Working Paper Series

Adaptive Information Systems and Modelling in Economics and Management Science

A Fully Bayesian Analysis of Multivariate Latent Class Models with an Application to Metric Conjoint Analysis

Sylvia Frühwirth-Schnatter
Thomas Otter
Regina Tüchler

Working Paper No. 89
June 2002

Working Paper Series

June 2002

SFB
'Adaptive Information Systems and Modelling in Economics and Management Science'

Vienna University of Economics
and Business Administration
Augasse 2–6, 1090 Wien, Austria

in cooperation with
University of Vienna
Vienna University of Technology

http://www.wu-wien.ac.at/am

This piece of research was supported by the Austrian Science Foundation (FWF) under grant SFB#010 ('Adaptive Information Systems and Modelling in Economics and Management Science').

A Fully Bayesian Analysis of Multivariate Latent Class Models with an Application to Metric Conjoint Analysis

by

Sylvia Frühwirth-Schnatter [1], Thomas Otter [2] and Regina Tüchler [3]

May 2000

Abstract

In this paper we head for a fully Bayesian analysis of the latent class model with a priori unknown number of classes. Estimation is carried out by means of Markov Chain Monte Carlo (MCMC) methods. We deal explicitely with the consequences the unidentifiability of this type of model has on MCMC estimation. Joint Bayesian estimation of all latent variables, model parameters, and parameters determining the probability law of the latent process is carried out by a new MCMC method called permutation sampling. In a first run we use the random permutation sampler to sample from the unconstrained posterior. We will demonstrate that a lot of important information, such as e.g. estimates of the subject-specific regression coefficients, is available from such an unidentified model. The MCMC output of the random permutation sampler is explored in order to find suitable identifiability constraints. In a second run we use the permutation sampler to sample from the constrained posterior by imposing identifiablity constraints.

The unknown number of classes is determined by formal Bayesian model comparison through exact model likelihoods. We apply a new method of computing model likelihoods for latent class models which is based on the method of bridge sampling.

The approach is applied to simulated data and to data from a metric conjoint analysis in the Austrian mineral water market.

Keywords. Bayesian analysis, conjoint analysis, latent class models, MCMC methods, model selection

1 Introduction

In the present paper we will address the problem of accounting for unobserved heterogeneity among repeated measurements of various subjects. We assume that the dependent data arise from a multivariate normal distribution where the mean depends on design variables through a multivariate regression model. A common way of including unobserved heterogeneity into such a model is the finite mixture or latent class model, where the unknown distribution of subject-specific regression coefficients is approximated by a discrete distribution with unknown support vectors and unknown group probabilities. This leads to a multivariate mixture of normals as marginal distribution for the data. Our interest for this model class has been motivated by work on accounting for unobserved heterogeneity among consumers within conjoint analysis. This issue has received considerable attention during the past years in the marketing community (Allenby and Ginter, 1995; DeSarbo *et al.*, 1992; Elrod and Häubl, 1997; Hagerty, 1985; Kamakura, 1988; Kamakura *et al.*, 1994; Lenk *et al.*, 1996; Wedel and Steenkamp, 1991).

The area, however, is also of greatest interest from a statistical point of view. Involving highly multivariate mixtures with an unknown number of components, it provides an excellent testing ground for statistical as well as computational advances made in these fields.

In this paper we head for a fully Bayesian analysis of the latent class model. Estimation is carried out by means of Markov Chain Monte Carlo (MCMC) methods (see e.g. Dieboldt and Robert, 1994 and Frühwirth-Schnatter, 1999a for a general discussion of MCMC estimation of mixture models). The unknown number of classes is determined by formal Bayesian model comparison through exact model likelihoods.

[1]Department of Statistics, University of Business Administration and Economics, Augasse 2-6, A-1090 Vienna, Austria, e-mail: sfruehwi@isis.wu-wien.ac.at

[2]Department of Advertising and Marketing Research, University of Business Administration and Economics, Augasse 2-6, A-1090 Vienna, Austria, e-mail: thomas.otter@wu-wien.ac.at

[3]Department of Statistics, University of Business Administration and Economics, Augasse 2-6, A-1090 Vienna, Austria, e-mail: regina.tuechler@wu-wien.ac.at

Such a fully Bayesian analysis of the latent class model has – at least in principle – been tried before, as it could be viewed as a special case of the finite mixture of generalized linear models with random effects discussed in Lenk and DeSarbo (1999) and Allenby *et al.* (1998). However, our approach differs from these in various respects. We deal in a different way with the unidentifiability of the latent class model and we use an alternative method for computing the model likelihood.

It is well known, that the latent class model, like any model including discrete latent variables, is only identified up to permutations of the labelling of the groups. The full unconstrained posterior of the latent class model with K classes is multimodal with at most $K!$ modes. When applying MCMC methods to such a posterior, we have to be aware of the problem of label switching which might render estimation of group specific quantities meaningless. Allenby *et al.* (1998) and Lenk and DeSarbo (1999) apply a standard order constraint on the weights of the mixture to circumvent the problem. The influence of constraints on the shape of the full constrained posterior has been investigated only recently (Frühwirth-Schnatter, 1999a). It turns out that only a constraint which respects the geometry of the posterior will restrict the posterior to a subspace with unique labelling.

Because of these problems with constrained estimation we start with MCMC estimation of the unconstrained latent class model using the random permutation sampler suggested in Frühwirth-Schnatter (1999a). We will demonstrate that a lot of important information, such as e.g. estimates of the subject-specific regression coefficients, is available from such an unidentified model. Furthermore, the MCMC output of the random permutation sampler is used to estimate the marginal model likelihood in order to compare models, differing e.g. in the number of classes. We apply a new method of computing model likelihoods for latent class models which is based on the method of bridge sampling (Meng and Wong, 1996). In Frühwirth-Schnatter (1999b) this method has been applied to computing the model likelihood for a general mixture and switching model and clearly outperformed alternative methods such as the candidate's formula (Chib, 1995) or reciprocal importance sampling (Gelfand and Dey, 1994). Only for the „best" model we start to think about identification. We explore the output of the random permutation sampler in order to find identifiability constraints which respect the geometry of the posterior. These constraints are then included into the permutation sampler in order to obtain information on group specific parameters and weights.

The outline of the paper is as follows. In Section 2 we discuss the Bayesian estimation of the latent class model and the impact of the unidentifiability problem. Section 3 presents MCMC estimation both of the unconstrained and the constrained latent class model. Model selection for latent class models and a new method of computing model likelihoods is outlined in Section 4. In Section 5 and Section 6, respectively, the suggested approach is applied to simulated data and data from a metric conjoint analysis in the Austrian mineral water market.

2 Bayesian Analysis of the Multivariate Latent Class Model

2.1 Notation and Specification

We start by defining the latent class model in a way that is well known from linear mixed modelling:

$$y_i = Z_i\alpha + W_i\beta_i + \varepsilon_i, \quad \varepsilon_i \sim \mathrm{N}(0, R_i) \tag{1}$$

where y_i is a vector of T_i repeated measurements for subject i, α are *fixed* effects which are constant for all subjects and β_i are *random* effects which due to heterogeneity are different for each subject. Z_i and W_i are the design matrices for the fixed effects α and the random effects β_i, respectively. Within latent class models the unknown distribution $\pi(\beta_i)$ of heterogeneity is approximated by a discrete distribution with unknown support vectors $\beta_1^G, \dots, \beta_K^G$ and unknown group probabilities $\eta = (\eta_1, \dots, \eta_K)$. Such a distribution could be written as:

$$\beta_i = \begin{cases} \beta_1^G, & \text{if } S_i = 1, \\ \vdots \\ \beta_K^G, & \text{if } S_i = K, \end{cases} \tag{2}$$

if we introduce a discrete latent group indicator S_i taking values in $\{1, \dots, K\}$ with unknown probability distribution $\Pr(S_i = k) = \eta_k$, $k = 1, \dots, K$.

In what follows we will assume that the model is identified up to relabelling the number of the group indicator. The general model appearing in (1) is not necessarely identified in this sense for arbitrary choices of Z_i and W_i. To give an example, the model is not identified, if some columns of Z_i and W_i are identical.

Unknown parameters which have to be estimated from the data are the fixed effects α, the group specific parameters $\beta_1^G, \dots, \beta_K^G$, the group probabilities $\eta = (\eta_1, \dots, \eta_K)$ and unknown parameters θ appearing in the definition of the observation variance R_i. These parameters will be summarized by ϕ: $\phi = (\alpha, \beta_1^G, \dots, \beta_K^G, \eta, \theta)$. Within a Bayesian approach the latent group indicator $S^N = (S_1, \dots, S_N)$ is viewed as missing data and is estimated along with the model parameter ϕ. This data augmentation is quite common within Bayesian analysis (see e.g. Tanner, 1993). Bayesian estimation of the model is based on the hierarchical structure of the model:

1. Conditional on ϕ and S^N, the complete data likelihood $f(y_1, \dots, y_N|S^N, \phi)$ factorizes into the product of normal distributions where the distribution of each y_i depends on S^N through the group specific parameter, only: $f(y_1, \dots, y_N|S^N, \phi) = \prod_{i=1}^N f(y_i|\alpha, \beta_{S_i}^G, \theta)$.
2. Conditional on ϕ the „prior" of S^N is given by the discrete probability distribution $\Pr(S_i = k) = \eta_k$ and the assumption that S_i and S_j are pairwise independent, therefore $\pi(S^N|\phi) = \prod_{k=1}^K \eta_k^{N_k}$, $N_k = \#\{S_i = k\}$.
3. Finally, ϕ has a prior distribution $\pi(\phi)$.

Note that the „prior" on S^N appearing in the second level is not a subjective prior, but part of the model. Only the prior on ϕ appearing on the third level has a subjective flavour. In this paper the focus lies on Bayesian estimation in situations where we lack strong prior information. From a theoretical point of view, being fully non-informative about ϕ is possible only for the fixed effects α and the variance parameter θ, if θ is class independent. Theoretically, being non-informative about $\beta_1^G, \dots, \beta_K^G$, η and class dependent variance parameters θ is not possible, as improper priors on $\beta_1^G, \dots, \beta_K^G$ and η result in improper posteriors (Diebolt and Robert, 1994; Roeder and Wasserman, 1997).

For mixture models η is commonly assumed to be independent from the remaining parameters of ϕ. A „natural" prior distribution $\pi(\eta)$ for η is a Dirichlet prior $D(e_{01}, \dots, e_{0K})$, which is the conjugate prior in the complete data setting, where S^N is assumed to be known. A common choice is $e_{0j} = 1$ which leads to a uniform prior on the unit simplex. We may select e_{0j} bigger than 1 to exclude empty classes *a priori*. For the fixed effects α we use a normal prior $N(c_0, C_0)$. Concerning the group specific parameters $\beta_1^G, \dots, \beta_K^G$, we assume that they are independent *a priori*. In the context of mixture modelling it is now common practice to use hierarchical priors for being weakly informative about group specific parameters (see e.g. Richardson and Green, 1997; Roeder and Wasserman, 1997; Stephens, 1997): $\pi(\beta_k^G) \propto N(b_0, B_0)$. This allows different parameters for the various groups, however with a slight restriction expressed by the prior. Furthermore this prior is invariant to relabelling the number of the groups. The prior on the variance parameter θ depends on the model chosen for the observation variance R_i. If $R_i = \sigma_\varepsilon^2 \cdot I$ with I being the identity matrix, then a "natural" prior for σ_ε^2 is an inverted gamma prior $\sim IG(\nu_{\varepsilon,0}, G_{\varepsilon,0})$, which is the conjugate prior in the complete data setting, where S^N is assumed to be known.

2.2 Posterior Analysis and Label Switching

Using Bayes' theorem we obtain from the hierarchical structure of the model that the non-normalized posterior distribution $\pi(\phi, S^N|y^N)$ of the augmented vector $\psi = (\phi, S^N)$ is proportional to the following product:

$$\pi(\phi, S^N|y^N) \propto f(y^N|S^N, \phi)\pi(S^N|\phi)\pi(\phi), \tag{3}$$
$$f(y^N|S^N, \phi) = \prod_{i=1}^N f(y_i|\alpha, \beta_{S_i}^G, \theta),$$

where $y^N = (y_1, \dots, y_N)$. For models including a latent, discrete structure such as S^N the unconstrained posterior has some characteristic properties (see Stephens, 1997; Celeux, 1998; Frühwirth-Schnatter, 1999a). The unconstrained parameter space contains $K!$ subspaces, each one corresponding to a different way of labelling the groups. The „complete data likelihood" $f(y^N|S^N, \phi)$, and the „prior" $\pi(S^N|\phi)$ are invariant to relabelling the groups. Therefore, if the prior $\pi(\phi)$, is invariant, too, the unconstrained posterior typically is multimodal and invariant to relabelling the groups. This special structure of the posterior has important consequences for estimation. The unconstrained model is not identified in a strict sense. In section 3 we use MCMC methods to obtain a sample $(S^N, \phi)^{(1)}, \dots, (S^N, \phi)^{(M)}$ from the unconstrained posterior. However, we do not know which of the labelling subspaces a sampled value $(S^N, \phi)^{(m)}$ belongs to, since label switching (jumping between the various labelling subspaces) might have occured. Thus we are not allowed to estimate functionals $f(\psi)$ of $\psi = (S^N, \phi)$ which are not invariant to relabelling the groups from MCMC simulations from the unconstrained posterior.

Note that a lot of information is available from MCMC simulations from the unconstrained posterior *without* introducing a unique labelling, as we are allowed to estimate functionals $f(\psi)$ of $\psi = (S^N, \phi)$ which are invariant to relabelling of S_i. Straightforward examples are parameters which are common to all groups, such as the fixed effects α or common variance parameters θ. Further examples are moments of the distribution of heterogeneity, e.g. the mean or the covariance matrix:

$$\alpha_\beta = \sum_{k=1}^{K} \beta_k^G \eta_k, \quad Q = \sum_{k=1}^{K} \beta_k^G (\beta_k^G)' \eta_k - \alpha_\beta \alpha_\beta'.$$

These moments may be estimated from the MCMC simulations by:

$$\hat{\alpha}_\beta = \frac{1}{M} \sum_{m=1}^{M} \sum_{k=1}^{K} (\beta_k^G)^{(m)} \eta_k^{(m)}, \quad \hat{Q} = \frac{1}{M} \sum_{m=1}^{M} \sum_{k=1}^{K} (\beta_k^G)^{(m)} ((\beta_k^G)^{(m)})' \eta_k^{(m)} - \hat{\alpha}_\beta \hat{\alpha}_\beta'.$$

Moreover, subject specific estimates of the random effects β_i which may be written as:

$$\beta_i = \sum_{k=1}^{K} \beta_k^G I_k(S_i),$$

where $I_k(S_i) = 1$ iff $S_i = k$, and obviously are invariant to relabelling can be derived from an unidentified model. An estimate of β_i is obtained from the MCMC simulations by:

$$\hat{\beta}_i = \frac{1}{M} \sum_{m=1}^{M} (\beta_s^G)^{(m)}, \quad s = S_i^{(m)}.$$

Finally, it is possible to predict the behaviour of each subject under designs $Z_i^\star$, $W_i^\star$ different from the ones used for estimation.

In order to estimate functionals $f(\psi)$ of $\psi = (S^N, \phi)$ which are not invariant to relabelling the groups such as $\beta_1^G, \dots, \beta_K^G$, η or the classification probabilities $Pr(S_i = k|y^N)$ we have to identify the model in the sense that we allow for MCMC simulations from a unique labelling subspace, only. A common way of dealing with the problem is to include an identifiability constraint. An arbitrary constraint, however, does not necessarily induce a unique labelling, if it ignores the geometry of the unconstrained posterior distribution (see Frühwirth-Schnatter, 1999b). Only a carefully selected constraint will separate the labelling subspaces and induce unique labelling. We will demonstrate in our case studies, how suitable identifiability constraints may be found by exploring MCMC simulations from the unconstrained posterior distribution.

If the model has been identified, estimates of group specific parameters are simply the mean of the simulations, whereas the classification probabilities $\text{Pr}(S_i = k|y^N), k = 1, \dots, K$ may be estimated for all subjects i from the sampled values $S_i^{(m)}, m = 1, \dots, M$ by

$$\hat{\text{Pr}}(S_i = k|y^N) = \frac{1}{M} \#\{S_i^{(m)} = k\}$$

3 Estimation of the Latent Class Model via MCMC Methods

3.1 MCMC Methods

A common way to deal with complex posterior distributions such as the posterior (3) is to sample from the posterior by some MCMC method (see e.g. Smith and Roberts, 1993, for a general introduction to MCMC methods). Applications of MCMC methods to classical mixture models appear e.g. in Diebolt and Robert (1994) and Richardson and Green (1997). MCMC techniques for sampling from a complicated posterior density split the joint unknown parameter into blocks and sample then from the conditonal posterior densities of each block given the fixed values for the other blocks. Sampling from the posterior of a latent class model is possible within the following four blocks:

(i) Sample S^N from $\pi(S^N|\eta, \alpha, \beta_1^G, \dots, \beta_K^G, \theta, y^N)$

(ii) Sample η from $\pi(\eta|S^N)$

(iii) Sample the fixed and the group specific effects from $\pi(\alpha, \beta_1^G, \dots, \beta_K^G|\theta, S^N, y^N)$

(iv) Sample the variance parameters θ from $\pi(\theta|\alpha, \beta_1^G, \dots, \beta_K^G, S^N, y^N)$

Details concerning the structure of these posteriors and the method for sampling from them will be given in subsection 3.2. Most of this material is standard with the exception of step (iii), where the fixed effects α and the group specific parameters $\beta_1^G, \dots, \beta_K^G$ are sampled jointly within one block. We do not recommend to sample these parameters in two different blocks from the conditional posteriors $\pi(\alpha|\beta_1^G, \dots, \beta_K^G, \theta, S^N, y^N)$ and $\pi(\beta_1^G, \dots, \beta_K^G|\alpha, \theta, S^N, y^N)$, respectively, as suggested e.g. in McCulloch and Tsay (1994) for the related MSAR-model. If there exist strong correlations between columns of Z_i and columns of W_i, such separate sampling will converge slowly. We will show in subsection 3.2 that joint sampling of all effects is possible, as conditional on S^N the latent class model may be rewritten as a classical regression model.

There exist various ways to run through this scheme and the suitable method for MCMC sampling depends on what kind of inference is of interest. An unconstrained model may be estimated by *unconstrained Gibbs sampling* running through step (i)–(iv) without any constraint on the group specific parameters. Unconstrained Gibbs sampling, however, does not explore the whole unrestricted parameter space, but tends to stick at the current labelling subspace with occasionally switching to other labelling subspaces. Some of the labelling subspaces will never be visited. An alternative method of estimating an unconstrained model is *random permutation sampling* (Frühwirth-Schnatter, 1999a). This method is simply an unconstrained Gibbs sampler concluded by a randomly selected permutation $\rho(1), \dots, \rho(K)$ of the current labelling $1, \dots, K$. After sampling ψ by an unconstrained Gibbs sampler, group dependent parameters are permuted in the following way:

$$\begin{aligned}
&(\beta_1^G, \dots, \beta_K^G) := (\beta_{\rho(1)}^G, \dots, \beta_{\rho(K)}^G), \\
&(\eta_1, \dots, \eta_K) := (\eta_{\rho(1)}, \dots, \eta_{\rho(K)}), \\
&(S_1, \dots, S_N) := (\rho(S_1), \dots, \rho(S_N)).
\end{aligned} \tag{4}$$

Parameters which are group independent such as α and θ are not permuted. The permuted parameters are the starting point for the next Gibbs step. This sampler is an appropriate method for exploring the whole space of the unconstrained posterior as it delivers a sample from the unconstrained posterior where balanced label switching occurs and all labelling subspaces are visited with the same probability.

To estimate a constrained model an identifiability constraint may be introduced into the sampling scheme. One way of imposing the identifiability constraint is to introduce some truncation or rejection method into step (iii) in order to obtain simulations which fulfill the constraint. This *constrained Gibbs sampling* is the standard method applied so far (Allenby *et al.*, 1998; Lenk and DeSarbo, 1999). An alternative method for constrained sampling is *permutation sampling under an identifiability constraint* (Frühwirth-Schnatter, 1999a). Unconstrained Gibbs sampling is concluded by a permutation as in (4), but this time the permutation is selected in such a way that the identifiability constraint is fulfilled. We mentioned already that an arbitrary constraint does not necessarily induce a unique labelling and a bias toward the constraint may be introduced. The poorness of the constraint may go undetected if we use constrained Gibbs sampling and the sampler sticks at the current labelling subspace. The permutation sampler, however, will indicate this fact and exhibit label switching. In this case a more suitable identifiability constraint may be derived from the MCMC output of the random permutation sampler.

3.2 The Structure of the Conditional Posteriors

We now discuss the structure of the various conditional posterior densities and the sampling from these posteriors in more detail.

Step (i) is a standard step occuring in mixture models and can be carried out as discussed in Diebolt and Robert (1994). From Bayes' Theorem we obtain:

$$\pi(S_1,\dots,S_N|y^N,\phi) \propto \prod_{i=1}^{N} f(y_i|\alpha,\beta^G_{S_i},\theta)\pi(S_i|\eta).$$

As $S_1,\dots,S_N$ are conditionally independent given ϕ and y^N, S_i may be sampled from the discrete distribution $\pi(S_i = k|y_i,\phi)$, $k = 1,\dots,K$:

$$\pi(S_i = k|y_i,\phi) \propto f(y_i|\beta^G_k,\alpha,\theta)\cdot\eta_k, \tag{5}$$

where $f(y_i|\beta^G_k,\alpha,\theta)$ is the density of a normal distribution with mean $Z_i\alpha + W_i\beta^G_k$ and variance $R_i(\theta)$.

Step (ii) is a standard Bayesian exercise. Given the Dirichlet prior $D(e_{01},\dots,e_{0K})$ the posterior $\pi(\eta|S^N)$ of η is $\mathcal{D}(e_{01}+N_1,\dots,e_{0K}+N_K)$, with $N_k = \#\{S_i = k\}$.

In step (iii) we sample the fixed effects α and the group specific parameters $\beta^G_1,\dots,\beta^G_k$ jointly within one block. Conditional on S^N the latent class model is a classical regression model:

$$y_i = X_i\alpha^\star + \varepsilon_i, \quad \varepsilon_i \sim \mathrm{N}(0,R_i),$$

with parameter $\alpha^\star = (\alpha,\beta^G_1,\dots,\beta^G_K)$ and

$$X_i = \begin{pmatrix} Z_i & W_iD_i^{(1)} & \cdots & W_iD_i^{(K)} \end{pmatrix},$$

where we used the coding $D_i^{(k)} = 1$ iff $S_i = k$, for $k = 1,\dots,K$. The posterior of $\alpha^\star = (\alpha,\beta^G_1,\dots,\beta^G_K)$ is given by $\alpha^\star|y^N,\theta \sim \mathrm{N}(a_N,A_N)$, where

$$A_N = (\sum_{i=1}^{N} X_i^{'}R_i^{-1}X_i + A_0^{-1})^{-1}, \tag{6}$$

$$a_N = A_N(\sum_{i=1}^{N} X_i^{'}R_i^{-1}y_i + A_0^{-1}a_0).$$

The joint normal prior $N(a_0,A_0)$ for $\alpha^\star$ is constructed in an obvious way from the normal priors $N(b_0,B_0)$ and $N(c_0,C_0)$ of the group specific parameters $\beta^G_1,\dots,\beta^G_K$, and the fixed effects α, respectively. The information matrix A_N^{-1} has a special structure which can be exploited for efficient sampling. If no fixed effects α are present, A_N^{-1} as well as A_N are block diagional and we may sample all group specific effects independently. If fixed effects α are present, then A_N^{-1} as well as A_N contain a submatrix which is block-diagonal. Therefore joint sampling of $\alpha^\star = (\alpha,\beta^G_1,\dots,\beta^G_K)$ is possible by sampling the fixed effects from the marginal posterior $N(c_N,C_N)$, where the group specific effects are integrated out, and by sampling the group specific effects independently from the conditional distributions $N(b_{N,k}(\alpha),B_{N,k}), k = 1,\dots,K$. The moments of these densities are given by:

$$b_{N,k}(\alpha) = B_{N,k}\left[\sum_{i=1}^{N} W_i^{'}R_i^{-1}D_i^{(k)}\tilde{y}_i + B_0^{-1}b_0\right],$$

$$B_{N,k} = \left[\sum_{i=1}^{N} W_i^{'}R_i^{-1}W_iD_i^{(k)} + B_0^{-1}\right]^{-1},$$

$$c_N = C_N\left[\sum_{i=1}^{N} Z_i^{'}R_i^{-1}(y_i - W_ib_{N,S_i}(0)) + C_0^{-1}c_0\right],$$

$$C_N = \left[\sum_{i=1}^{N} Z_i^{'}R_i^{-1}(R_i - W_iB_{N,S_i}W_i^{-1})R_i^{-1}Z_i + C_0^{-1}\right]^{-1},$$

where $\tilde{y}_i = y_i - Z_i\alpha$.

Step (iv) depends on the model chosen for the observation variance R_i. If $R_i = \sigma_\varepsilon^2 \cdot I$ with σ_ε^2 being group independent, then:

$$\sigma_\varepsilon^2|\alpha, \beta_1^G, \dots, \beta_K^G, y^N \sim IG(\nu_{\varepsilon,N}, G_{\varepsilon,N}),$$
$$\nu_{\varepsilon,N} = \nu_{\varepsilon,0} + N \cdot (\sum_{i=1}^{N} T_i)/2, \quad G_{\varepsilon,N} = G_{\varepsilon,0} + 1/2(\sum_{i=1}^{N} ||y_i - Z_i\alpha - W_i\beta_{S_i}^G||_2^2).$$

4 Issues in Model Selection

Model selection is based on the Bayesian model discrimination procedure where various models $\mathcal{M}_1, \dots, \mathcal{M}_K$ are compared through the posterior probability of each model (Bernardo and Smith, 1994):

$$P(\mathcal{M}_l|y^N) \propto f(y_1, \dots, y_N|\mathcal{M}_l)P(\mathcal{M}_l). \tag{7}$$

The factor $L(y^N|\mathcal{M}_l) := f(y_1, \dots, y_N|\mathcal{M}_l)$ is called model likelihood and quantifies evidence in favour of a model given the data. For a latent class model the model likelihood is given by the following integral of the marginal likelihood $L(y_1, \dots, y_N|\phi)$ with respect to the prior $\pi(\phi)$:

$$L(y^N) = \int L(y_1, \dots, y_N|\phi)\pi(\phi)d\phi, \tag{8}$$

where an explicit formula for the marginal likelihood $L(y_1, \dots, y_N|\phi)$ is available:

$$L(y_1, \dots, y_N|\phi) = \prod_{i=1}^{N} \left(\sum_{k=1}^{K} f(y_i|\beta_k^G, \alpha, \theta) \cdot \eta_k \right). \tag{9}$$

The computation of the model likelihood has proven to be challenging for models with latent processes such as the latent class model. Model likelihoods have been estimated from the MCMC output using methods such as the candidate's formula (Chib, 1995), importance sampling based on mixture approximations (Frühwirth-Schnatter, 1995), combining MCMC simulations and asymptotic approximations (Gelfand and Dey, 1994; DiCiccio *et al.*, 1997) and bridge sampling (Meng and Wong, 1996). The application of these methods to compute the model likelihood from the MCMC output for switching and mixture models has been discussed in detail in Frühwirth-Schnatter (1999b) with the following main results: first, estimation of the model likelihood turns out to be sensitive to the problem of label switching. Especially the candidate's formula (Chib, 1995) should not be applied, if label switching is present, and is a suitable estimation method only for identified models. Second, it is not necessary to identify the model in order to compute the model likelihood. This is very convenient, if we want to compare a wide range of values of K. Third, the best result with the lowest standard error is obtained by using the method of bridge sampling where the MCMC sample obtained by random permutation sampling is combined with an iid sample from an importance density $q(\phi)$. The importance density $q(\phi)$ is constructed in an unsupervised manner from the MCMC output $(\phi^{(1)}, \dots, \phi^{(M)})$ of the random permutation sampler using a mixture of complete data posteriors:

$$q(\phi) = 1/M_L \sum_{m=1}^{M_L} \pi(\phi|(S^N)^{(m)}, \phi^{(m)}, y^N). \tag{10}$$

The bridge sampling estimator outperforms other methods such as importance sampling or reciprocal importance sampling for the following reason: whereas importance sampling as well as reciprocal importance sampling are known to be sensitive to the tail behaviour of the importance density $q(\phi)$, the bridge sampling estimator turns out to be much more robust in this concern. In the present paper we will apply this method to latent class models. For further details the reader is referred to Frühwirth-Schnatter (1999b).

Within the latent class model the Bayesian approach could be applied to various issues arising in model selection, the most important one being the selection of the the number of classes. Note that this testing situation may result in a non-regular problem. Selecting the number of groups is not possible within the classical framework of maximum likelihood. Although a mixture model with K classes could be viewed as that special case of a mixture model with $K+1$ classes, the regularity conditions for justifying the χ^2-approximation to the likelihood ratio statistic do not hold, as the group specific parameters are unidentified under the hypothesis that there are really K groups.

In one of our case studies the Bayesian approach will be applied to further issues arising in model selection such as testing for heterogeneity of selected components. If the marginal densities of a certain component $\beta^G_{\cdot,r}$ of the group specific parameters $\beta^G_1, \ldots, \beta^G_K$ overlap for all groups such as in figure 13, we could formulate the hypothesis that the component $\beta^G_{\cdot,r}$ is fixed rather than random: $\beta^G_{1,r} = \ldots = \beta^G_{K,r}$. This hypothesis is tested against the hypothesis of a random component by comparing the model likelihoods. In order to obtain the model likelihood of the new hypothesis, we have to rerun MCMC estimation under the assumption that $\beta^G_{\cdot,r}$ is fixed (the corresponding column in model (1) has to be deleted from W_i and added to Z_i). Hypotheses involving more than one component such as $\beta^G_{1,r} = \ldots = \beta^G_{K,r}, \beta^G_{1,s} = \ldots = \beta^G_{K,s}$ are tested in a similar way.

A final application will be variable selection. If the marginal density of a certain component α_s of the fixed effects cover 0, we could formulate the hypothesis that this effect is not significant and should be deleted from the model: $\alpha_s = 0$. Again, this hypothesis is tested against the hypothesis that α_s should be kept in the model by comparing the model likelihoods. In order to obtain the model likelihood of the new hypothesis, we have to rerun MCMC estimation subject to $\alpha_s = 0$ (the corresponding column in model (1) has to be deleted from Z_i). Similarily, if all marginal densities of a certain component $\beta^G_{\cdot,r}$ of the group specific parameters $\beta^G_1, \ldots, \beta^G_K$ cover 0 such as in figure 13, we could test the hypothesis that all components $\beta^G_{1,r}, \ldots, \beta^G_{K,r}$ should be deleted from the model.

5 Application to Simulated Data

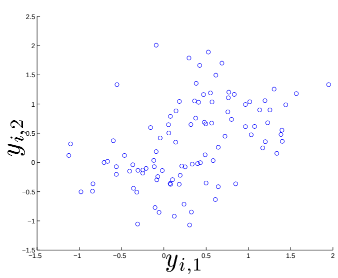

Figure 1: Simulated Data

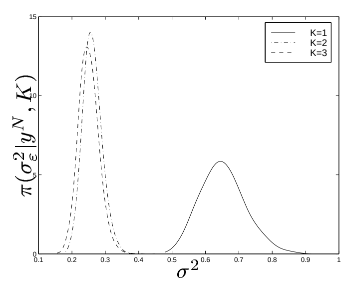

Figure 2: Marginal densities $\pi(\sigma^2_\varepsilon|y^N, K)$ estimated from a random permutation sampler

5.1 The Data

For illustration, we apply the methods of the previous sections to 100 bivariate observations simulated from the model

$$y_i = \begin{bmatrix} 1 & u_{1,i} \\ 1 & u_{2,i} \end{bmatrix} \cdot \beta^G_{S_i} + \varepsilon_i, \qquad \varepsilon_i \sim N(0, 0.25 \cdot I),$$

where $u_{1,i} \sim U(0.6, 0.8)$, $u_{2,i} \sim U(0.8, 1)$, $S_i = 1$ with probability $\eta_1 = 0.45$, $\beta^G_1 = (0.25, -0.7)'$ and $\beta^G_2 = (0.3, 0.7)'$. Figure 1 shows a scatter plot of the data.

5.2 Selecting the number of classes

We used the bridge sampling estimator based on a random permutation sampler (see section 4) with the priors $b_0 = 0$ and $B_0^{-1} = 0.04 \cdot I$ on the group specific parameters and D(1, ..., 1) on the group probabilities to compute the model likelihoods for models with one to three classes (the number M_L of mixtures in the construction of the importance density is selected to be 500). The results reported in table 1 clearly reject a homogeneous model with just one group. The model likelihood points towards a model with two classes. For completeness, we also report the estimates obtained from reciprocal importance sampling and importance sampling which serve as a starting value for the bridge sampling estimator. The standard errors clearly indicate that the bridge sampling estimator leads to a more precise estimate than the other methods.

Estimator	$K = 1$	$K = 2$	$K = 3$
$\log \hat{L}_{BS}(y^N)$	-204.59	-178.90 (0.0064)	-186.02 (0.064)
$\log \hat{L}_{IS}(y^N)$	-204.59	-178.89 (0.011)	-181.55 (0.822)
$\log \hat{L}_{RI}(y^N)$	-204.59	-178.94 (0.037)	-185.38 (0.131)

Table 1: Various estimates of the model likelihood $L(y^N|K)$ for various values of K (relative standard errors are given in parenthesis); $\hat{L}_{BS}(y^N)\ldots$ bridge sampling estimator, $\hat{L}_{IS}(y^N)\ldots$ importance sampling estimator, $\hat{L}_{RI}(y^N)\ldots$ reciprocal importance sampling estimator

An exploratory cross-check for the number of classes selected by the formal Bayesian procedure is plotting the posterior densities of the variance σ_ε^2 of the error term ε_i for various numbers of K. Being a group independent parameter, σ_ε^2 may be estimated from the MCMC simulations without caring about identifiability and label switching. The posterior densities $\pi(\sigma_\varepsilon^2|y^N, K)$ estimated from the random permutation sampler show a significant decrease of σ_ε^2 only up to two classes (see figure 2). The addition of the third class does not help to reduce the unexplained variance σ_ε^2 indicating that the number of classes is too big. Figure 2 could be regarded as a kind of a Bayesian scree plot.

5.3 Explorative Bayesian Analysis of the Unconstrained Model

Explorative Bayesian analysis is based on the following special structure of the posterior density of the unconstrained model. Due to invariance of the posterior to relabelling the classes, the marginal posterior of all group specific parameters is identical: $\pi(\beta_1^G|y^N) = \pi(\beta_k^G|y^N)$ for all $k = 2, \ldots, K$. Therefore for a sampler with balanced label switching, the simulations for all groups contain the same information. This property could be used to check convergence of the random permutation sampler by comparing the marginal posterior densities estimated from the MCMC output for each component (see figure 3).

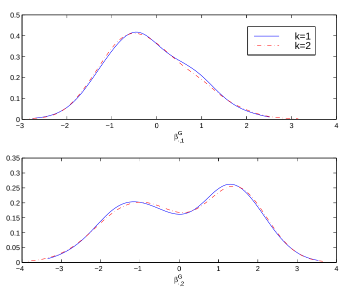

Figure 3: Checking convergence of the random permutation sampler

If convergence has occured, it is possible to consider the MCMC simulations of β_k^G for any k as coming from one density $\pi(\beta^G|y^N)$. As we are sampling from the unconstrained posterior, the simulations will switch between all possible labelling subspaces and might belong to any of the groups. By exploring the MCMC simulations we may learn a lot about differences between the groups. We found this kind of visualization of the unconstrained posterior extremely helpful with respect to selecting useful identifiability constraints.

Figure 4 and figure 5, for instance, show a scatter plot of the first component versus the second component of all simulated values $(\beta_{\cdot,1}^G, \beta_{\cdot,2}^G)^{(m)}$ for a two- and three class model. These figures contain further evidence in favour of a model with two groups. For the model with $K = 3$, where we allowed for three classes, the unconstrained marginal posterior distribution of $\pi(\beta^G|y^N)$ practically shows the correct number of classes, but is less informative due to the superfluent parameters introduced for the additional class. Furthermore, from the figures 4 and 5, we may learn a lot about differences between the classes. The first

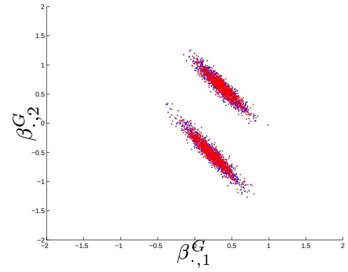

Figure 4: Posterior simulations obtained from the random permutation sampler for $K = 2$

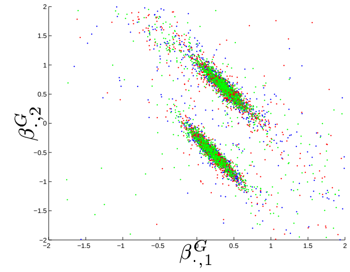

Figure 5: Posterior simulations obtained from the random permutation sampler for $K = 3$

component $\beta^G_{\cdot,1}$ does not differ between the groups, whereas the second component $\beta^G_{\cdot,2}$ clearly differentiates the two groups. This information will be helpful to identify the model.

5.4 Identifying the Selected Model

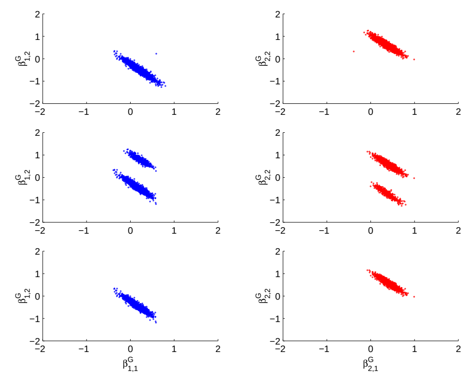

Figure 6: Influence of various identifiability constraints on the MCMC simulations simulations for β^G_1 (left column) and β^G_2 (right column); first row: $\beta^G_{1,2} < \beta^G_{2,2}$ (permutation sampling), second row: $\beta^G_{1,1} < \beta^G_{2,1}$ (permutation sampling), third row: $\beta^G_{1,1} < \beta^G_{2,1}$ (constrained Gibbs sampling)

If we want to estimate group specific parameters such as β^G_1, β^G_2, and η_1, we have to identify the model. Data-driven identifiability constraints, which respect the geometry of the posterior, are obtained from explorative Bayesian analysis of the MCMC output of the random permutation sampler. From figure 4 we found that the second components clearly differs between the groups, suggesting the constraint $\beta^G_{1,2} < \beta^G_{2,2}$. Permutation sampling under this constraint no longer showed any signs of label switching (see the first row of figure 6). Estimates of group specific parameters from the MCMC output of the identified model are reported in column three of table 2.

Table 2: Estimates of the group specific parameters under various constraints

Parameter	true	constrained estimation	
		$\beta^G_{1,2} < \beta^G_{2,2}$	$\beta^G_{1,1} < \beta^G_{2,1}$
$\beta^G_{1,1}$	0.25	0.221 (0.177)	0.181 (0.150)
$\beta^G_{1,2}$	–0.7	–0.519 (0.237)	–0.110 (0.579)
$\beta^G_{2,1}$	0.3	0.363 (0.160)	0.403 (0.140)
$\beta^G_{2,2}$	0.7	0.623 (0.201)	0.213 (0.601)
η_1	0.45	0.502 (0.059)	0.502 (0.059)

Note that the constraint is a device for defining a unique labelling subspace by describing geometrically the difference between the groups we made visible by allowing for balanced label switching in the sampled parameters. An alternative constraint would be the composite constraint $\beta^G_{1,1} + \beta^G_{1,2} < \beta^G_{2,1} + \beta^G_{2,2}$.

In order to demonstrate how an identifiability constraint which does not respect the geometry of the posterior density affects Bayesian analysis, we impose a constraint on the first component of the group specific parameter: $\beta^G_{1,1} < \beta^G_{2,1}$. The constraint $\beta^G_{1,1} < \beta^G_{2,1}$ ignores the geometry of this posterior insofar, as samples from *both* labelling subspaces will contain values for which this constraint is valid and *both* labelling subspaces will contain values for which this constraint is violated (see figure 4); the constraint does not separate the labelling subspaces. For such a constraint a well-mixing constrained sampler will *necessarily* exhibit label switching rendering estimation of group specific parameter from the MCMC output meaningless. If permutation sampling is combined with this "poor" identifiability constraint $\beta^G_{1,1} < \beta^G_{2,1}$, label switching is actually still present in the MCMC output (see the second row of figure 6). Like label switching occuring during MCMC estimation of the unconstrained model, this failure to identify the model must not be viewed as a weakness of the sampler, but as a property of the constrained posterior density itself. Label switching occuring during constrained sampling should always be viewed as a hint that the constraint selected for identification actually fails to separate the labelling subspaces. Label switching under a "poor" constraint will not occur for a constrained sampler which sticks at the current labelling subspace. This property, however, is not necessarily an advantage over a well-mixing sampler like permutation sampling. As such a constrained sampler will deliver a sample only from that region of the labelling subspace, where the constraint is fulfilled, we obtain an identified model with unique labelling where a bias toward the constraint might be introduced. The bias may affect the constrained components as well as all parameters correlated with the constrained components.

This actually happens for the present case study, if a truncation sampling method is included into the Gibbs sampler for the constraint $\beta^G_{1,1} < \beta^G_{2,1}$. The third row of figure 6 demonstrates that no label switching occured, whereas table 2 demonstrates that a bias is introduced for both components of the group specific parameter, when estimating these parameters from the output of the constrained sampler. Whereas permutation sampling indicates the failure to separate the labelling subspaces by this constraint (see the second row of figure 6), the bias introduced will go undetected if the sticky constrained sampler is employed instead.

6 Application to Metric Conjoint Analysis – A Case Study from the Austrian Mineral Water Market

6.1 Introduction

Previous research investigating the problem of efficiently modelling consumer heterogeneity in a conjoint setting went into two directions. One based on the assumption that heterogeneous consumer preferences are best described by an arbitrary discrete distribution with possibly very few mass points (e.g. DeSarbo *et al.*, 1992). The other building on the notion that the distribution of consumer preferences is unimodal and can be approximated by a parametric continuous distribution (e.g. Allenby and Ginter, 1995; Lenk *et al.*, 1996). Recently Allenby *et al.* (1998) as well as Lenk and DeSarbo (1999) indicated that the combination of the two approaches allowing for continuous heterogeneity within discrete latent classes even more adequately portrays preference heterogeneity.

In light of the available evidence that models assuming a continuous parametric distribution of consumer preferences outperform the latent class approach with respect to internal validity (Allenby and Ginter, 1995; Natter and Feuerstein, 1999; Moore *et al.*, 1998) it should be stressed that we do not argue in favour of the latent class approach. However, applied researchers continue to use the model and need to know about the opportunities and intricacies of its fully Bayesian analysis.

6.2 The data

The data come from a brand-price trade-off study in the mineral-water category conducted as part of an ongoing research project on brand equity (Schweiger, 1996; Strebinger *et al.*,1998). Each of 213 Austrian consumers stated their likelihood of purchasing 15 different product-profiles offering five brands of mineral water (Römerquelle, Vöslauer, Juvina, Waldquelle, and one brand not available in Austria, Kronsteiner) at 3 different prices (2.80, 4.80, and 6.80 [all prices in ATS]) on 20 point rating scales (higher values indicate

K	$\log \hat{L}_{BS}(y^N)$
2	-9892.88 (0.022)
3	-9756.02 (0.049)
4	-9701.54 (0.144)

K	$\log \hat{L}_{BS}(y^N)$
5	-9636.50 (0.294)
6	-9620.82 (0.393)
7	-9610.11 (0.396)

K	$\log \hat{L}_{BS}(y^N)$
8	-9589.53 (0.722)
9	-9582.66 (1.254)
10	-9582.89 (1.105)

Table 3: Estimates of the model likelihood $L(y^N|K)$ for various values of K (relative standard errors are given in parenthesis)

greater likelihood of purchasing). In an attempt to make the full brand by price factorial less obvious to consumers, the price levels varied in the range of ± 0.1 ATS around the respective design levels such that mean prices of brands in the design were not affected (Elrod *et al.*, 1992).

The data come from a brand-price trade-off study in the mineral-water category conducted as part of an ongoing research project on brand equity (Schweiger, 1996; Strebinger *et al.*,1998).

We used a fully parameterized matrix W_i with 15 columns corresponding to the constant, four brand contrasts, a linear and a quadratic price effect, four brand by linear price and four brand by quadratic price interaction effects, respectively. We used dummy-coding for the brands. The unknown brand Kronsteiner was chosen as baseline. We subtracted the smallest price from the linear price column in matrix W_i, and computed the quadratic price contrast from the centred linear contrast. Therefore the constant corresponds to the purchase likelihood of Kronsteiner at the lowest price level, if quadratic price effects are not present. Theory did not suggest excluding any effect for all consumers.

At the level of an individual consumer the model would be saturated since only 15 data points are available to estimate 15 parameters leaving zero degrees of freedom. In Frühwirth-Schnatter and Otter (1999) a random-effects model was fitted to the data. Here, we discuss modelling of heterogeneity by a latent class model with a priori unknown numbers K of groups.

6.3 Selecting the number of classes

Selection of the number of classes is based on the same tools as discussed in section 5.2 for the simulated data. First, we used the bridge sampling estimator based on a random permutation sampler with proper priors to compute the model likelihoods for models from two to ten classes (the number M_L of mixture in the construction of the importance density is selected to be 200). The prior mean b_0 of the group specific parameters is taken to be equal to the mean of the normal prior of the random effects estimated within the random-effects model reported in Frühwirth-Schnatter and Otter (1999). If such results were not available, we could have used the parameters estimated from a model with one class as a prior mean. The information matrix of the prior is equal to: $B_0^{-1} = 0.02 \cdot I$. The prior on η is $D(1, \dots, 1)$. The prior on σ_ε^2 is $IG(0,0)$.

The estimated model likelihoods together with relative standard errors are reported in table 3. The model likelihood clearly points towards a model with nine classes. This result is supported by the Bayesian scree plot in figure 7: the posterior densities of σ_ε^2 estimated from the random permutation sampler showed a significant reduction of σ_ε^2 only up to nine classes.

6.4 Estimation within Unidentified Models

In section 2.2 it was demonstrated that subject specific parameter estimates may be obtained from an unidentified model. Here we illustrate how the latent class model captures consumer heterogeneity for different numbers of classes. We take advantage of our Bayesian approach and investigate the posterior densities of implied choice probabilities for different offers. Throughout this section it is assumed that choice probabilities may be derived from preferences directly using a multinomial logit model. The following is based on a choice set offering the two major brands Römerquelle and Vöslauer at a price of 5.9, their competitors Juvina and Waldquelle at a price of 3.9 and finally the dummy brand Kronsteiner at a the price of 3.2 (all prices in ATS).

The columns of Figure 8 contain the marginal choice probabilities for the Römerquelle, the Juvina and the Kronsteiner offer, respectively. The rows correspond to different numbers of classes. Comparing the

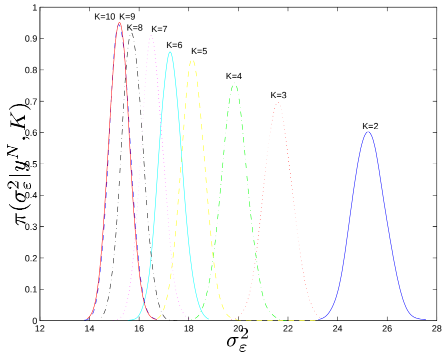

Figure 7: Marginal densities $\pi(\sigma_\varepsilon^2|y^N, K)$ estimated from a random permutation sampler

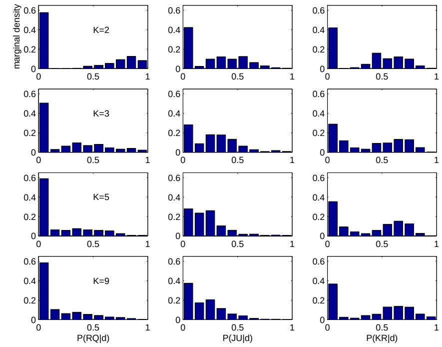

Figure 8: Choice probabilities from a logit model (design: (RO, VO, 5.9), (JU,WA, 3.9), (KR, 3.2))

distributions of choice probabilities for the Römerquelle offer in the first column, the detrimental effect of choosing too low a number of classes becomes obvious. The marginal density obtained from the model with two classes suggests that the major part of the sample has a near zero choice probability for the Römerquelle offer. On the other hand substantial mass of the distribution is in the region with a choice probability greater than 0.6. The optimal choice of nine classes, in contrast, reveals that only little mass of the distribution can be found in the region of high choice probability for the Römerquelle offer. Similar but less extreme differences between the different solutions arise for the Juvina and the Kronsteiner offer.

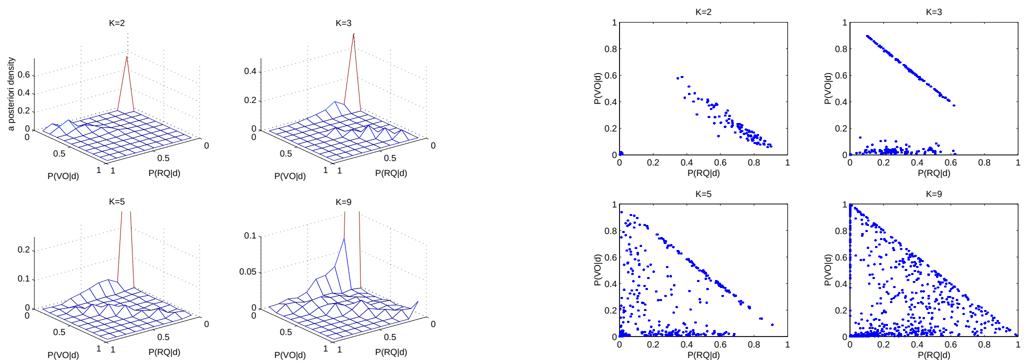

Figure 9: Choice probabilities from a logit model - bivariate representations

Figure 9 contains bivariate marginal densities for the choice probabilities of the Römerquelle and the Vöslauer offer. Notice the 45 degree frontier due to the fact that we are dealing with choice probabilities from a multinomial logit model. Moreover, the scaling of the density axis is adapted to the number of classes assumed in the three dimensional plots. Again it is obvious, possibly even more than from the univariate density plots that too low a number of classes will lead to very different conclusions than the optimal choice. The solution with two classes suggests that most of the mass is concentrated at the point of near zero choice probabilities for the Römerquelle and the Vöslauer offer. Moreover, there seems to be some mass in the area of high choice probabilities for the Römerquelle offer accompanied by low choice probabilities of the Vöslauer offer. Increasing the number of classes to three changes the picture dramatically. Again most of the distributional mass is concentrated at the zero/zero point. However, the distribution indicates some support for the combination of high choice probability for the Vöslauer offer accompanied by lower choice probabilities for the Römerquelle offer. Moreover, there also is some support for the combination of small choice probabilities for the Römerquelle offer and a near zero probability for the Vöslauer offer. Naturally, the optimal solution with nine classes offers the most detailed picture. Notice that now some support for the combination of choice probabilities near one for the Vöslauer offer accompanied by such near zero for the Römerquelle offer can be found. This feature of the distribution is not present in the solution relying on five classes only.

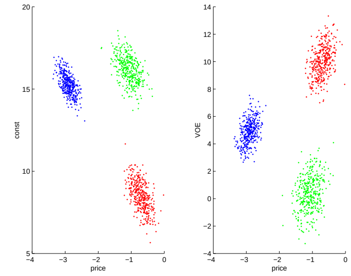

Figure 10: Scatter plots of the MCMC simulations for $K = 3$

6.5 Model Identification

So far the models are identified only up to permutations of the labels of the groups. In order to identify group specific parameters, we need to introduce identifiability constraints that guarantee unique labelling. In what follows we use the following notation for the group specific parameters: const_k refers to the constant, RQ_k, VO_k, JU_k, and WA_k refer to the main effects for the various brands and p_k to the linear price effect for group k. We carefully searched for data-driven identification constraints using plots of the marginal parameter densities as well as two- and three-dimensional scatterplots of MCMC simulations from the unidentified model. As one might expect, simple constraints will only suffice to identify models with few classes. For instance in the case of a model with three classes, we found from the scatter plots in figure 10 that the linear price effect differentiates one class from the remaining two and that the constant is useful to tell these apart: $p_1 < \min(p_2, p_3)$, $\text{const}_2 < \text{const}_3$. Sensible identifiability constraints are not necessarily unique. For the model with three classes, an equivalent set of constraints turned out to be $p_1 < \min(p_2, p_3)$, $\text{VO}_2 < \text{VO}_3$.

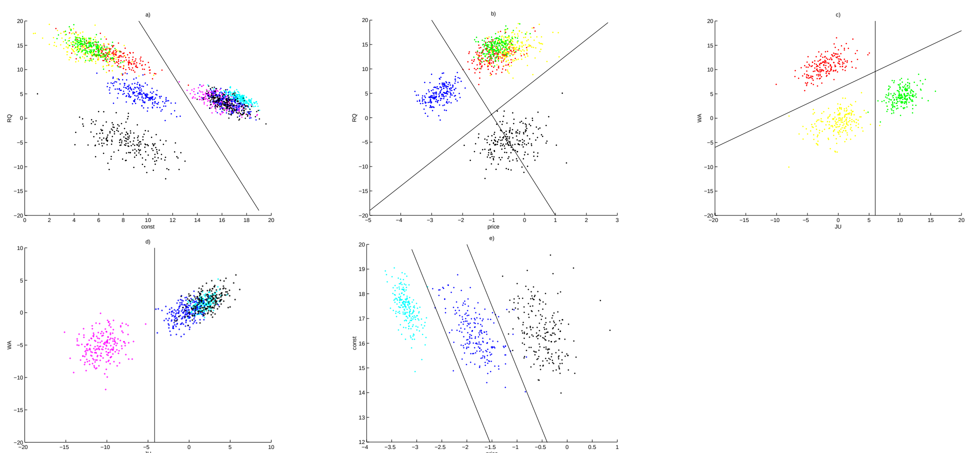

Figure 11: Scatter plots of the MCMC simulations for $K = 9$

Estimation of the model likelihoods pointed to a model with nine classes (see table 3). Due to the relatively high number of classes we proceed in a stepwise manner to identify this model. The constraint $\max_{k=1,\dots,5}(4\cdot\text{const}_k + \text{RQ}_k) < \min_{k=6,\dots,9}(4\cdot\text{const}_k + \text{RQ}_k)$ divides the nine classes into two subgroups of five and four classes, respectively (see figure 11a). These two subgroups are treated separately now, the one with the five classes first. In figure 11b we see that the constraint $\max_{k=1,\dots,4}(10 \cdot p_k - 2 \cdot \text{RQ}_k) < 10 \cdot p_5 - 2 \cdot \text{RQ}_5$ splits off the first group and $10 \cdot p_1 + \text{RQ}_1 < \min_{k=2,3,4}(10 \cdot p_k + \text{RQ}_k)$ splits off the second group. In figure 11c the third group is separated by $3 \cdot \text{JU}_3 - 5 \cdot \text{WA}_3 < \min_{k=4,5}(3 \cdot \text{JU}_k - 5 \cdot \text{WA}_k)$ and finally, the fourth and the fifth group are identified by $\text{JU}_4 < \text{JU}_5$. To split off one group from the subgroup with the four classes we use the constraint $\text{JU}_6 < \min_{k=7,8,9} \text{JU}_k$ (see figure 11d). The restriction $5 \cdot \text{price}_7 + \text{const}_7 < 5 \cdot \text{price}_8 + \text{const}_8 < 5 \cdot \text{price}_9 + \text{const}_9$ separates the three classes that are still left. These constraints were imposed to obtain the group specific estimates reported in table 4. They will be interpreted from a marketing point of view in subsection 6.7.

Next we discuss alternative identifiability constraints. Lenk and DeSarbo (1999) and Allenby *et al.* (1998)

$E(\beta_k^G\|y^N)$ (Standard deviation in parenthesis)									
Effect	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9
const	9.32	7.31	4.91	5.34	8.41	15.8	17.4	16.4	16.4
	(1.2)	(1.7)	(1.4)	(1.1)	(1.8)	(1.1)	(0.64)	(0.89)	(0.89)
RQ	5.0	12.6	14.3	14.6	–4.92	3.53	4.15	3.28	3.10
	(1.8)	(2)	(2.2)	(1.5)	(2.7)	(1.6)	(0.87)	(1.2)	(1.2)
VO	5.61	11.8	15.0	13.5	6.15	3.12	4.30	0.68	–0.38
	(1.7)	(2.1)	(2.2)	(1.5)	(2.7)	(1.6)	(0.85)	(1.3)	(1.3)
JU	1.55	–1.94	0.10	10.4	3.25	–10.5	1.86	–0.27	2.30
	(1.8)	(2.4)	(2.3)	(1.6)	(2.6)	(1.6)	(0.89)	(1.4)	(1.4)
WA	2.44	10.8	-0.92	4.62	4.23	–5.08	1.42	0.016	1.80
	(1.7)	(2.0)	(2.2)	(1.7)	(2.8)	(2.0)	(0.88)	(1.3)	(1.3)
p	–2.70	–0.97	–0.35	–0.91	–0.55	–0.93	–3.19	–1.79	–0.51
	(0.35)	(0.45)	(0.45)	(0.30)	(0.56)	(0.35)	(0.16)	(0.32)	(0.32)
p^2	0.75	0.048	–0.22	0.092	–0.089	–0.062	–0.14	0.11	–0.17
	(0.27)	(0.28)	(0.36)	(0.25)	(0.44)	(0.25)	(0.13)	(0.20)	(0.20)
RQ$\cdot p$	–1.25	–0.75	–2.47	–0.41	0.10	0.13	–0.54	0.014	0.12
	(0.49)	(0.57)	(0.67)	(0.46)	(0.81)	(0.46)	(0.24)	(0.35)	(0.35)
VO$\cdot p$	–1.53	–1.05	–2.67	–0.20	0.69	0.24	–0.59	–0.009	0.51
	(0.48)	(0.6)	(0.69)	(0.47)	(0.81)	(0.47)	(0.24)	(0.36)	(0.36)
JU$\cdot p$	–0.63	0.14	–0.59	–0.027	1.10	0.36	–0.53	–0.021	0.084
	(0.48)	(0.55)	(0.66)	(0.45)	(0.82)	(0.46)	(0.24)	(0.36)	(0.36)
WA$\cdot p$	–0.72	–1.30	–0.11	–0.51	–0.56	–0.066	–0.50	0.061	0.23
	(0.48)	(0.53)	(0.62)	(0.47)	(0.81)	(0.44)	(0.24)	(0.35)	(0.35)
RQ$\cdot p^2$	0.25	–0.24	0.048	–0.43	–0.022	–0.11	–0.39	–0.25	–0.11
	(0.40)	(0.43)	(0.54)	(0.36)	(0.68)	(0.36)	(0.19)	(0.28)	(0.28)
VO$\cdot p^2$	0.37	–0.062	0.26	–0.032	0.49	–0.02	–0.47	–0.24	0.05
	(0.41)	(0.42)	(0.59)	(0.37)	(0.67)	(0.38)	(0.19)	(0.30)	(0.30)
JU$\cdot p^2$	0.37	0.094	0.39	–0.31	–0.20	–0.038	–0.28	–0.056	–0.034
	(0.40)	(0.41)	(0.53)	(0.37)	(0.68)	(0.39)	(0.19)	(0.30)	(0.30)
WA$\cdot p^2$	0.20	–0.15	0.41	0.22	–0.26	0.044	–0.19	–0.091	0.0068
	(0.38)	(0.42)	(0.51)	(0.37)	(0.68)	(0.37)	(0.19)	(0.28)	(0.28)

$E(\eta_k\|y^N)$ (Standard deviations in parenthesis)									
	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9
	0.086	0.078	0.046	0.091	0.027	0.089	0.33	0.15	0.10
	(0.022)	(0.022)	(0.015)	(0.021)	(0.011)	(0.024)	(0.037)	(0.030)	(0.026)

Table 4: Estimates of group specific parameters for $K = 9$ (RQ, VO, etc are the various brands, p and p^2 denote the linear and the quadratic price effect; RQ $\cdot p$, etc are interaction effects)

apply the standard constraint $\eta_1 < \ldots < \eta_K$ within their MCMC procedure to identify group specific parameters. The marginal posterior densities of η_k estimated from the MCMC output of our identified model show that there are three groups of practically the same size (see the upper part of figure 12). Thus the constraint is not supported by the data. We will now discuss the consequence of including this constraint in comparison to the model identified above. Constrained Gibbs sampling produces a substantially different posterior for the weights $\eta_1, \ldots, \eta_9$ compared to the model identified above (compare the lower and the upper part of figure 12), with an obvious bias toward the constraint for those groups which are of about the same size. Interestingly, there is little to no effect on estimation of the group specific parameters (estimation results not reported here). This may be explained by the rather low correlations between the various weights η_k and the group specific parameters β_k^G which range between –0.28 for η_2 and p_2 and 0.31 for WA$_6$ and η_6. We did find differences in the classification probabilities for some consumers.

Overall the different solutions obtained were practically identical in terms of predicting holdout data. The holdout design consisted of all five brands at prices meant to create a pareto optimal set after integrating over consumers. The two major brands Römerquelle and Vöslauer were offered at 5.9, Juvina and Waldquelle were offered at 3.9 and finally the dummy brand Kronsteiner at 3.2 (all prices in ATS). Again consumers rated their purchase likelihood on the same 20 point scale. The mean squared errors for the

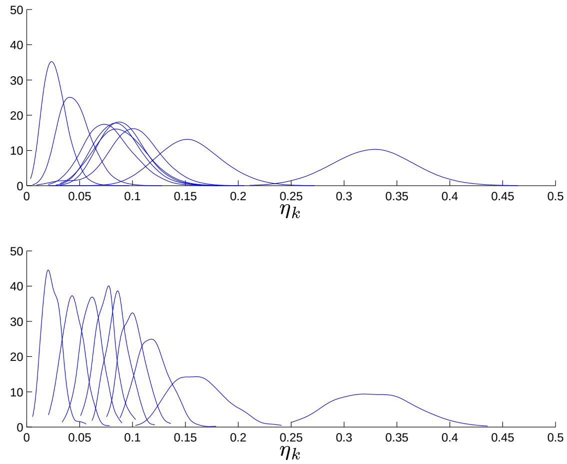

Figure 12: Posterior densities of the weights $\eta_1, \ldots, \eta_9$ (upper row: identified model, lower row: order constraint on the weights)

unidentified model and the model with order constraints on η_k were 25.48 and 25.25, respectively. When every consumer is assigned a parameter vector according to the class with the highest a posteriori classification probability mean squared errors for the models with data driven constraints and with order contraints on η_k were 26.99 and 26.79, respectively. Finally, the first choice hit rates for the unidentified and the η_k constrained model did not differ and amounted to 51.64%.

6.6 Testing for Heterogeneity and Variable Selection

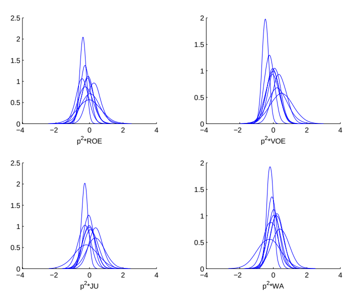

Figure 13: Marginal denities of all quadratic interaction effects

From all results reported above heterogeneity in consumer preferences is present without doubt. It is, however, not clear *apriori* that heterogeneity affects all components of the group specific parameters. As the marginal densities the quadratic interaction effects plotted in figure 13 overlap for all groups to a high degree, we formulate the hypothesis that the quadratic interaction effects RQ$\cdot p^2$, VO$\cdot p^2$, JU$\cdot p^2$, and WA$\cdot p^2$ are fixed rather than random. As the marginal density of the quadratic interaction effects in figure 13 not only overlap for all groups, but most of them also cover 0, we further hypothesize that the quadratic interaction affects are not significant and should be deleted from the model: RQ$\cdot p^2$ = VO$\cdot p^2$= JU$\cdot p^2$ = WA$\cdot p^2 = 0$.

All hypotheses are tested against the full model by comparing the model likelihoods. The log of the model likelihood for a model with fixed quadratic interaction effects is equal to –9309.4 with an standard error of 0.714, whereas the log of the model likelihood for a model without quadratic interaction effects is equal to -9399.9 with an standard error of 1.182. In comparison to the full model (see table 3) excluding the quadratic interaction effects decreases the model likelihood, whereas assuming fixed rather than random quadratic interaction effects increases the model likelihood substantially. It turned out that the same identifiability constraints that were formulated for the full model, where all components were heterogeneous, applied to this mixed effects latent model.

6.7 Interpretation of the nine classes model from a marketing point of view

An obvious starting point for the interpretation of the classes would be the class specific parameter estimates reported in table 4. However, due to the dimensionality of the designmatrix and the presence of quadratic price and interaction effects it is not an easy task to derive a coherent interpretation. Therefore, we settled for the following procedure: Given the class specific parameters we formulated three designs. The three designs offered all five brands at a low price (ATS 2.7), a medium price (ATS 4.8) and a high price (ATS 6.9), respectively. Then we computed the purchase likelihood ratings to be expected in the 9 classes for all three designs. Figure 14 illustrates the result.

Class 7 with the highest a posteriori size of all classes (approximately 33%) and class 1 are very price sensitive with only minor brand differentiation. Whereas class 7 would still accept a medium price class 1 shows a strong tendency to avoid all offers but the cheapest. Interestingly there is some differentiation between brands offered at the lowest price in class 1. The dummy brand Kronsteiner is evaluated less favourably. Classes 2, 4 and 8 are moderately price sensitive. Again, there is little brand differentiation in class 8 with a slight advantage of Römerquelle over its competitors. Class 2 clearly dislikes the Juvina brand and prefers Römerquelle, Vöslauer and Waldquelle to the dummy brand Kronsteiner. A price increase seems to affect Römerquelle to a lesser extent than Vöslauer and Waldquelle in this class. Class 4 prefers Römerquelle and Vöslauer to the other brands and clearly disapproves of the Kronsteiner brand. Class 3 again favours Römerquelle and Vöslauer over the other brands. Interestingly the advantage of Römerquelle and Vöslauer diminishes substantively at higher price levels. Classes 5, 6 and 9 reveal only little sensitivity to price. In the case of the very small class 5 the Juvina brand even is evaluated more favourably at the higher price levels. Moreover this class is the only one to clearly reject the Römerquelle brand. Finally, class 6 again favours Römerquelle and Vöslauer and clearly disapproves of Juvina. Also, the dummy brand Kronsteiner is preferred to the established brand Waldquelle in this class.

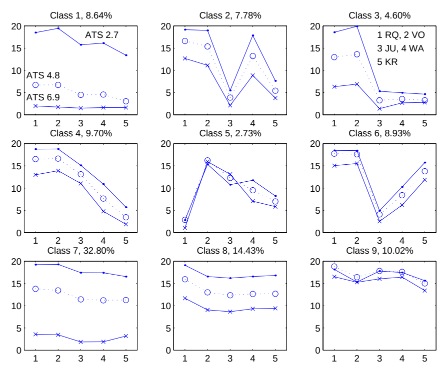

Figure 14: Class specific purchase likelihoods for three designs: all brands 2.7, 4.8 and 6.9 ATS

Overall, a major portion of consumers seems to be very price sensitive with only little brand differentiation. Despite intensive marketing activity especially by Römerquelle and Vöslauer these consumers behave like in a commodity market. Even an up to the interview unknown dummy brand would be readily accepted. Römerquelle and Vöslauer seem to be generally accepted (with the exception of Römerquelle in one very small class). Juvina and Waldquelle did not succeed in establishing classes that favoured their brand over the competitors.

7 Discussion

We discussed a fully Bayesian analysis of the latent class model using a new approach towards MCMC estimation in the context of mixture models. This approach starts with estimating unidentified models for various numbers of classes. We compute exact Bayes' factors by the bridge sampling estimator and developed a Bayesian scree plot to compare different models and select the number of classes. Estimation of the unidentified model is carried out using the random permutation sampler. From the unidentified model we estimate model parameters that are not class specific such as fixed effects, the mean and the covariance of the random effects, consumer specific effects (with the obvious exception of class membership), and the model likelihood. We explored the MCMC output from the unconstrained model in order to find identifiability constraints. The constrained version of the permutation sampler was used to identify group specific parameters. We discussed the problems associated with identifiability constraints which ignore the geometry of the posterior distribution.

The suggested approach was applied to simulated data and to data from a brand-price trade off conjoint study. Within the conjoint study we did not compare the latent class solution to the parametric random coefficients model in this paper. However, we would like to add that the random coefficients model outperformed even the latent class solution with nine classes in terms of the model likelihood. Nevertheless, the latent class approach could be used in a similar way as nonparametric maximum likelihood estimation to parsimoniously account for multimodal preference distributions. The characteristics of such a distribution can be sensibly described *a posteriori* without the need to identify a unique labelling. Furthermore, this paper should be useful in the context of Bayesian estimation of general finite mixtures that in a way combine the latent class approach and parametric models for consumer heterogeneity (Allenby *et al.*, 1998; Lenk and DeSarbo, 1999).

Acknowledgments

Special thanks go to Josef Mazanec, Vice-Rector of the University of Business Administration and Economics, to Günter Schweiger, Chair of the Department of Advertising and Marketing Research, and to Römerquelle Ges.m.b.H. The work of the second author was partly supported by the Austrian Science Foundation (FWF), Project No 12025; the work of the third author was supported by the Kuratorium der Wirtschaftsuniversität Wien.

References

Allenby, G. M. and Ginter, J. L. (1995): Using Extremes to Design Products and Segment Markets. *Journal of Marketing Research*, 32, 392-403.

Allenby, G. M., Arora, N. and Ginter, J. L. (1998): On the Heterogeneity of Demand. *Journal of Marketing Research*, 35, 384-89.

Bernardo, J.M., and Smith, A.F.M. (1994), *Bayesian Theory*. New York: Wiley.

Celeux, G. (1998), Bayesian Inference for Mixture: The Label Switching Problem. Preprint Inria.

Chib, S. (1995), Marginal Likelihoods from the Gibbs Output. *Journal of the American Statistical Assosiation*, 90, 1313-21.

DeSarbo, W. S., Wedel, M., Vriens, M., and Ramaswamy, V. (1992): Latent Class Metric Conjoint Analysis. *Marketing Letters*, 3, 273-88.

DiCiccio, T.J., Kass, R., Raftery, A. and L. Wasserman (1997): Computing Bayes Factors By Combining Simulations and Asymptotic Approximations. *Journal of the American Statistical Assosiation*, **92**, 903-15.

Diebolt, J. and Robert, C.P. (1994), Estimation of finite mixture distributions through Bayesian sampling, *Journal of the Royal Statististical Society*, Ser. B, 56, 363-75.

Elrod, T., and Häubl, G. (1997), An Extended Random Coefficients Model, With Application to Metric Conjoint Analysis. Working paper, University of Alberta, Alberta.

Frühwirth-Schnatter, S. (1995), Bayesian Model Discrimination and Bayes Factors for Linear Gaussian State Space Models, *Journal of the Royal Statistical Society*, B 57, 237-246.

Frühwirth-Schnatter, S. (1999a): MCMC Estimation of Classical and Dynamic Switching and Mixture Models. Second revision submitted to Journal of the American Statistical Assosiation.

Frühwirth-Schnatter, S. (1999b): Model Likelihoods and Bayes Factors for Switching and Mixture Models. Preprint Vienna University of Economics and Business Administration.

Frühwirth-Schnatter, S. and Otter, Th. (1999): Conjoint-Analysis Using Mixed Effect Models. In: Friedl, H., Berghold, A. and Kauermann, G. (Eds.): *Statistical Modelling. Proceedings of the Fourteenth International Workshop on Statistical Modelling*. Graz 1999, pp.181–191.

Gelfand, A.E., and D.K. Dey (1994): Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society*, Ser. B, 501-514.

Hagerty, M. R. (1985): Improving the Predictive Power of Conjoint Analysis: The Use of Factor Analysis and Cluster Analysis. *Journal of Marketing Research*, 22, 168-184.

Kamakura, W. A. (1988): A Least Squares Procedure for Benefit Segmentation with Conjoint Experiments. *Journal of Marketing Research*, 25, 157-67.

Kamakura, W. A., Wedel, M., and Agrawal, J. (1994): Concomitant variable latent class models for conjoint analysis. *International Journal of Research in Marketing*, 11, 451-464.

Lenk, P. J., DeSarbo, W. S., Green, P. E. and Young, M. R. (1996): Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs. *Marketing Science*, 15, 173-91.

Lenk, P. J. and DeSarbo, W. S. (1999): Bayesian Inference for Finite Mixture of Generalized Linear Models with Random Effects. To appear in *Psychometrika*.

McCulloch, R.E. and Tsay, R.S. (1994), Statistical analysis of economic time series via Markov switching models, *Journal of Time Series Analysis*, 15, 523-39.

Meng, X.L. and W.H. Wong (1996): Simulating Ratios of Normalising Constants via a Simple Identidy. *Statistica Sinica*, 6, 831-60.

Moore, W., Gray-Lee, J., and Louviere, J. J. (1998): A Cross Validity Comparison of Conjoint Analysis and Choice Models at Different Levels of Aggregation. *Marketing Letters*, 9(2), 195-207.

Natter, M., and Feuerstein, M. (1999): Individual Level or Segmentation Based Market Simulation? Working Paper Vienna University of Economics and Business Administration, Vienna.

Richardson, S. and Green, P. (1997), On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society*, Ser. B, 59, 731-92.

Roeder, K. and Wasserman, L. (1997), Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association*, 92, 894-902.

Schweiger, G. (1996), Brand Equity - A Consumer Information Processing Approach, Grant No. P-12025-OEK, Austrian Science Foundation (FWF).

Smith, A.F.M. and Roberts, G. (1993), Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods, *Journal of the Royal Statistical Society*, Ser. B, 55, 3-23.

Stephens, M. (1997), Bayesian Methods for mixtures of normal distributions. Ph.D.Thesis, University of Oxford.

Strebinger, A., Schweiger,G., and Th. Otter (1998), Brand Equity and Consumer Information Processing: A Proposed Model, paper presented at AMA's Marketing Exchange Colloquium, Vienna.

Tanner, M.A. (1993), *Tools for Statistical Inference: Methods of the Exploration of Posterior Distributions and Likelihood Functions*. Second edition, New York: Springer-Verlag.

Wedel, M., and Steenkamp, J.-B. (1991): A Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Benefit Segmentation. *Journal of Marketing Research*, 28, 385-96.