represent the posterior mean coefficient plot of an unselected variable (odds ratio ≤ 20). The coefficient of the product reliability variable in Segment 2 and price variable in Segment 4 (Figure 1, Panels A and B) appear clearly significant and were selected with sampled mean value of approximately .35–.4. In contrast, the coefficients of the records variable in Segment 1 and the acctrep variable in Segment 4 (Figure 1, Panels E and F) were not selected because they are mostly sampled with zero. Here, for moderately strongly selected variables, we sampled the coefficient of the maintenance variable in Segment 1 (Figure 1, Panel C) with mean value of near .5, but the cumulative slope is less steep than product reliability variable in Segment 2. Last, customer service in Segment 3 (Figure 1, Panel D) shows somewhat mixed sampling from approximately .25 and zero values. In summary, these diagnostic tools graphically portray the degree of influence of the various independent variables on the dependent variable within their segments.

[3]Unfortunately, predictive validation in such segmented regression models with real data is difficult to implement because predictions for holdout samples are conditional on (known) segment membership, which is rarely known for post hoc segmentation schemes a priori in practice.

### Table 9
MODEL SELECTION RESULTS FOR THE CSM EXAMPLE USING THE PROPOSED BAYESIAN FINITE MIXTURE REGRESSION MODEL WITH VARIABLE SELECTION

| | *K = 1* | *K = 2* | *K = 3* | *K = 4* | *K = 5* |
|---|---|---|---|---|---|
| Log marginal likelihood | −1298 | −1181 | −1127 | −413[a] | −729 |

[a]Indicates the most parsimonious solution selected by this criterion.

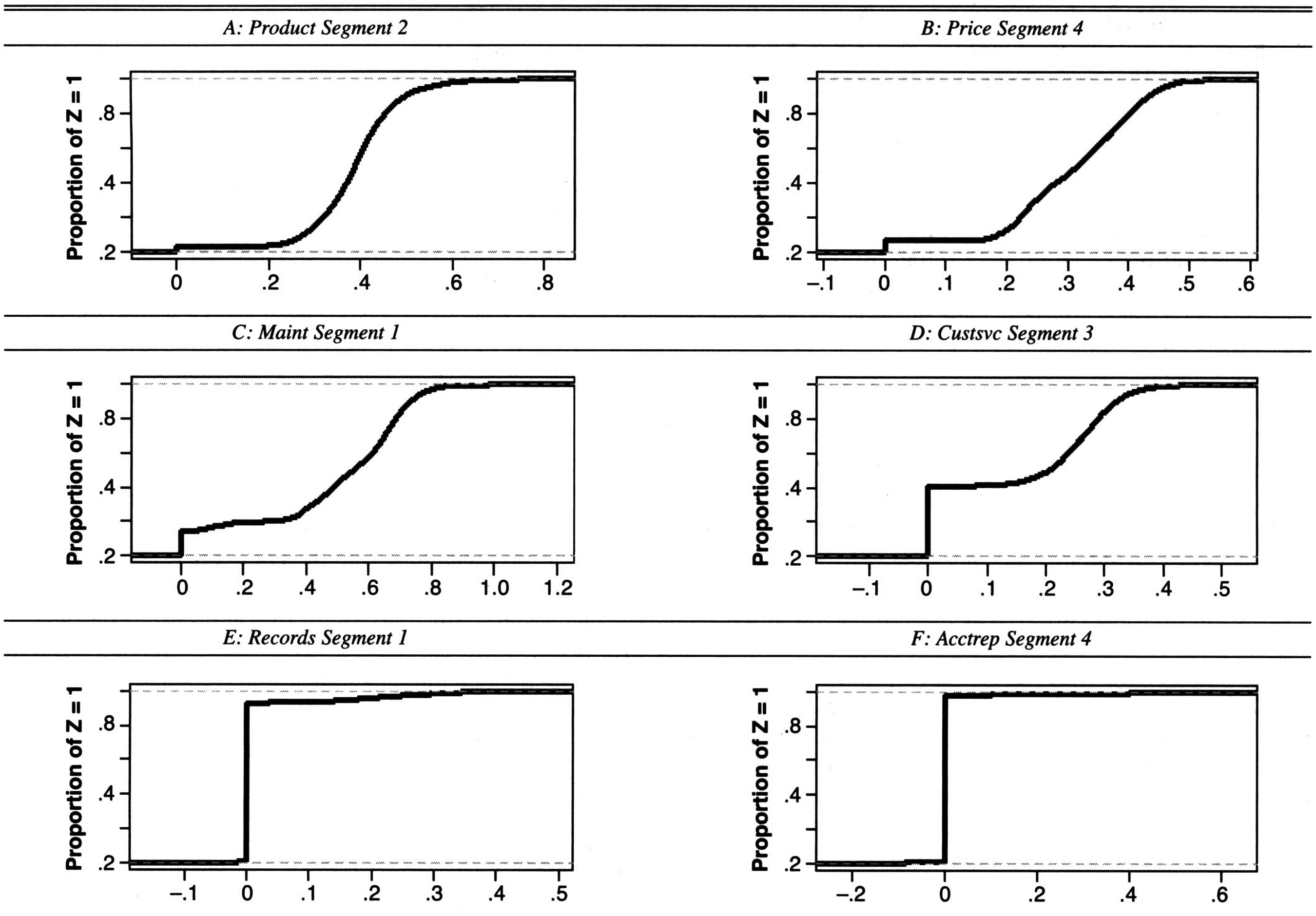
### Table 10
BAYESIAN FINITE MIXTURE REGRESSION MODEL WITH VARIABLE SELECTION RESULTS (K = 4) FOR THE CSM EXAMPLE

| | *K = 4 Solution* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *Segment 1* | | *Segment 2* | | *Segment 3* | | *Segment 4* | |
| | *Coefficient* | *Odds Ratio* | *Coefficient* | *Odds Ratio* | *Coefficient* | *Odds Ratio* | *Coefficient* | *Odds Ratio* |
| Intercept | −.007 | .93 | .000 | .02 | .002 | .24 | .088 | 5.17 |
| Price | .002 | .26 | .000 | .02 | .020 | 1.61 | .302[a] | 245.44 |
| Product | .142 | 15.87 | .384[a] | 618.92 | .496[a] | 232.83 | .020 | 1.34 |
| Maint | .482[a] | 94.80 | .067 | 6.04 | .031 | 1.42 | .028 | 1.84 |
| Repair | .018 | 1.82 | .008 | 1.43 | .021 | 3.34 | .032 | 3.75 |
| Acctrep | .048 | 2.96 | .216[a] | 2.85 | .122 | 7.21 | .006 | .58 |
| Techsup | .280[a] | 88.58 | .150[a] | 21.35 | .036 | 3.60 | .075 | 8.38 |
| Custsvc | .019 | 1.29 | .115 | 8.70 | .156[a] | 23.78 | .008 | 1.05 |
| Records | .016 | 1.33 | .073 | 8.54 | .095[a] | 2.81 | .005 | .78 |
| Billing | .002 | .37 | .002 | .53 | .016 | 1.53 | .232[a] | 248.46 |
| Memb | 287 (19%) | | 433 (28.7%) | | 466 (30.9%) | | 323 (21.4%) | |

[a]Odds ratio > 20.

Figure 1

SELECT CUMULATIVE DISTRIBUTION PLOTS OF POSTERIOR MEAN COEFFICIENTS IN THE CSM STUDY




*DISCUSSION*

We describe the technical details of a new model-based segmentation approach involving a Bayesian latent structure regression model with variable selection. We also systematically test the proposed model with two existent approaches: the standard latent class regression model (Flexmix) and the Bayesian finite mixture regression model with no variable selection (RegmixMH). Our proposed Bayesian model simultaneously provides segmentation, regression analysis, and variable selection results, and it is particularly useful in situations in which sample sizes are somewhat limited and there are many independent variables. Through the performance examination from the Monte Carlo analysis (described in Web Appendix E at www.marketingpower.com/jmr_webappendix), we find that the proposed Bayes model dominates the traditional competing methods by displaying stable performances in three of the four dependent measures over various levels of the six experimental factors. From the CSM study application, the proposed Bayesian variable selection approach appears to have higher face validity and practicality/usefulness from a managerial perspective.

As with any research, there are potential limitations. First, more sensitivity analysis is required with respect to the robustness of the proposed procedure to different stochastic specifications of the prior distributions. Second, further research should explore predictive validation in more earnest with calibration and holdout samples in a comparative fashion for the various competing procedures. Third, the stability issue of the proposed solution for label switching should be examined more closely when any of our model assumptions are violated. For further research, the proposed Bayesian finite mixture regression model with variable selection could be extended to accommodate various marketing applications in which there is a need to constrain or restrict the pattern of variable selection. For example, some variables or features can be grouped a priori for specified reasons (e.g., the inherent correlation between independent variables, managerial purposes). In other situations, there may be a competing relationship between predictors such that either one predictor or the other is active, but not both. Future extensions of the model may reflect such competing relations between predictors and constrain the selection of nonoverlapping variables during the classification process. In addition, it would be worthwhile to explore the effect of multicollinearity on the performance of these various procedures. Finally, extending the proposed methodology to choice and conjoint settings would prove beneficial.

REFERENCES

Allenby, M. Greg, Neeraj Arora, and James L. Ginter (1998), "On the Heterogeneity of Demand," *Journal of Marketing Research*, 35 (August), 384–89.

Andrews, Rick L., Asim Ansari, and Imran S. Currim (2002), "Hierarchical Bayes Versus Finite Mixture Conjoint Analysis Models: A Comparison of Fit, Prediction, and Partworth Recovery," *Journal of Marketing Research*, 39 (February), 87–98.

Benaglia, Tatiana, Didier Chauveau, David R. Hunter, and Derek S. Young (2009), "Mixtools: An R Package for Analyzing Finite Mixture Models," *Journal of Statistical Software*, 32 (6), 1–29.

Berkhof, Johannes, Iven Van Mechelen, and Andrew Gelman (2003), "A Bayesian Approach to the Selection and Testing of Mixture Models," *Statistica Sinica*, 13 (2), 423–42.

Brusco, Michael J. and Dennis J. Cradit (2001), "A Variable-Selection Heuristic for K-Means Clustering," *Psychometrika*, 66 (2), 249–70.

Büschken, Joachim, Thomas Otter, and Greg Allenby (2010), "Do We Halo or Form? A Bayesian Mixture Model for Customer Satisfaction Data," working paper, Catholic University of Eichstätt-Ingolstadt, (accessed June 18, 2012), [available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1620863].

Chen, Ming-Hui, Qi-Man Shao, and Joseph George Ibrahim (2000), *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.

Dean, Nema and Adrian E. Raftery (2010), "Latent Class Analysis Variable Selection," *Annals of the Institute of Statistical Mathematics*, 62 (1), 11–35.

DeSarbo, Wayne S., A. Selin Atalay, David LeBaron, and Simon J. Blanchard (2008), "Estimating Multiple Consumer Segment Ideal Points from Context-Dependent Survey Data," *Journal of Consumer Research*, 35 (1), 142–53.

———, Douglas J. Carroll, Linda A. Clark, and Paul E. Green (1984), "Synthesized Clustering: A Method for Amalgamating Alternative Clustering Bases with Differential Weighting of Variables," *Psychometrika*, 49 (1), 57–78.

——— and William L. Cron (1988), "A Maximum Likelihood Methodology for Clusterwise Linear Regression," *Journal of Classification*, 5 (2), 249–82.

——— and Christian F. DeSarbo (2003), "A Generalized Normative Segmentation Methodology Employing Conjoint Analysis," in *Conjoint Measurement: Methods and Applications*, Andreas Herrman and Frank Huber Anders Gustafsson, eds. Berlin: Springer-Verlag, 321–45.

———, Rajdeep Grewal, and Crystal J. Scott (2008), "A Clusterwise Bilinear Multidimensional Scaling Methodology for Simultaneous Segmentation and Positioning Analyses," *Journal of Marketing Research*, 45 (August), 280–92.

——— and Douglas Grisaffe (1998), "Combinatorial Optimization Approaches to Constrained Market Segmentation: An Application to Industrial Market Segmentation," *Marketing Letters*, 9 (2), 115–34.

———, Wagner A. Kamakura, and Michel Wedel (2006), "Latent Structure Regression," in *The Handbook of Marketing Research*, Rajiv Grover and Marco Vriens Grover, eds. Thousand Oaks, CA: Sage Publications, 394–417.

——— and Vijay Mahajan (1984), "Constrained Classification: The Use of A Priori Information in Cluster Analysis," *Psychometrika*, 49 (2), 187–215.

———, Richard L. Oliver, and Arvind Rangaswamy (1989), "A Simulated Annealing Methodology for Clusterwise Linear Regression," *Psychometrika*, 54 (4), 707–736.

De Soete, Geert, Wayne S. DeSarbo, and Douglas J. Carroll (1985), "Optimal Variable Weighting for Hierarchical Clustering: An Alternating Least-Squares Algorithm," *Journal of Classification*, 2 (1), 173–92.

Diebolt, Jean and Christian P. Robert (1994), "Estimation of Finite Mixture Distributions Through Bayesian Sampling," *Journal of the Royal Statistical Society*, 56 (2), 363–75.

Fan, Jianqing and Runze Li (2001), "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of American Statistical Association*, 96 (456), 1348–60.

Fong, Duncan K.H. and Wayne S. DeSarbo (2007), "A Bayesian Methodology for Simultaneously Detecting and Estimating Regime Change Points and Variable Selection in Multiple Regression Models for Marketing Research," *Quantitative Marketing and Economics*, 5 (4), 427–53.

Fowlkes, Edward B., Ram Gnanadesikan, and Jon R. Kettenring (1988), "Variable Selection in Clustering," *Journal of Classification*, 5 (2), 205–228.

——— and Collin L. Mallows (1983), "A Method for Comparing Two Hierarchical Clusterings (with Comments and Rejoinder)," *Journal of American Statistical Association*, 78 (383), 553–69.

Frank, Ronald E., William F. Massy, and Yoram Wind (1972), *Market Segmentation*. Englewood Cliffs, NJ: Prentice Hall.

George, Edward I. and Robert E. McCulloch (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88 (423), 881–89.

——— and ——— (1997), "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7 (2), 339–74.

Gilbride, Timothy J., Greg M. Allenby, and Jeff D. Brazell (2006), "Models for Heterogeneous Variable Selection," *Journal of Marketing Research*, 43 (August), 420–30.

Gnanadesikan, Ram, Jon R. Kettenring, and Shiao Li Tsao (1995), "Weighting and Selection of Variables for Cluster Analysis," *Journal of Classification*, 12 (1), 113–36.

Grün, Bettina and Friedrich Leisch (2008), "Flexmix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters," *Journal of Statistical Software*, 28 (4), 1–35.

Gupta, Mayetri and Joseph G. Ibrahim (2007), "Variable Selection in Regression Mixture Modeling for the Discovery of Gene Regulatory Networks," *Journal of the American Statistical Association*, 102 (479), 867–80.

Hurn, Merrilee, Ana Justel, and Christian P. Robert (2003), "Estimating Mixtures of Regressions," *Journal of Computational and Graphical Statistics*, 12 (1), 55–79.

Ishwaran, Hemant and J. Sunil Rao (2005), "Spike and Slab Variable Selection: Frequentist and Bayesian Strategies," *The Annals of Statistics*, 33 (2), 730–73.

Jasra, A., C.C. Holmes, and D.A. Stephens (2005), "Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling," *Statistical Science*, 20 (1), 50–67.

Jeffreys, Harold (1961), *Theory of Probability*. Oxford: Oxford University Press.

Khalili, Abbas and Jiahua Chen (2007), "Variable Selection in Finite Mixture of Regression Models," *Journal of the American Statistical Association*, 102 (479), 1025–38.

Kim, Sinae, Mahlet G. Tadesse, and Marina Vannucci (2006), "Variable Selection in Clustering via Dirichlet Process Mixture Models," *Biometrika*, 93 (4), 877–93.

Lee, Kate, Jean-Michel Marin, Kerrie Mengersen, and Christian Robert (2009), "Bayesian Inference on Finite Mixtures of Distributions," in *Statistical Science and Interdisciplinary Research, Vol. 7: Perspectives in Mathematical Sciences I: Probability and Statistics*, N.S. Narasimha Sastry, T.S.S.R.K. Rao, Mohan Delampady, and B. Rajeev, eds. Hackensack, NJ: World Scientific, 165–202.

Lenk, J. Peter and Wayne S. DeSarbo (2000), "Bayesian Inference for Finite Mixtures of Generalized Linear Models with Random Effects," *Psychometrika*, 65 (1), 93–119.

Marin, Jean-Michel and Christian P. Robert (2007), *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. New York: Springer-Verlag.

Mitchell, T.J. and J.J. Beauchamp (1988), "Bayesian Variable Selection in Linear Regression," *Journal of the American Statistical Association*, 83 (404), 1023–32.

Naik, Prasad A., Peide Shi, and Chih-Ling Tsai (2007), "Extending the Akaike Information Criterion to Mixture Regression Models," *Journal of the American Statistical Association*, 102 (477), 244–54.

Otter, Thomas, Regina Tüchler, and Sylvia Frühwirth-Schnatter (2004), "Capturing Consumer Heterogeneity in Metric Conjoint Analysis Using Bayesian Mixture Models," *International Journal of Research in Marketing*, 21 (3), 285–97.

Quandt, Richard E. (1972), "A New Approach to Estimating Switching Regressions," *Journal of the American Statistical Association*, 67 (338), 306–310.

——— and James B. Ramsey (1978), "Estimating Mixtures of Normal Distributions and Switching Regressions," *Journal of the American Statistical Association*, 73 (364), 730–38.

Raftery, Adrian E. and Nema Dean (2006), "Variable Selection for Model-Based Clustering," *Journal of the American Statistical Association*, 101 (473), 168–78.

Roberts, John H., Ujwal Kayande, and Stefan Stremersch (2009), "From Academic Research to Marketing Practice: Exploring the Marketing Science Value Chain," working paper, ANU College of Business & Economics, Australian National University.

Smith, Wendell R. (1956), "Product Differentiation and Market Segmentation as Alternative Marketing Strategies," *Journal of Marketing*, 21 (January), 3–8.

Späth, Helmuth (1982), "Algorithm 48: A Fast Algorithm for Clusterwise Linear Regression," *Computing*, 29 (2), 175–81.

Sperrin, Matthew, Thomas Jaki, and Ernst Wit (2010), "Probabilistic Relabelling Strategies for the Label Switching Problem in Bayesian Mixture Models," *Statistics and Computing*, 20 (3), 357–66.

Städler, Nicolas, Peter Bühlmann, and Sara van de Geer (2010), "$\ell_1$-Penalization for Mixture Regression Models," *Test*, 19 (2), 209–285.

Stephens, Matthew (1997), "Bayesian Methods for Mixtures of Normal Distributions," doctoral dissertation, Magdalen College, University of Oxford.

Tadesse, Mahlet G., Naijun Sha, and Marina Vannucci (2005), "Bayesian Variable Selection in Clustering High-Dimensional Data," *Journal of the American Statistical Association*, 100 (470), 602–617.

Tibshirani, Robert (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, 58 (1), 267–88.

Wedel, Michel and Wayne S. DeSarbo (1994), "A Review of Recent Developments in Latent Class Regression Models," in *Advanced Methods of Marketing Research*, Richard P. Bagozzi, ed. Malden, MA: Blackwell Publishers.

——— and ——— (1995), "A Mixture Likelihood Approach for Generalized Linear Models," *Journal of Classification*, 12 (1), 21–55.

——— and Wagner A. Kamakura (2000), *Market Segmentation: Conceptual and Methodological Foundations*. Norwell, MA: Kluwer.

——— and Cor Kistemaker (1989), "Consumer Benefit Segmentation Using Clusterwise Linear Regression," *International Journal of Research in Marketing*, 6 (1), 45–59.

——— and Jan-Benedict E.M. Steenkamp (1991), "A Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Benefit Segmentation," *Journal of Marketing Research*, 28 (November), 385–96.