ECONSTOR

Aßmann, Christian; Boysen-Hogrefe, Jens

Working Paper

A bayesian approach to model-based clustering for panel probit models

a bayesian approach to model-based clustering for panel probit models

by Chrisitan Aßmann and Jens Boysen-Hogrefe

A Bayesian approach to Model-Based Clustering for Panel Probit Models

Christian Aßmann[*]

Institute of Statistics and Econometrics, Christian-Albrechts-Universität Kiel, Germany

Jens Boysen-Hogrefe

Kiel Institute for the World Economy, Germany

April 15, 2009

Abstract

Consideration of latent heterogeneity is of special importance in non linear models for gauging correctly the effect of explaining variables on the dependent variable. This paper adopts the stratified model-based clustering approach for modeling latent heterogeneity for panel probit models. Within a Bayesian framework an estimation algorithm dealing with the inherent label switching problem is provided. Determination of the number of clusters is based on the marginal likelihood and out-of-sample criteria. The ability to decide on the correct number of clusters is assessed within a simulation study indicating high accuracy for both approaches. Different concepts of marginal effects incorporating latent heterogeneity at different degrees arise within the considered model setup and are directly at hand within Bayesian estimation via MCMC methodology. An empirical illustration of the developed methodology indicates that consideration of latent heterogeneity via latent clusters provides the preferred model specification compared to a pooled and a random coefficient specification.

JEL classification: C11; C23; C25

Keywords: Bayesian Estimation; MCMC Methods; Panel Probit Model; Mixture Modelling

---

[*]Corresponding author. Tel.: +49-431-8803399; fax: +49-431-8807605. *E-mail address:* christian.assmann@stat-econ.uni-kiel.de (C. Aßmann)

1   Introduction

The consideration of latent heterogeneity in panel models with relative short time horizons is of special importance in order to gauge correctly the influence of variables suggested by theory on the dependent variable. The often adapted strategy for linear models of pooling and computation of heteroscedastic or cluster robust variance estimates can not be applied to panel probit models for several reasons. In the presence of latent heterogeneity the pooled estimator is not guaranteed to be consistent, see Cameron and Trivedi (2005), and the computation of cluster robust variance estimates requires a sufficiently large number of observations per cross section. Furthermore, the natural approach of using fixed effects is often problematic since it causes the occurrence of an incidental parameter problem, see Lancaster (2000) for a general review and Greene (2004a) for a discussion focused on limited dependent variable models. Other common approaches to model latent heterogeneity are random coefficient specifications, which employ distributional assumptions on the heterogeneity and assume the orthogonality of latent heterogeneity and explaining variables, see Revelt and Train (1998), Mehndiratta (1996) and Ben-Akiva and Bolduc (1996) for applications.[1] Alternatively, Frühwirth-Schnatter and Kaufmann (2008) suggest a stratified model-based clustering approach to deal with latent heterogeneity in the data generating process.[2]

Within this paper, the strategy to model latent heterogeneity via clustering, where the identification of clusters and affiliation of panel members to clusters is simultaneously undergone to parameter estimation, is provided for panel probit models. Latent clusters for panel probit models have been considered by Greene (2004b) and Greene and Hensher (2003) in the context of maximum likelihood estimation. However, maximum likelihood estimation possibly fails to provide an accurate assessment of parameter uncertainty arising from genuine multimodality of the likelihood as discussed by Celeux (1998).

Hence, within this paper a Bayesian estimation approach is pursued, which is able to gauge parameter uncertainty correctly in the presence of genuine multimodality via inspection of the posterior distribution. These advantages of the Bayesian approach are accompanied by the difficulty of label switching stemming from the invariance of the likelihood to relabeling of the clusters.[3] As noted by Stephens (2000) dealing with label switching in Bayesian estimation is often done via incorporation of an identifiability constraint ensuring formal separation of the symmetric modes of the parameter space. Unfortunately, an identifiability restriction will possibly not suppress label switching suffi-

---

[1] Note that latent heterogeneity can also be addressed in non-parametric estimation environments, which are not subject of this paper.

[2] See Fraley and Raftery (2002) for a review of model based clustering in non-time series data.

[3] In a maximum likelihood approach this problem is avoided as a single maximum (out of all symmetric ones) is regarded, thereby providing identification en passant.

ciently, since separation of the parameter space may only be weak causing poor parameter estimates. This paper employs a relabeling algorithm based on Stephens (2000) to deal with label switching, which is extended to deal with the considered model feature of stratified cluster probabilities.[4]

A further advantage of Bayesian estimation via MCMC methodology is the direct accessibility of a wide range of estimates of marginal effects incorporating latent heterogeneity at different degrees. Since the MCMC methodology provides draws from the posterior distributions of parameters, estimates of marginal effects as moments thereof are easily calculated. In contrast to maximum likelihood based estimation also the distributions of the marginal effects are directly accessible allowing to gauge the robustness of theoretical implications under latent heterogeneity.

Next to estimation conditionally on the number of clusters, the problem of determining the number of clusters is addressed within this paper. Several strategies to determine the number of cluster components are analyzed. Based on the MCMC output we use the marginal likelihood as a natural benchmark for assessing model fit. However, computation of the marginal likelihood asks for special efforts to ensure the accuracy of the involved numerical integration techniques. Therefore, we contribute an easy implementable alternative approach to assess model fit via a cross validation experiment.[5] Thereby, our approach extends the range of available cluster strategies towards a Bayesian analysis of multivariate panel probit mixture models. We provide a simulation study highlighting the properties of these model selection devices.

Our results suggest via the conducted simulation study the ability of the marginal likelihood and the cross validation approach to select the correct number of clusters. In the empirical application using the data set of Bertschek and Lechner (1998) on firm innovation, we find that a model with 3 clusters is preferred and that it provides a better fit compared to a pooled or random coefficient specification. Implications for gauging theoretical issues are discussed in the context of marginal effects incorporation latent heterogeneity at different degrees.

This paper is hence organized as follows. Section 2 reviews the panel probit framework with random coefficients and stratified model based clustering. In Section 3 the Bayesian estimation methodology, the relabeling algorithm, the approach to calculate the marginal likelihood, and the considered cross validation design for determining the number of clusters are provided. The methods to identify the number of clusters are assessed within a simulation study in Section 4. An empirical

---

[4]See Handcock et al. (2007) for an application of the relabeling algorithm in the absence of meaningful identifiability constraints.

[5]Procedures for determining the number of clusters in finite mixture models are discussed in related contexts. Chen and Khalili (2008) discuss a penalized likelihood approach for univariate mixtures. Dunson et al. (2008) analyze a semi-parametric approach for univariate mixture models without stratified cluster probabilities. Ihswaran et al. (2001) employ approximate Bayes factors. Heard et al. (2006) contribute a hierarchical analysis for Bayesian curve fitting, while Ray and Lindsay (2008) analyze determination with a non-parametric quadratic risk approach.

illustration is given in Section 5, where the developed methodology is applied on the data set of Bertschek and Lechner (1998). Section 6 concludes.

2 Model Formulation

The setup of a pooled panel probit model is given as follows. Let y_{it} denote the observed dichotomous variable, where $i = 1, \ldots, N$, and $t = 1, \ldots, T$. A link between observed explaining factors and the observed binary variables is provided via the latent variable y_{it}^*

$$y_{it} = \begin{cases} 1, & \text{if } y_{it}^* > 0, \\ 0, & \text{if } y_{it}^* \leq 0, \end{cases} \tag{1}$$

where

$$y_{it}^* = X_{it}\beta + e_{it} \tag{2}$$

and e_{it} is an independent identically normal distributed error term with unit variance. Pooling yields the likelihood

$$\mathcal{L}_P(Y|\theta, X) = \prod_{i=1}^{N} \prod_{t=1}^{T} \Phi\left[(2y_{it} - 1)X_{it}\beta\right], \tag{3}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal distribution.

An often considered approach to account for latent heterogeneity is to model random coefficients, see Train (2003). Hence β is modeled as unit specific random variable

$$\beta_i \overset{\text{iid}}{\sim} \mathcal{N}(b, W), \tag{4}$$

where W denotes the covariance matrix of the random coefficients and b denotes the common mean vector for all individuals. This results in a likelihood function given as

$$\mathcal{L}_{RC}(Y|\theta, X) = \prod_{i=1}^{N} \int_{R^p} \left[\prod_{t=1}^{T} \Phi\left[(2y_{it} - 1)X_{it}\beta_i\right]\right] f(\beta_i|b, W)d\beta_i, \tag{5}$$

where p denotes the dimension of the vector of random coefficients. We will refer to this kind of modeling heterogeneity as a benchmark for modeling latent heterogeneity within the empirical illustration.

Alternatively, latent heterogeneity can be incorporated via latent clusters. Model-based clustering assumes a given number of cluster, where members of a cluster share the same parameters. Define $S = \{S_i\}_{i=1}^{N}$, $k = 1, \ldots, K$ indicating the cluster membership for each individual. Conditional on the cluster membership the latent model is given as

$$y_{it}^* = X_{it}\beta_{S_i} + e_{it}, \tag{6}$$

providing the joint density function

$$f(Y|\theta, X, S) = \prod_{i=1}^{N} \prod_{t=1}^{T} \Phi\left[(2y_{it} - 1)X_{it}\beta_{S_i}\right]. \tag{7}$$

Frühwirth-Schnatter and Kaufmann (2008) propose two distinct ways in modeling the probabilistic structure of cluster membership. An a priori ignorant approach is to assume

$$\Pr(S_i = k|\eta_1, \dots, \eta_K) = \eta_k, \tag{8}$$

where $\eta_k,\ k = 1, \dots, K - 1$ and $\eta_K = 1 - \sum_{k=1}^{K-1} \eta_k$ are the relative group sizes assumed to be unknown parameters to be estimated. Alternatively, the probability of cluster membership may depend on certain unit specific factors, hence the probability is parameterized as a multinomial logit model given as

$$\Pr(S_i = k|\{\gamma_k\}_{k=1}^{K-1}, z_i) = \frac{\exp\{z_i\gamma_k\}}{1 + \sum_{k=1}^{K-1} \exp\{z_i\gamma_k\}}. \tag{9}$$

The individual specific variables z_i stratifies hence the probability for a panel unit to belong to cluster k.[6] The logit structures coincides with the unconditional approach, when z_i includes only a constant. Variables in z_i may help to assign cluster membership. The model likelihood is given as

$$\mathcal{L}_C(Y|\theta, X) = \prod_{i=1}^{N} \sum_{k=1}^{K} \left[\prod_{t=1}^{T} \Phi\left((2y_{it} - 1)X_{it}\beta_k\right) \Pr(S_i = k|\{\gamma_k\}_{k=1}^{K-1}, z_i)\right]. \tag{10}$$

Given these possibilities to consider latent heterogeneity within probit models different concepts for marginal effects arise. Conceptually, the marginal effect summarizing all parameters within θ is given as

$$\frac{\partial}{\partial x}\Pr(y = 1|x = \overline{x}, \theta), \tag{11}$$

where the function form of the derivative depends on the considered model structure for incorporation of latent heterogeneity. For the pooled model specification, the marginal effect is well known to be

$$ME_P = \phi(\overline{x}\beta)\beta, \tag{12}$$

where $\phi(\cdot)$ denotes the density of the standard normal distribution. For the model specification incorporating latent heterogeneity via random coefficients, the marginal effect takes the form

$$ME_{RC} = \int \phi(\overline{x}\beta_i)\beta_i f(\beta_i|b, W)d\beta_i, \tag{13}$$

---

[6]Note that for identification reasons only $K - 1$ parameter vectors are specified, see Train (2003) for a discussion.

where the above given integral can be solved numerically or analytically, when assuming a normal distribution for the distribution $f(\beta_i|b, W)$.[7] When latent heterogeneity is considered via modeling of latent clusters, two concepts apply for assessment of marginal effects. First, a cluster specific measure is given as

$$ME_{CS} = \phi(\overline{x}\beta_k)\beta_k, \qquad k = 1, \ldots, K, \tag{14}$$

while a cluster robust concept is[8]

$$ME_{CR} = \sum_{k=1}^{K} \phi(\overline{x}\beta_k)\beta_k \Pr(S = k|\overline{z}, \{\gamma_k\}_{k=1}^{K-1}). \tag{15}$$

This set of different concepts are derived as measures for marginal effects and allow to gauge the influence of variables on the probability of an event under consideration of different forms of latent heterogeneity.[9]

3 Model Estimation

The inclusion of model structures capturing latent heterogeneity provides estimation problems, which are especially accessible to Bayesian estimation via MCMC methodology namely Gibbs sampling including Metropolis-Hastings updates. Using data augmentation, see Tanner and Wong (1987), the parameter vector is augmented to include the panel member specific cluster indices, which simplifies sampling from all other full conditional distributions.

Genuine estimation problems in the context of the model based clustering framework as a special type of mixture model are referred to in the literature as label switching and genuine multi modality, see Stephens (2000) and Frühwirth-Schnatter (2004). Label switching refers to the invariance of the likelihood under relabeling of clusters. The involved problem of label identification can be handled in different ways. The first solution is based on setting an identifying restriction on the parameter space, which hinders label switching. While the restrictions may be easily implemented it is important to find a restrictions, which efficiently separates the parameter space. If one chooses a restriction on parameters, which are almost identical within the clusters, the label switching problem is possibly still

---

[7]Note that the numerical solution is provided as a byproduct of the employed estimation routine, namely Gibbs sampling. However, the analytical solution is to be preferred for reasons of storing capacities and prevention of approximation errors.

[8]Alternatively to conditioning on a point $\overline{z}$, the unconditional probability can be used ($\Pr(S = k)$ instead of $\Pr(S = k|\overline{z})$).

[9]Note that also individual specific marginal effects might be interest, which can readily calculated based on individual specific expectation of the parameter vector.

present especially within the multivariate clustering considered here.[10] Alternatively, as suggested by Stephens (2000), the output of the unconstrained Gibbs sampler can be post screened. These relabeling algorithms provide a decision theoretic tool to decide on the identification of clusters. While an identifiability constraint hindering label switching is not a priori known and hence chosen arbitrarily, the use of an relabeling device for post screening is computationally more demanding, but based on a decision theoretic criterion to deal with label switching.

Based on a finite sample argument, genuine multi modality is not accessibly to maximum likelihood estimation and would hinder a correct assessment of parameter uncertainty. In contrast, posterior inference allows to display the occurrence of multimodality and hence provides a correct assessment of sample uncertainty in estimation yielding a further advantage of a Bayesian estimation approach.

The following sections provide the unconstrained sampler for the considered model framework. Furthermore, the relabeling algorithm for post screening, and the computation of the marginal likelihood using bridge sampling and the out-of-sample experiment are presented. One might argue that the need for the additional calculation steps - the relabeling algorithm, the bridge sampler, and for the out-of-sample experiment - to get interpretable parameter estimates and a model selection device is a drawback of the Bayesian approach. It is true that relabeling is an additional complication which is prevented by the use of maximum likelihood estimation techniques but model selection, namely the determination of the number of clusters causes additional conceptional problems with the maximum likelihood approach, too. Testing for the number of clusters within the maximum likelihood approach would run into a nuisance parameter problem and e.g. the asymptotic χ^2-distribution for the LR test statistic would not be valid.[11] These drawbacks of the frequentistic approach have already been noted by Geisser and Eddy (1979) who introduce in the literature the use of resampling strategies in a Bayesian context allowing for non-nested model comparison.

Furthermore, the Bayesian approach also allows direct comparison with non-nested models like the comparison between a cluster and a random coefficient model.

---

[10]To soften the problem of choosing a restriction a priori, the empirical literature suggests hence to choose the identifying restrictions on the basis of pre runs of the unconstrained Gibbs sampler.

[11]In contrast to the nuisance parameter problem occurring for structural break analysis where general asymptotic results have been derive by e.g. Andrews and Ploberger (1994), derivation of general results for testing the number of clusters is possibly hindered via the non availability of a natural ordering of clusters.

3.1 Estimation algorithm

This section presents the analysis of the considered model framework for a given number of clusters.[12] Comparison of model specifications is then based on the marginal likelihood of different models. Bayesian estimation is concerned about the posterior distribution of the parameter vector θ summarizing all model parameters

$$p(\theta|Y, X, Z) \propto \mathcal{L}(Y|X, Z, \theta)\pi(\theta), \tag{16}$$

where $\pi(\theta)$ denotes the prior distribution. Assuming prior independence of parameters concerning the conditional mean and parameters governing cluster membership allows to specify a multivariate normal prior of the parameters concerning the conditional mean.

Given the cluster indices estimation of the parameters within a cluster corresponds to the case of the pooled panel probit model which is straightforward to estimate in terms of the approach suggested by Albert and Chib (1993). The posterior distribution is augmented to include the latent variable y_{it}^*. Given this augmentation Gibbs sampling can be applied to obtain draws from the posterior distribution via iterative sampling from the closed form full conditional distributions. Additionally for the cluster model the Gibbs sampler is enhanced by two steps where the cluster indices are augmented and the parameters governing the cluster probabilities are drawn. The Gibbs sampling scheme has the following structure.

Step I Sample the latent variable y_{it}^* from a truncated normal with mean and variance

$$\mu_{y_{it}^*} = X_{it}\beta_{S_i}, \quad \sigma_{y_{it}^*} = 1, \tag{17}$$

where the truncation sphere is $(-\infty, 0)$ if $y_{it} = 0$ and $(0, \infty)$ if $y_{it} = 1$. β_k denotes the parameter vector of cluster k, while individual i belongs to cluster k: $\beta_{S_i} = \beta_k$.

Step II Sample β_k from the linear regression setup given by $Y_k^* = X_k\beta_k + E$ from a multivariate normal with moments

$$\Sigma_{\beta_k} = (X_k'X_k + \Omega_{\beta_k}^{-1})^{-1}, \qquad \mu_{\beta_k} = \Sigma_{\beta_k}^{-1}(X_k'Y_k^* + \Omega_{\beta_k}^{-1}\psi_{\beta_k}), \tag{18}$$

whereby Ω_{β_k} denotes the prior variance covariance matrix of $beta_k$ and ψ_{β_k} its prior mean vector. Y_k^* represents the latent variables of all individuals in cluster k. Step II is done for each cluster.

Step III Sample the cluster indicator S_i for each individual from a discrete distribution, where the full conditional probability is given as

$$\Pr(S_i = k|Y, X, Z, \theta) \propto \left[\prod_{t=1}^T \Phi\left((2y_{it} - 1)X_{it}\beta_k\right)\right] \frac{\exp\{z_i\gamma_k\}}{1 + \sum_{k=1}^{K-1} \exp\{z_i\gamma_k\}}. \tag{19}$$

---

[12]Note that the Gibbs scheme for the random coefficient panel probit model is documented in Aßmann (2007).

Step IV Simulation of the parameters governing the cluster probabilities is straightforward in case of the ignorant setup. The full conditional distribution is proportional to a Dirichlet distribution with parameters given as

$$p_1 = \sum_{i=1}^{N} I(S_i = 1), \quad \ldots, \quad p_K = 1 - \sum_{k=1}^{K-1} p_k \tag{20}$$

In case of the multinomial logit parameterization no direct sampling is possible. Hence, we adopt a Metropolis-Hastings scheme, see Chib and Greenberg (1995) for an introductive review, and use as a jumping distribution a normal distribution, where the mean μ^* is obtained via maximization of the posterior distribution given by the implicit likelihood of cluster probabilities and prior distribution over $\gamma_1, \ldots, \gamma_{K-1}$ and the covariance Σ^* as the corresponding inverted Hessian. Denote $\gamma = (\gamma_1, \ldots, \gamma_{K-1})$. A candidate draws is then accepted with probability

$$\min \left\{ \frac{\prod_{i=1}^{N} \frac{\exp\{z_i \gamma_{S_i}^*\}}{1+\sum_{k=1}^{K-1} \exp\{z_i \gamma_k^*\}} \exp\{-.5(\gamma^* - \psi_\gamma)' \Omega_\gamma^{-1}(\gamma* - \psi_\gamma) - .5(\gamma - \mu^*)'\Sigma^{*-1}(\gamma - \mu^*)\}}{\prod_{i=1}^{N} \frac{\exp\{z_i \gamma_k\}}{1+\sum_{k=1}^{K-1} \exp\{z_i \gamma_k\}} \exp\{-.5(\gamma - \psi_\gamma)' \Omega_\gamma^{-1}(\gamma - \psi_\gamma) - .5(\gamma^* - \mu^*)'\Sigma^{*-1}(\gamma^* - \mu^*)\}}, 1 \right\} . \tag{21}$$

This specific choice of the jumping distribution showed favorable acceptance rates compared to simpler random walk chains and provided only moderate autocorrelation within the draws.

3.2 Relabeling algorithm

Relabeling algorithms have been introduced within the literature on finite mixture models by Celeux (1998) and can be motivated via a decision theoretic approach. Relabeling is performed via minimizing the risk to misreport a draw from the Gibbs output. We adapt the relabeling algorithm suggested by Stephens (2000) for clustering inference in the context of stratified clustering within the panel probit model. Nevertheless, some discussion with respect to the severity of the label switching problem shall be provided. Label switching is connected to the following stylized sampling event. By incident two parameter vectors linked to two distinct clusters characteristics change, i.e. denoting one as the first and one as second, the first reflects the properties of the second and vice versa. The probability of such an event is the smaller the larger the parameter space is and the more distinct the cluster characteristics are. Since the parameter vectors have changing characteristics, individuals within the panel are regrouped into the clusters. Label switching can also be induced via an incidental regrouping of individuals into clusters. However, the larger the number of individuals is, the more unlikely is a complete regrouping of all members belonging to a certain cluster. Thus, the larger the parameter space and the more panel members are considered the less frequent is label switching to be observed within the Gibbs sampling sequences. Given this, we provide the relabeling

8

algorithm in the following for the most general model specification, where latent heterogeneity is modeled via clusters.

Let $P(\theta)$ denote the matrix of classification probabilities $(p_{ik}(\theta))$ given as

$$p_{ik}(\theta) = \Pr(S_i = k | Y, X, Z, \theta) \propto \left[\prod_{t=1}^{T} \Phi\left((2y_{it} - 1)X_{it}\beta_k\right) \right] \frac{\exp\{z_i\gamma_k\}}{1 + \sum_{k=1}^{K-1} \exp\{z_i\gamma_k\}} f(\beta_k | b_k, W_k). \quad (22)$$

The Kullback-Leibler divergence is operationalized to measure the loss involved in reporting and cluster assignment $\mathcal{Q}$ given as the matrix q_{ik} when the true distribution on clustering is $P(\theta)$. It has the form

$$KL(\mathcal{Q}; \theta) = \sum_{i=1}^{n} \sum_{k=1}^{K} p_{ik}(\theta) \log\left\{ \frac{p_{ik}(\theta)}{q_{ik}} \right\}. \quad (23)$$

Given an initial choice of relabeling $v_1, \ldots, v_R$, e.g. using the raw output from the Gibbs sampler, the relabeling algorithm consists then out of the following steps.

Step 1: Choose $\mathcal{Q}$ as

$$\mathcal{Q} = \arg\min \sum_{r=1}^{R} \sum_{i=1}^{n} \sum_{k=1}^{K} p_{ik}(v_r(\theta^{(r)})) \log\left\{ \frac{p_{ik}(v_r(\theta^{(r)}))}{q_{ik}} \right\}, \quad (24)$$

which is achieved via setting

$$q_{ik} = \frac{1}{R} \sum_{r=1}^{R} p_{ik}(v_r(\theta^{(r)})). \quad (25)$$

Step 2: Choose v_r to minimize

$$\sum_{i=1}^{n} \sum_{k=1}^{K} p_{ik}(v_r(\theta^{(r)})) \log\left\{ \frac{p_{ik}(v_r(\theta^{(r)}))}{q_{ik}} \right\}, \quad (26)$$

which is achieved in the present context of application via consideration of all $K!$ possibilities for each v_r, $r = 1, \ldots, R$.

The algorithm converges to the optimal fix point in the present context, where 20 iterations are found to ensure convergence. Some caveats apply with respect to storing requirements. Since the full augmented parameter vector must be stored, storing capacities must be cautiously provided, since the most general specification with shrinkage within clusters requires saving of $R \times N \times K \times p$ draws. Based on the relabeled and screened MCMC output, estimates of parameters and marginal effects are directly available as averages from this output.

3.3 Calculating the Marginal Likelihood via Bridge Sampling

Bridge sampling offers a conceptually straightforward method to calculate the Marginal Likelihood, see Frühwirth-Schnatter (2004). It has its particular advantages in the presence of multimodality and label switching as it does not depend on a certain density region of the posterior like the method of Chib (1995) and Chib and Jeliazkov (2001). The Marginal Likelihood is the normalizing constant of the posterior distribution of the model parameters. The non-normalized posterior can be denoted as

$$p^*(\theta|Y) = \mathcal{L}(Y|\theta)\pi(\theta), \tag{27}$$

where $\mathcal{L}(Y|\theta)$ represents the likelihood and $\pi(\theta)$ represents the prior. Then the normalized posterior is given as

$$p(\theta|Y) = \frac{p^*(\theta|Y)}{p(Y)}. \tag{28}$$

Now assume $q(\theta)$ which is a simple approximation of the posterior $p(\theta|Y)$ with known normalizing constant. Furthermore, assume $\alpha(\theta)$ to be an arbitrary function fulfilling:

$$\int \alpha(\theta)p(\theta|Y)q(\theta)\nu(d\theta) > 0. \tag{29}$$

Bridge-sampling is based on the following result:

$$1 = \frac{E_q(\alpha(\theta)p(\theta|Y))}{E_p(\alpha(\theta)q(\theta))}, \tag{30}$$

which can be transformed into

$$p(Y) = \frac{E_q(\alpha(\theta)p^*(\theta|Y))}{E_p(\alpha(\theta)q(\theta))}, \tag{31}$$

due to Equation (28). A consistent estimator of the marginal likelihood is then given by

$$\hat{p}(Y) = \frac{L^{-1}\sum_{l=1}^{L}\alpha(\tilde{\theta}^{(l)})p^*(\tilde{\theta}^{(l)}|Y)}{M^{-1}\sum_{m=1}^{M}\alpha(\theta^{(m)})q(\theta^{(m)})}, \tag{32}$$

where expectations are replaced by sample averages, $\tilde{\theta}^{(l)}$ are draws from the auxiliary density $q(\theta)$ and $\theta^{(m)}$ are draws from the before noted posterior sampler namly the same draws of the MCMC-Algorithm used for estimation of the model parameters. The estimator in Equation (32) is called a general bridge-sampler. By chosing different functions $\alpha(\cdot)$ some special cases arise.[13] If one applies $\alpha(\theta) = 1/q(\theta)$ the bridge-sampling estimator is an importance sampling estimator:

$$\hat{p}(Y) = L^{-1}\sum_{l=1}^{L}\frac{p^*(\tilde{\theta}^{(l)}|Y))}{q(\tilde{\theta}^{(l)})}. \tag{33}$$

---

[13]The approach of Gelfand and Dey (1994) for calculating the marginal likelihood corresponds to the choice $\alpha(\theta) = 1/p^*(\theta|Y)$.

Meng and Wong (1996) discuss the asymptotically optimal choice of $\alpha(\theta)$ and propose the use of

$$\alpha(\theta) \propto \frac{1}{Lq(\theta) + Mp(\theta|Y)}. \tag{34}$$

While $p(\theta|Y)$ is unknown as the marginal likelihood is unknown and has to be replaced by its estimate Meng and Wong (1996) propose an iterative procedure to estimate the marginal likelihood and thereby the posterior integrational constant. In the k-th step of the iteration the posterior is estimated by

$$\hat{p}(\theta|Y) = \frac{p^*(\theta|Y)}{\hat{p}_{k-1}(Y)}, \tag{35}$$

where $\hat{p}_{k-1}(Y)$ results from the following recursion:

$$\hat{p}_k(Y) = \hat{p}_{k-1}(Y) \frac{L^{-1} \sum_{l=1}^{L} \frac{\hat{p}(\tilde{\theta}^{(l)}|Y)}{Lq(\tilde{\theta}^{(l)}) + M\hat{p}(\tilde{\theta}^{(l)}|Y)}}{M^{-1} \sum_{m=1}^{M} \frac{q(\theta^{(m)})}{Lq(\theta^{(m)}) + M\hat{p}(\theta^{(m)}|Y)}}. \tag{36}$$

The necessary importance density $q(\cdot)$ is constructed using Rao-Blackwellisation as proposed by Gelfand and Smith (1990). This ensures the capability of the employed importance density to deal with the possibly occurring multi-modality of the underlying function. It is hence given as

$$q(\tilde{\theta}^{(l)}) = \frac{1}{H} \sum_{h=1}^{T} \prod_{k=1}^{K-1} f(\tilde{\gamma}_k^{(l)}|m_{\gamma_k}^{(h)}) \prod_{k=1}^{K} f(\tilde{\beta}_k^{(l)}|m_{\beta_k}^{(h)}), \tag{37}$$

where $f(\cdot|\cdot)$ denote the full conditional distribution and $\{m_{\gamma_k}^{(h)}, m_{\beta_k}^{(h)}\}_{h=1}^{H}$ denote a random resample of the MCMC output to gain the all integrating constants of the of the full conditional distributions. Since integrating constant is not known analytically for $f(\tilde{\gamma}_k^{(l)}|\cdot)$, it is approximated via the involved MH-algorithm, compare Step IV, Section 3.1. Mind, the L draws (of S) to calculate the importance density can be obtained by resampling the MCMC output. As Frühwirth-Schnatter (2004) points out convergence is achieved quickly (after 5 iterations in our application), where starting values for the recursion can be gained from the importance sampling approach. Formally, we assume convergence achieved, when the difference in marginal likelihood values is less than $1e-3$.

3.4   Cross validation for determining the number of clusters

The marginal likelihood provides a consistent measure of the goodness of fit, which is adequate for non-nested model comparison. However, the high dimensional integration problem arising within the computation of the marginal likelihood, which is necessarily based on numerical approximations via simulation methods, makes it attractive to consider alternative measures of goodness of fit. A common approach is to use the predictive ability of a model to assess its fit.[14] Typically, to forecast

---

[14]See e.g. Geisser and Eddy (1979).

a binary variable in a probit model the cdf of the latent model (here: given the cluster indication, see Equation 38) is calculated and checked whether it exceeds 0.5. If so the model predicts a one, i.e.

$$I\left[\Pr(y_{it}^* \geq 0 | y_{it} = 1, S_i) = \Phi(X_{it}\beta_{S_i}) > 0.5\right] \tag{38}$$

By the ROC measure Egan (1975) provides an extension. Based on the predictive performance of the model under consideration, it comprises the set $\{a(w), b(w)\}$, $\{w : 0 \leq w \leq 1\}$, where

$$a(w) = 1 - \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} I\left[\Pr(y_{it}^* \geq 0 | y_{it} = 1, S_i) > w\right] \tag{39}$$

$$b(w) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} I\left[\Pr(y_{it}^* \geq 0 | y_{it} = 0, S_i) > w\right]. \tag{40}$$

The ROC graph assesses therefore the predictive performance for alternative values of the prediction threshold ranging from 0 to 1, and not just 0.5.[15]

We apply the simple forecasting criterion as well as the ROC in a cross validation setup to prevent overfitting.[16] This is done as follows: The sample is split along its time dimension into a estimation and a prediction part. We run a modified Gibbs sampler where the distributions of the parameters and the cluster indicators solely depend on the observations of the estimation part. Additionally within each Gibbs run the values of the latent models of the observations of the prediction part are calculated (given the current draw of the parameter vector as well as of the cluster indicators). Afterwards, these values are averaged over all Gibbs runs and are than used to gain forecasts and the ROC measure. Thereby, the problem of relabeling is prevented, as the value of the latent model is unchanged by relabeling.

4 Simulation Study

We assess via a simulation study the accuracy of the marginal likelihood concept to identify the correct number of clusters and compare it with an alternative model selection device based on forecasting accuracy in an out-of-sample experiment. The simulation study is performed for two panel sizes A and B. Within panel size A time dimension is $T = 10$ and the number of individuals is $N = 50$, whereas panel size B is $T = 5$ and $N = 500$. For each of the two panel sizes we assume a probit model with two explaining variables and a constant of the right hand side of the latent model, i.e.

$$y_{it}^* = X_{it}\beta_{S_i} + e_{it}, \tag{41}$$

---

[15]In fact the above mentioned probabilities are replaced in empirical analysis by the relative frequencies obtained via computation of the predictive probabilities for each single observations.

[16]See Stone (1974).

with $X_{it} = [1 \ x_{it}^{(1)} \ x_{it}^{(2)}]$, where we generally assume that the individuals belong to either of three unobservable clusters denoted by S_i. All together, we generate for the two panel sizes 500 repetitions from 11 parameter scenarios. Thereby, the 11 parameter scenarios reflect different cluster constellations ranging from a pooled scenario up to a scenario with 3 clusters and stratified cluster probabilities. For each repetition in each parameter scenario we estimate panel probit models assuming 1 through 4 clusters. The prior assumptions employed throughout the simulation study are given in Table (1).

The scenarios are generated in the following manner. For panel size A and B the corresponding regressors $x_{it}^{(1)}$, $x_{it}^{(2)}$ and z_i are drawn once from standard normal distributions each. Firstly, we draw from the logit distribution (with parameters $\{\gamma_k\}_{k=1}^{K-1}$) the cluster number of individual i. After obtaining the cluster membership the dependent variable is drawn from the corresponding probit model. By setting the cluster specific parameters in some scenarios to the same values the number of clusters are reduced to one (pooled) and two clusters. We consider furthermore different degrees of inhomogeneity between clusters to illustrate how much inhomogeneity is necessary for proper identification of different clusters. We measure the degree of inhomogeneity present within the data sets via computation of a global measure of inhomogeneity given as

$$IH = \sum_{p=1}^{\mathrm{con}} \sum_{k=1}^{K} (\beta_k^p - \overline{\beta}^p)^2 \eta_k, \tag{42}$$

where $^{\mathrm{con}}$ denotes the number of regressors and η_i denotes unconditional cluster probability of cluster i. However, to provide a more accurate view on the present latent heterogeneity, we accompany the global measure of inhomogeneity via the set of pairwise inhomogeneity measures, i.e. the inhomogeneity between two distinct clusters. Note also in case of stratified cluster probabilities inhomogeneity is present also within the parameters of these probabilities, which we assess similarly as above via the logit coefficients.

4.1 Marginal Likelihood

Table (2) gives the results of the simulation study in terms of the relative frequency, the marginal likelihood criterium chooses the corresponding number of clusters (1 through 4) for panel size A. Column 10 gives the true number of clusters. In Scenario I the marginal likelihood estimate for the true (pooled) model is in almost all cases higher than for the models with clusters. The model selection approach is at least well suited to reject the hypotheses of clustered data. If the simulated data is clustered the ability of the model selection approach to detect the right number of clusters is mixed and depends on the degree of homogeneity between the clusters and on whether one regards explaining variables for the (stratified) clustering.

13

Data in Scenarios II through IV are simulated from two clusters. In almost all cases the right number of clusters are estimated even when as in Scenario III and IV the heterogeneity between clusters is rather low. In Scenario V three clusters exist, however the parameters β are identical for two of them. Thus, these clusters are only different with respect to the logit part or the individuals in two of the three cluster behave like each other but for different reasons. The model selection can hardly detect the right number of clusters. It seems to be mainly driven by differences in β. In Scenarios VI and VII three clusters different in γ and β are assumed. However, the heterogeneity between two of theses clusters is rather low ($MIH_{13} =$ compare to scenario III). Again the model selection procedure identifies only two clusters in most cases, but results improve if the conditioning information in the logit regression is taken into account. Thus it is advisable to do an one step analysis instead of a two step procedure, where in the first step clusters are estimated and in a second step logit parameters are estimated conditional on inferred states. This is even stressed in the assessment of Scenarios $VIII$ and IX where the heterogeneity between the β coefficients of the clusters is even lower, so that in a reasonable number of simulation runs (about 10 %) the marginal likelihood criterium even favors a pooled model in both scenarios. In Scenario $VIII$ the true number of clusters is detected only in very few simulation runs due to the high similarity between β_1 and β_2 ($MIH_{12} = 0.015$). In Scenario IX when the logit specification is taken into account the number of right model selections rises above 50 %. Finally, Scenarios X and XI provide an analysis for three rather heterogeneous cluster. Accordingly, the model selection procedure detects in almost all cases the true number of cluster. Qualitatively these results do not heavily depend on the structure of the data as the results for panel size B given in Table (3) show. They allow mainly the same conclusions as the results given for panel size A.

4.2 Cross validation approach

In this paragraph the results of the assessment of the marginal likelihood are compared to out-of-sample prediction criteria for model selection. We run an out-of-sample prediction exercises. The sample is spilt along time into an estimation sample containing 80 % of the observations of the whole sample while the remaining 20 % are to be forecasted. To expand the number of forecasts this partition (and thereby the forecasting exercise) is done five times in that way that all observations are once part of the forecasting sample. Results are averaged over this five different partitions.

Note that this approach allows the use of the unrestricted Gibbs-Sampler without application of relabeling to reach sensible results. In each Gibbs run we apply the set of sampled parameters on the covariates of the observations to be forecasted to obtain a forecast of the binary variable in each run.[17]

---

[17]Mind, the observations as well as its covariates do not enter the other parts of the Gibbs sampler as there only

The forecasting distributions are thus byproducts of the Gibbs output. If the estimated probability exceeds 0.5, a one is taken as forecast for this observation and compared to the actually observed. Additionally, the ROC measure is calculated allowing for different probability borders in forecasting (hence, the popular approach with 0.5 is a special case of the ROC approach). The forecasts are done for the same models and scenarios as in the case of the marginal likelihood assessment. The number of clusters is then determined by the best model according to the forecasting and the ROC criterion respectively. Tables (4) and (6) report the relative frequencies the above explained criteria decide on the specific number of clusters in Panel size A.

Both, the forecasting and the ROC criterion provide very similar conclusions compared to the marginal likelihood. Thus the case of homogenity with no clusters is detected almost with certainty (scenario I). The same is true with two true clusters (scenarios II through IV) and with three true clusters if the heterogenity of the clusters is high enough (scenarios X and XI), whereby the marginal likelihood shows considerable less dispersion in the two latter scenarios. In scenarios V through VII both criteria give advice to two clusters in more than 90 percent of the cases. While the marginal likelihood also has a high tendency to advice two clusters in these scenarios the rate of true selections is higher than for both competitors. Finally, in scenarios VII and IX the the most pronounced differences between the model selection criteria occur. The forecasting criterion has a high tendency to neglect the heterogeneity in 63 percent and 36 percent, respectively, it tends to just one cluster. The marginal likelihood and the ROC approach both have the same median: In scenario $VIII$ both prefer two clusters mostly. Whereby the ROC approach has a higher dispersion. In 19 percent of the cases it chooses the right number of cluster, but in 26 percent of the cases it assumes just one clsuter and in 10 percent even four clusters. In scenario IX the number of cases, where the true cluster dimension is detected is with 58 percent similar to the marginal likelihood approach, but here the ROC has a tendency to overfitting and more often signals four cluster. Tables (5) and (7) provide the results for Panel size B. Qualitatively the results are mainly unchanged. To some degree the dispersion of the decisions is a bit less and results are a bit improved in that sense for the ROC approach.

5   Empirical Illustration

The usefullness of model based clustering for modeling latent heterogeneity shall be illustrated using the data set studied in Bertschek and Lechner (1998). The data set is well known in the empirical literature and Greene (2004b) documents the presence of a considerable degree of latent heterogeneity within the data. The following listing gives the variables used for analysis, for a detailed description,

observations of the estimation sample are regarded.

see Bertscheck and Lechner (1998):

$y_{it} = 1$ if a product innovation was realized by firm i in year t, 0 otherwise; $x_{2_{it}}$ Log of industry sales in DM; $x_{3_{it}}$ Import share=ratio of industry imports to (industry sales plus imports; $x_{4_{it}}$ Relative firm size = ratio of employment in business unit to employment in the industry (times 30); $x_{5_{it}}$ FDI share = Ratio of industry foreign direct investment to (industry sales plus imports); $x_{6_{it}}$ Productivity = Ratio of industry value added to industry employment; $x_{7_{it}}$ Raw materials sector=1 if the firm is in this sector; $x_{8_{it}}$ Investment goods sector=1 if the firm is in this sector.

The data set contains 1270 firms, where the time dimension covers the five years from 1984 to 1988. Based on the insight delivered by the simulation study in the previous section, we consider the marginal likelihood analysis and the out-of-sample analysis with partitioning along the time dimension for determination of the number of latent clusters. The results are provided in Table (8).

Comparison of the pooled specification with the specification incorporating random coefficients via the marginal likelihood shows the random coefficients specification to be very strongly preferred according to Jeffreys' scale, see Jeffreys (1961). Consideration of latent heterogeneity via model based clustering asks for specifying the number of clusters. Table (8) gives the corresponding marginal likelihoods for models with two, three, and four clusters. For each number of cluster stratified and non stratified cluster specific probabilities have been considered. The marginal likelihood reveals that the consideration of three cluster in conjunction with non stratified cluster probabilities is the preferred model specification.

A similar conclusion is drawn on the basis of the performed out-of-sample prediction experiment. For both out-of-sample experiments with sample partitioning along the time dimension the ratio of correctly predicted observations and the ROC measure is the highest for the model specification incorporating latent heterogeneity via estimation of three latent clusters with non stratified cluster probabilities. Thus, also the empirical illustration provides evidence for the accuracy of the here proposed cross validation strategy for deciding on the preferred number of latent clusters.

To reveal the impact of the considered forms of latent heterogeneity on gauging the impact of variables on the dependent variable firm innovation, we focus discussion on the implied different marginal effects. With respect to empirical results, Table (9) provides the pooled estimation results. All variables show significant influence on the dependent variable. Furthermore, marginal effects show that higher sales, larger size, higher import ratio and foreign direct investment, as well as membership in investment good branch have positive effect on the patent activity. Negative effects on patent activity are documented for productivity and the raw materials sector.

Incorporation of latent heterogeneity via random coefficients provides some significant changes with respect to the influence some variables exhibit on patent activity. Bayesian estimation results

16

are shown in Table (10).[18] In specific, the productivity of a firm has no longer positive influence on the patent activity. The corresponding heterogeneity robust (in the sense of the random coefficient specification) marginal effect is not different from zero at conventional levels. The marginal effect for this specification using a numerical solution to the involved integral is computed as

$$\widetilde{ME_{RC}} = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{N}\sum_{i=1}^{N}\phi(\overline{x}\beta_i^{(r)})\beta_i^{(r)}, \tag{43}$$

where $\beta_i^{(r)}$, $r = 1, \ldots, R$ denote the posterior draws including the random coefficients. Using an analytical solution to the integration provides an estimate computed as

$$\widetilde{ME_{RC}} = \frac{1}{R}\sum_{r=1}^{R}\exp\{-.5(b^{(r)\prime}W^{-1,(r)}b^{(r)}) - (W^{-1,(r)}b^{(r)})\prime[\overline{x}\prime\overline{x} + W^{-1,(r)}]^{-1}(W^{-1,(r)}b^{(r)}))\}$$
$$(2\pi)^{-.5}\det(W^{(r)})^{-.5}(\det(\overline{x}\prime\overline{x} + W^{-1,(r)}))^{.5}(\overline{x}\prime\overline{x} + W^{-1,(r)})^{-1}(W^{-1,(r)}b^{(r)}). \tag{44}$$

The estimation results corresponding to the estimation of three clusters are given in Table (11). Estimates are based on post screened Gibbs output. Figure (1) gives the posterior draws before post screening exhibit potentially multimodality and label switching. The post screened Gibbs output is given in Figure (2). It is seen that label switching has not been present in the Gibbs output, however slight genuine multi modality might be present, as shown in Figure (3) displaying the estimated posterior densities.[19]

With respect to empirical results, the preferred cluster specification provides an alternative account of latent heterogeneity and recommends alternative interpretation of results. Consideration of latent heterogeneity via latent clusters allows to gain several insights with respect to the marginal effects when latent heterogeneity is present. At first, cluster specific marginal effects can be considered. These can be readily derived from the screened Gibbs output as

$$\widetilde{ME_C} = \frac{1}{R}\sum_{r=1}^{R}\phi(\overline{x}\beta_k^{(r)})\beta_k^{(r)}, \quad k = 1, \ldots, K \tag{45}$$

and are based on cluster specific means $\overline{x}_k$. However, generally no knowledge on cluster membership is available. Therefore, of interest is also the distribution of marginal effects within the whole sample. Note that an heterogeneity robust average effect can be computed based on weighted averages of

---

[18]The random coefficient model has two blocks of parameters. The first block are the means of the random parameters and the second its variances. For the first block of parameters we assume a priori a multivariate normal distribution and for the second block we assume a Wishart distribution, the prior moments are given in Table (1). The Gibbs sampler is used to get inference on parameters and marginal effects.

[19]Numerical optimization routines to perform a Maximum Likelihood estimation yield differing results depending on their starting conditions. It can hardly be guaranteed to reach the global maximum with a low number of trials.

the draws given as

$$\widetilde{ME_{CR}} = \frac{1}{R} \sum_{r=1}^{R} \sum_{k=1}^{K} w_k^{(r)} \phi(\overline{x}\beta_k^{(r)})\beta_k^{(r)}, \tag{46}$$

where $w_k^{(r)}$ denotes the fraction of individuals in cluster k in Gibbs iteration r. Alternatively to $w_k^{(r)}$ a conditional probability can be used, for example $\Pr(S_i = k | \overline{x}, \{\gamma_k^{(r)}\}_{k=1}^{K-1})$. Draws of the cluster specific marginal effects are hence weighted with the number of individuals, which are in this cluster for the considered draw. Table (12) gives the corresponding results.

While both, the random coefficient and the model-based clustering approach aim at a general representation of latent heterogeneity, several differences between the two specification are revealed, when comparing marginal effects. While within the population robust marginal effects given in Table (12) based on the latent cluster specification only log sales, relative firm size, and investment good sector have substantial influence on firm innovation, the random coefficient approach also exhibits substantial effect of imports and foreign direct investment. Furthermore, besides these qualitative differences, also quantitative differences for relative size and imports are revealed. In contrast very similar marginal effects are documented for log sales and the investment good sector indicator, see also the pooled specification in Table (9). These findings suggest that not all variables capture latent heterogeneity to the same extent. When looking at cluster specific marginal effects with estimation results being provided in Table (13), these provide insight into to what extent the clusters are distinct. Interestingly, within the cluster labeled as second all variables show substantial influence on firm innovation, however at a quantitative level differing from the pooled specification. For the first cluster only log sales and relative firm size are documented to have substantial impact, while for the third cluster log sales and the investment good sector influence firm innovation substantially. These different patterns provide insight into different latent cluster specific firm types, for which a pooled specification does not provide valid inference of the relationship between firm innovation and economic determinants thereof.

Given the latent cluster specification being the preferred model, comparison between cluster specific and random coefficient marginal effects suggests that the underlying normal distribution for latent heterogeneity does not allow to represent the full extent of the latent heterogeneity present within the considered empirical illustration. Thus, the empirical example reveals the importance to consider modeling latent heterogeneity in different manners.

6 Conclusion

This paper provides Bayesian estimation procedures for panel probit models incorporating model based clustering. Based on the different forms of incorporating latent heterogeneity, we provide

a discussion of different concepts of marginal effects. Furthermore, we discuss the issue of model selection with respect to the number of clusters. While the marginal likelihood is typically the preferred device for model selection in a Bayesian framework we propose the use of a cross-validation approach based on the ROC measure. The latter approach is less demanding than the marginal likelihood as it only needs the implementation of the Gibbs sampler and no further algorithms like the bridge sampler which is used for calculating the marginal likelihood. The results of a simulation study show that given the presence of a certain degree of inhomogeneity, both concepts correctly identifies the true number of latent clusters. Due to the high degree of complication in calculation marginal likelihoods for panel probit models the cross-validation approach seems to be a reasonable model selection tool. As a side result we find that stratified cluster probabilities perform overall better than non-stratified cluster probabilities, thus a one step procedure is preferred against possible two step procedures, where in a first step clusters are determined and in a second step a multinomial model is estimated.

Within the chosen empirical application we find strong evidence for latent heterogeneity captured via latent clusters. Based on the different forms of latent heterogeneity one arrives at different conclusions concerning the impact of explaining variables on the dependent variable conceptualized via marginal effects. This is a strong case for the class of models applied here and for the Bayesian approach to handle this class of models, as the Bayesian approach is able to deal with genuine multi modality and provides a consistent tool for model selection.

References

[1] Albert, J. and S. Chib: 1993, 'Bayesian Analysis of Binary and Polychotmous Response Data'. *Journal of the American Statistical Association* **88**, 669–679.

[2] Andrews, D. W. K. and W. Ploberger: 1994, 'Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative'. *Econometrica* **62**(6), 1383–1414.

[3] Aßmann, C.: 2007, 'Determinants and Costs of Current Account Reversals under Heterogeneity and Serial Correlation'. Economic Working Paper No. 2007-17, CAU Kiel.

[4] Ben-Akiva, M. and D. Bolduc: 1996, 'Multinomial Probit with a logit kernel and a general parametric specification of the covariance structure'. Working paper, Department of Civil Engineering, MIT.

[5] Bertschek, I. and M. Lechner: 1998, 'Convenient estimators for the panel probit model'. *Journal of Econometrics* **87**(2), 329–372.

[6] Cameron, C. and P. Trivedi: 2005, *Microeconometrics: Methods and Applications*. Cambridge University Press.

[7] Celeux, G.: 1998, 'Bayesian inference for mixtures: the label switching problem'. *COMPSTAT* pp. 227–232.

[8] Chen, J. and A. Khalili: 2008, 'Order Selection in Finite Mixture Models With a Nonsmooth Penalty'. *Journal of the American Statistical Association* **103**(484), 1674–1683.

[9] Chib, S.: 1995, 'Marginal Likelihood from the Gibbs output'. *Journal of the American Statistical Association* **90**(432), 1313–1321.

[10] Chib, S. and E. Greenberg: 1995, 'Understanding the Metropolis-Hastings Algorithm'. *The American Statistician* **49**(4), 327–335.

[11] Chib, S. and I. Jeliazkov: 2001, 'Marginal Likelihood from the Metropolis-Hastings Output'. *Journal of the American Statistical Association* **96**(453), 270–281.

[12] Dunson, D., A. Herring, and A. Siega-Riz: 2008, 'Bayesian Inference on changes in response densities over predictor clusters'. *Journal of the American Statistical Association* **103**(484), 1508–1517.

[13] Egan, J.: 1975, *Signal Detection Theory and ROC analysis*, Series in Cognition and Perception. Academic Press, New York.

[14] Fraley, C. and A. Raftery: 2002, 'Model-Based Clustering, Discriminant Analysis, and Density Estimation'. *Journal of the American Statistical Association* **97**, 611–631.

[15] Frühwirth-Schnatter, S.: 2004, 'Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques'. *The Econometrics Journal* **7**, 143–167.

[16] Frühwirth-Schnatter, S. and S. Kaufmann: 2008, 'Model-Based Clustering of Multiple Time Series'. *Journal of Business & Economic Statistics* **26**(1), 78–89.

[17] Geisser, S. and W. Eddy: 1979, 'A predictive approach to model selection'. *Journal of the American Statistical Association* **74**(365), 153–160.

[18] Gelfand, A. E. and D. K. Dey: 1994, 'Bayesian model choice: Asymptotics and exact calculations'. *Journal of Royal Statistical Society B* **56**, 501514.

[19] Gelfand, A. E. and A. Smith: 1990, 'Sampling based approaches to calculating marginal densities'. *Journal of the American Statistical Society* **85**, 398–409.

[20] Greene, W.: 2004a, 'The behaviour of the maximum likelihood estimator for limited dependent variable models in the presence of fixed effects'. *Econometrics Journal* **7**, 98–119.

[21] Greene, W.: 2004b, 'Convenient estimators for the panel probit model: Further results'. *Empirical Economics* **29**, 29–47.

[22] Greene, W. and D. Hensher: 2003, 'A latent class model for discrete choice analysis: contrast to a mixed logit'. *Transportation Research Part B* **37**, 681–698.

[23] Handcock, M., A. E. Raftery, and J. Tantrum: 2007, 'Model-based clustering for social networks'. *Journal of the Royal Statistical Society A* **170**(2), 301–354.

[24] Heard, N. A., C. C. Holmes, and D. A. Stephens: 2006, 'A Quantitative Study of Gene Regulation Involved in the Immune Response of Anopheline Mosquitoes: An Application of Bayesian Hierarchical Clustering of Curves'. *Journal of the American Statistical Society* **101**(473), 18–29.

[25] Ishwaran, H., L. F. James, and J. Sun: 2001, 'Bayesian Model Selection in Finite Mixtures by Marginal Density Decomposition'. *Journal of the American Statistical Association* **96**(456), 1316–1332.

[26] Jeffreys, H.: 1961, *Theory of probability*. Oxford: Clarendon Press.

[27] Lancaster, T.: 2000, 'The incidental parameter problem since 1948'. *Journal of Econometrics* **95**, 391–413.

[28] Mehndiratta, S.: 1996, 'Time-of-day effects in inter-city business travel'. Ph.D. thesis, University of California, Berkeley.

[29] Meng, X.-L. and W. H. Wong: 1996, 'Simulating ratios of normalizing constants via a simple identity: A theoretical exploration'. *Statistica Sinica* **6**, 831860.

[30] Ray, S. and B. G. Lindsay: 2008, 'Model selection in high dimensions: a quadratic-risk-based-approach'. *Journal of the Royal Statistical Society, Series B* **70**(1), 95–118.

[31] Revelt, D. and K. Train: 1998, 'Mixed Logit with repeated choices'. *Review of Economics and Statistics* **80**, 647–657.

[32] Stephens, M.: 2000, 'Dealing with the label switching in mixture models'. *Journal of the Royal Statistical Society B* **62**(4), 795–809.

[33] Stone, M.: 1974, 'Cross-validatory choice and assessment of statistical predictions'. *Journal of the Royal Statistical Society B* **36**, 111–147.

[34] Tanner, M. and W. Wong: 1987, 'The calculation of posterior distributions by data augmentation'. *Journal of the American Statistical Association* **92**(398), 528–540.

[35] Train, K. E.: 2003, *Discrete Choice Methods with Simulation*. Cambridge.

Tables

Table 1: Prior distributions

parameter			
	pooled specification		
β	$\mathcal{N}$	0	$10I$
	random coefficient specification		
b	$\mathcal{N}$	0	$10I$
W_b	$\mathcal{IW}$	$k+1$	I
	clustered specification		
$\beta_k, \ k=1,\ldots,K$	$\mathcal{N}$	0	$10I$
$\gamma_k, \ k=1,\ldots,K$	$\mathcal{N}$	0	$1I$

Table 2: Simulation Study - Model selection via marginal likelihood - Panel size A

Scenario	1	2	3	4	MIH/GIH	MIH_{12}/GIH_{12}	MIH_{13}/GIH_{13}	MIH_{23}/GIH_{23}	K true	stratified
I	0.990	0.010	0	0	0/–	–	–	–	1	no
II	0	0.991	0.010	0	1.325 / 0	1.325/ 0	–/–	–/–	2	no
III	0	1.000	0	0	2.250 /0.116	2.250/ 0.116	–/–	–/–	2	no
IV	0	0.998	0.002	0	2.250 / 0.373	2.250/ 0.373	–/–	–/–	2	yes
V	0	0.952	0.048	0	2.270 / 0.1342	0 / 0.420	2.250/0.250	2.250/0.260	3	yes
VI	0	0.948	0.052	0	1.251 /0	1.290 / 0	0.125 / 0	1.165 / 0	3	no
VII	0	0.782	0.216	0.002	1.251 /0.097	1.290 / 0.123	0.125 /0.023	1.165 / 0.040	3	yes
$VIII$	0.102	0.890	0.008	0	0.240 /0	0.0150/0	0.0900 /0	0.1050 /0	3	no
IX	0.098	0.326	0.574	0.002	0.240 /0.097	0.045 / 0.123	0.270 / 0.023	0.315 /0.040	3	yes
X	0	0	1.000	0	1.500 /0	1.313/0	0.750/0	1.313 /0	3	no
XI	0	0.008	0.990	0.002	0.606 /0.1333	0.250 / 0.123	0.750 / 0.023	0.500 / 0.040	3	yes

Notes: Simulation of 500 data sets for each scenario I–XI. Scenarios are described in Section xx. For each scenario the columns denotes the relative frequency the marginal likelihood decided for a specific number of clusters. K true denotes the true number of clusters within the scenarios. IH denotes the inhomogeneity of the specified cluster specific parameters.

Table 3: Simulation Study - Model selection via marginal likelihood - Panel size B

Scenario	1	2	3	4	MIH/GIH	MIH_{12}/GIH_{12}	MIH_{13}/GIH_{13}	MIH_{23}/GIH_{23}	K true	stratified
I	0.990	0.010	0	0	0/0	0/0	0/0	0/0	1	no
II	0	0.996	0.004	0	0.3874/-	-/-	-/-	-/-	2	no
III	0	1.000	0	0	0.0556 /-	0/-	0.7500/-	0.7500/-	2	no
IV	0	0.996	0.004	0	0.0556 /0.2068	0/0	0.7500/0.1863	0.7500/0.1863	2	yes
V	0	0.972	0.024	0.004	0.0556 /0.1311	0/0.2050	0.7500/0.1250	0.7500/0.1350	3	yes
VI	0	0.738	0.262	0	0.3840 /-	0.4300/-	0.0417 /-	0.3883/-	3	no
VII	0	0.510	0.490	0	0.3840 /0.1327	0.4300/0.0612	0.0417 /0.0113	0.3883/ 0.0200	3	yes
$VIII$	0.046	0.904	0.050	0	0.0783 /-	0.0150/-	0.0900 /-	0.1050 /-	3	no
IX	0.006	0.226	0.768	0	0.0783 /0.1327	0.0150/0.0612	0.0900 /0.0113	0.1050 /0.0200	3	yes
X	0	0	0.994	0.006	0.4815 /-	0.4375/-	0.2500/-	0.4375/-	3	no
XI	0	0	0.994	0.006	0.2528 /0.1327	0.0833/0.0612	0.2500 /0.0113	0.1667/0.0200	3	yes

Notes: Simulation of 500 data sets for each scenario $I–XI$. Scenarios are described in Section xx. For each scenario the columns denotes the relative frequency the marginal likelihood decided for a specific number of clusters. K true denotes the true number of clusters within the scenarios. IH denotes the inhomogeneity of the specified cluster specific parameters.

Table 4: Simulation Study - Model selection via Forecasting Criterion (T) - Panel size A

Scenario	1	2	3	4	MIH/GIH	MIH_{12}/GIH_{12}	MIH_{13}/GIH_{13}	MIH_{23}/GIH_{23}	K true	stratified
I	0.9900	0.0100	0	0	0/–	–	–	–	1	no
II	0	0.9950	0.0050	0	1.325 / 0	1.325/ 0	–/–	–/–	2	no
III	0	1.0000	0	0	2.250 /0.116	2.250/ 0.116	–/–	–/–	2	no
IV	0	1.0000	0	0	2.250 / 0.373	2.250/ 0.373	–/–	–/–	2	yes
V	0	0.9800	0.0200	0	2.270 / 0.1342	0 / 0.420	2.250/0.250	2.250/0.260	3	yes
VI	0	0.9850	0.0150	0	1.251 /0	1.290 / 0	0.125 / 0	1.165 / 0	3	no
VII	0	0.9000	0.0900	0.0100	1.251 /0.097	1.290 / 0.123	0.125 /0.023	1.165 / 0.040	3	yes
$VIII$	0.6300	0.2450	0.0800	0.0450	0.240 /0	0.0150/0	0.0900 /0	0.1050 /0	3	no
IX	0.3600	0.0450	0.3950	0.2000	0.240 /0.097	0.045 / 0.123	0.270 / 0.023	0.315 /0.040	3	yes
X	0	0	0.9900	0.0100	1.500 /0	1.313/0	0.750/0	1.313 /0	3	no
XI	0.0150	0.0050	0.9100	0.0700	0.606 /0.1333	0.250 / 0.123	0.750 / 0.023	0.500 / 0.040	3	yes

Notes: Simulation of 500 data sets for each scenario I–XI. Scenarios are described in Section xx. For each scenario the columns denotes the relative frequency the marginal likelihood decided for a specific number of clusters. K true denotes the true number of clusters within the scenarios. IH denotes the inhomogeneity of the specified cluster specific parameters.

Table 5: Simulation Study - Model selection via Forecasting Criterion (T) - Panel size B

Scenario	1	2	3	4	MIH/GIH	MIH_{12}/GIH_{12}	MIH_{13}/GIH_{13}	MIH_{23}/GIH_{23}	K true	stratified
I	1.0000	0	0	0	0/–	–	–	–	1	no
II	0	0.9938	0	0.0062	1.325 / 0	1.325/ 0	–/–	–/–	2	no
III	0	0.9877	0.0123	0	2.250 /0.116	2.250/ 0.116	–/–	–/–	2	no
IV	0	0.9630	0.0309	0.0062	2.250 / 0.373	2.250/ 0.373	–/–	–/–	2	yes
V	0	0.9691	0.0309	0	2.270 / 0.1342	0 / 0.420	2.250/0.250	2.250/0.260	3	yes
VI	0	0.8704	0.1235	0.0062	1.251 /0	1.290 / 0	0.125 / 0	1.165 / 0	3	no
VII	0	0.7346	0.2284	0.0370	1.251 /0.097	1.290 / 0.123	0.125 /0.023	1.165 / 0.040	3	yes
$VIII$	0.3272	0.5000	0.1111	0.0617	0.240 /0	0.0150/0	0.0900 /0	0.1050 /0	3	no
IX	0.1049	0.0802	0.5617	0.2531	0.240 /0.097	0.045 / 0.123	0.270 / 0.023	0.315 /0.040	3	yes
X	0	0	0.9938	0.0062	1.500 /0	1.313/0	0.750/0	1.313 /0	3	no
XI	0	0	0.9630	0.0370	0.606 /0.1333	0.250 / 0.123	0.750 / 0.023	0.500 / 0.040	3	yes

Notes: Simulation of 500 data sets for each scenario I–XI. Scenarios are described in Section xx. For each scenario the columns denotes the relative frequency the marginal likelihood decided for a specific number of clusters. K true denotes the true number of clusters within the scenarios. IH denotes the inhomogeneity of the specified cluster specific parameters.

Table 6: Simulation Study - Model selection via ROC Graph - Panel size A

Scenario	1	2	3	4	MIH/GIH	MIH_{12}/GIH_{12}	MIH_{13}/GIH_{13}	MIH_{23}/GIH_{23}	K true	stratified
I	1.0000	0	0	0	0/−	−	−	−	1	no
II	0	1.0000	0	0	1.325 / 0	1.325/ 0	−/−	−/−	2	no
III	0	1.0000	0	0	2.250 /0.116	2.250/ 0.116	−/−	−/−	2	no
IV	0	1.0000	0	0	2.250 / 0.373	2.250/ 0.373	−/−	−/−	2	yes
V	0	0.9800	0.0200	0	2.270 / 0.1342	0 / 0.420	2.250/0.250	2.250/0.260	3	yes
VI	0	1.0000	0	0	1.251 /0	1.290 / 0	0.125 / 0	1.165 / 0	3	no
VII	0	0.9500	0.0450	0.0050	1.251 /0.097	1.290 / 0.123	0.125 /0.023	1.165 / 0.040	3	yes
$VIII$	0.2600	0.4550	0.1850	0.1000	0.240 /0	0.0150/0	0.0900 /0	0.1050 /0	3	no
IX	0.0950	0.0250	0.5800	0.3000	0.240 /0.097	0.045 / 0.123	0.270 / 0.023	0.315 /0.040	3	yes
X	0	0	0.9850	0.0150	1.500 /0	1.313/0	0.750/0	1.313 /0	3	no
XI	0.0100	0.0100	0.8850	0.0950	0.606 /0.1333	0.250 / 0.123	0.750 / 0.023	0.500 / 0.040	3	yes

Notes: Simulation of 500 data sets for each scenario I–XI. Scenarios are described in Section xx. For each scenario the columns denotes the relative frequency the marginal likelihood decided for a specific number of clusters. K true denotes the true number of clusters within the scenarios. IH denotes the inhomogeneity of the specified cluster specific parameters.

Table 7: Simulation Study - Model selection via ROC Graph - Panel size B

Scenario	1	2	3	4	MIH/GIH	MIH_{12}/GIH_{12}	MIH_{13}/GIH_{13}	MIH_{23}/GIH_{23}	K true	stratified
I	1.0000	0	0	0	0/–	–	–	–	1	no
II	0	1.0000	0	0	1.325 / 0	1.325/ 0	–/–	–/–	2	no
III	0	1.0000	0	0	2.250 /0.116	2.250/ 0.116	–/–	–/–	2	no
IV	0	1.0000	0	0	2.250 / 0.373	2.250/ 0.373	–/–	–/–	2	yes
V	0	0.9877	0.0123	0	2.270 / 0.1342	0 / 0.420	2.250/0.250	2.250/0.260	3	yes
VI	0	0.9753	0.0247	0	1.251 /0	1.290 / 0	0.125 / 0	1.165 / 0	3	no
VII	0	0.9074	0.0802	0.0123	1.251 /0.097	1.290 / 0.123	0.125 /0.023	1.165 / 0.040	3	yes
$VIII$	0.0679	0.5926	0.1852	0.1543	0.240 /0	0.0150/0	0.0900 /0	0.1050 /0	3	no
IX	0.0062	0.0988	0.6728	0.2222	0.240 /0.097	0.045 / 0.123	0.270 / 0.023	0.315 /0.040	3	yes
X	0	0	1.0000	0	1.500 /0	1.313/0	0.750/0	1.313 /0	3	no
XI	0	0.0185	0.9198	0.0617	0.606 /0.1333	0.250 / 0.123	0.750 / 0.023	0.500 / 0.040	3	yes

Notes: Simulation of 500 data sets for each scenario I–XI. Scenarios are described in Section xx. For each scenario the columns denotes the relative frequency the marginal likelihood decided for a specific number of clusters. K true denotes the true number of clusters within the scenarios. IH denotes the inhomogeneity of the specified cluster specific parameters.

Table 8: Model Comparison – Log Marginal Likelihood and Out-of-Sample Prediction

	Log Marginal Likelihood		ROC		Prediction	
	pooled		pooled		pooled	
	-4143.8		0.5383		0.6203	
	random coefficient – diagonal		random coefficient – diagonal		random coefficient – diagonal	
	-3608.4		0.7506		0.7378	
K	cluster	cluster–stratified	cluster	cluster–stratified	cluster	cluster–stratified
2	-3663.4	-3655.2	0.7208	0.7279	0.7509	0.7528
3	-3590.8	-3593.9	0.7762	0.7420	0.7549	0.7543
4	-3611.5	-3609.0	0.7526	0.7533	0.7542	0.7541

Table 9: Pooled Specification

	estimates			$\partial\mathrm{Pr}(y_{it}=1|\overline{x})/\partial\overline{x}$		
variable	mean	std	95% HDI	mean	std	95% HDI
constant	-1.9402	0.2354	$[-2.3997; -1.4775]$	–	–	–
log sales	0.1742	0.0227	$[0.1295; 0.2182]$	0.0671	0.0087	$[0.0500; 0.0840]$
rel size	1.0702	0.1380	$[0.8042; 1.3495]$	0.4122	0.0530	$[0.3099; 0.5197]$
imports	1.1339	0.1503	$[0.8429; 1.4298]$	0.4368	0.0579	$[0.3248; 0.5508]$
FDI	2.8039	0.4004	$[2.0304; 3.5987]$	1.0801	0.1541	$[0.7816; 1.3851]$
Prod.	-2.1993	0.7086	$[-3.5946; -0.8125]$	-0.8472	0.2729	$[-1.3858; -0.3131]$
Raw. Mtl.	-0.2875	0.0794	$[-0.4435; -0.1319]$	-0.1108	0.0306	$[-0.1709; -0.0509]$
Inv. good	0.1897	0.0389	$[0.1117; 0.2643]$	0.0731	0.0150	$[0.0430; 0.1020]$

Table 10: Random Coefficient Specification – diagonal

variable	mean	95% HDI	variance	95% HDI
constant	-3.1453 $_{0.6629}$	$[-4.4234; -1.8718]$	0.1622 $_{0.0534}$	$[0.0849; 0.2975]$
log sales	0.2687 $_{0.0648}$	$[0.1432; 0.3963]$	0.0160 $_{0.0012}$	$[0.0137; 0.0185]$
rel size	3.0470 $_{0.7205}$	$[1.7504; 4.2945]$	7.0640 $_{4.3072}$	$[0.9063; 17.5230]$
imports	1.8798 $_{0.3931}$	$[1.0837; 2.6329]$	0.4842 $_{0.2219}$	$[0.1913; 1.0648]$
FDI	2.7348 $_{1.0269}$	$[0.7468; 4.7695]$	1.5251 $_{2.0589}$	$[0.2649; 8.5134]$
Prod.	-2.0111 $_{1.7093}$	$[-5.4175; 1.3175]$	0.6597 $_{0.5082}$	$[0.1718; 2.1504]$
Raw. Mtl.	-0.5696 $_{0.2336}$	$[-1.0208; -0.1009]$	0.4461 $_{0.2043}$	$[0.1686; 0.9843]$
Inv. good	0.3956 $_{0.1179}$	$[0.1654; 0.6270]$	0.1979 $_{0.0632}$	$[0.0994; 0.3425]$
	marginal effect	95% HDI	marginal effect	95% HDI
–	$\widetilde{ME_{RC}}$		$\widetilde{ME_{RC}}$	
log sales	0.0565 $_{0.0147}$	$[0.0277; 0.0855]$	0.0531 $_{0.0141}$	$[0.0257; 0.0807]$
rel size	0.6829 $_{0.1557}$	$[0.3974; 0.9680]$	0.6496 $_{0.1490}$	$[0.3774; 0.9200]$
imports	0.4300 $_{0.0902}$	$[0.2481; 0.6036]$	0.4079 $_{0.0855}$	$[0.2354; 0.5722]$
FDI	0.6293 $_{0.2375}$	$[0.1708; 1.1050]$	0.5977 $_{0.2257}$	$[0.1620; 1.0469]$
Prod.	-0.4670 $_{0.3957}$	$[-1.2622; 0.3048]$	-0.4439 $_{0.3760}$	$[-1.2008; 0.2888]$
Raw. Mtl.	-0.1328 $_{0.0535}$	$[-0.2358; -0.0258]$	-0.1262 $_{0.0511}$	$[-0.2246; -0.0239]$
Inv. good	0.0878 $_{0.0267}$	$[0.0350; 0.1397]$	0.0833 $_{0.0257}$	$[0.0328; 0.1334]$

Table 11: Parameter Estimates Cluster Specification – $K = 3$

	$k = 1$		$k = 2$		$k = 3$	
variable	mean / std	95% HDI	mean / std	95% HDI	mean / std	95% HDI
constant	-2.0456 0.7034	$[-3.4881; -0.7273]$	-2.4621 0.7340	$[-3.8931; -1.0214]$	-6.3776 1.4934	$[-9.8281; -3.8364]$
log sales	0.2650 0.0706	$[0.1315; 0.4092]$	0.1661 0.0659	$[0.0353; 0.2937]$	0.3601 0.1227	$[0.1375; 0.6265]$
rel size	4.6454 1.0988	$[2.8024; 7.1186]$	0.8793 0.4529	$[0.1927; 2.0339]$	1.2375 0.6160	$[0.1941; 2.5098]$
imports	0.8145 0.4472	$[-0.0596; 1.6886]$	2.2386 0.5168	$[1.2560; 3.2895]$	1.5618 1.3235	$[-1.3657; 3.6694]$
FDI	1.1766 1.4908	$[-1.7619; 4.2059]$	2.2670 1.3647	$[0.1173; 5.7940]$	3.2779 2.9830	$[-3.6103; 7.9629]$
Prod.	-1.0736 2.6381	$[-6.0471; 4.0568]$	-1.6256 2.5424	$[-7.3004; 1.9024]$	0.1337 2.2981	$[-4.5534; 5.2717]$
Raw. Mtl.	-0.3785 0.2245	$[-0.8244; 0.0430]$	-0.9065 0.3909	$[-1.7077; -0.2057]$	0.6109 1.1467	$[-2.5092; 2.8301]$
Inv. good	0.1725 0.1274	$[-0.0830; 0.4233]$	0.4355 0.1291	$[0.1833; 0.6896]$	0.7598 0.5579	$[0.1884; 2.8471]$
$\gamma_1/\gamma_2/\gamma_3$	0.4597 0.0449	$[0.3687; 0.5441]$	0.3387 0.0385	$[0.2640; 0.4144]$	0.2016 0.0340	$[0.1316; 0.2666]$

Table 12: Marginal Effects in Population

	marginal effect	95% HDI
constant	–	–
log sales	0.0550 _{0.0183}	$[0.0193; 0.0898]$
rel size	0.1922 _{0.0955}	$[0.0362; 0.3861]$
imports	0.2712 _{0.1665}	$[-0.0654; 0.5680]$
FDI	0.6361 _{0.4553}	$[-0.3405; 1.4273]$
Prod.	0.0023 _{0.3433}	$[-0.6755; 0.8056]$
Raw. Mtl.	0.1102 _{0.1152}	$[-0.0944; 0.3345]$
Inv. good	0.0922 _{0.0330}	$[0.0293; 0.1581]$

Table 13: Cluster Specific Marginal Effects – $K = 3$

|	$k = 1$		$k = 2$		$k = 3$	
variable	mean / std	95% HDI	mean / std	95% HDI	mean / std	95% HDI
constant	−0.3157 0.1082	[−0.5234; −0.1017]	−0.9766 0.2861	[−1.5493; −0.4234]	−0.5672 0.2870	[−1.0147; −0.0500]
log sales	0.0406 0.0112	[0.0186; 0.0619]	0.0635 0.0252	[0.0147; 0.1129]	0.0320 0.0182	[0.0017; 0.0644]
rel size	0.7132 0.1190	[0.4770; 0.9391]	0.3442 0.1652	[0.0925; 0.7422]	0.1135 0.0790	[−0.0148; 0.2680]
imports	0.1280 0.0675	[−0.0010; 0.2660]	0.8921 0.1879	[0.5253; 1.2609]	0.1557 0.1347	[−0.0593; 0.4086]
FDI	0.1886 0.2135	[−0.2018; 0.6451]	0.8580 0.5052	[0.0491; 2.0094]	0.3023 0.3551	[−0.2702; 1.0138]
Prod.	−0.1921 0.4089	[−1.0165; 0.5783]	−0.5491 0.9490	[−2.7600; 0.7454]	0.0193 0.2293	[−0.4603; 0.5594]
Raw. Mtl.	−0.0562 0.0305	[−0.1179; 0.0043]	−0.3319 0.1430	[−0.6459; −0.0837]	0.0364 0.0959	[−0.1945; 0.2123]
Inv. good	0.0276 0.0187	[−0.0091; 0.0638]	0.1724 0.0478	[0.0787; 0.2660]	0.0614 0.0285	[0.0113; 0.1192]

Figures
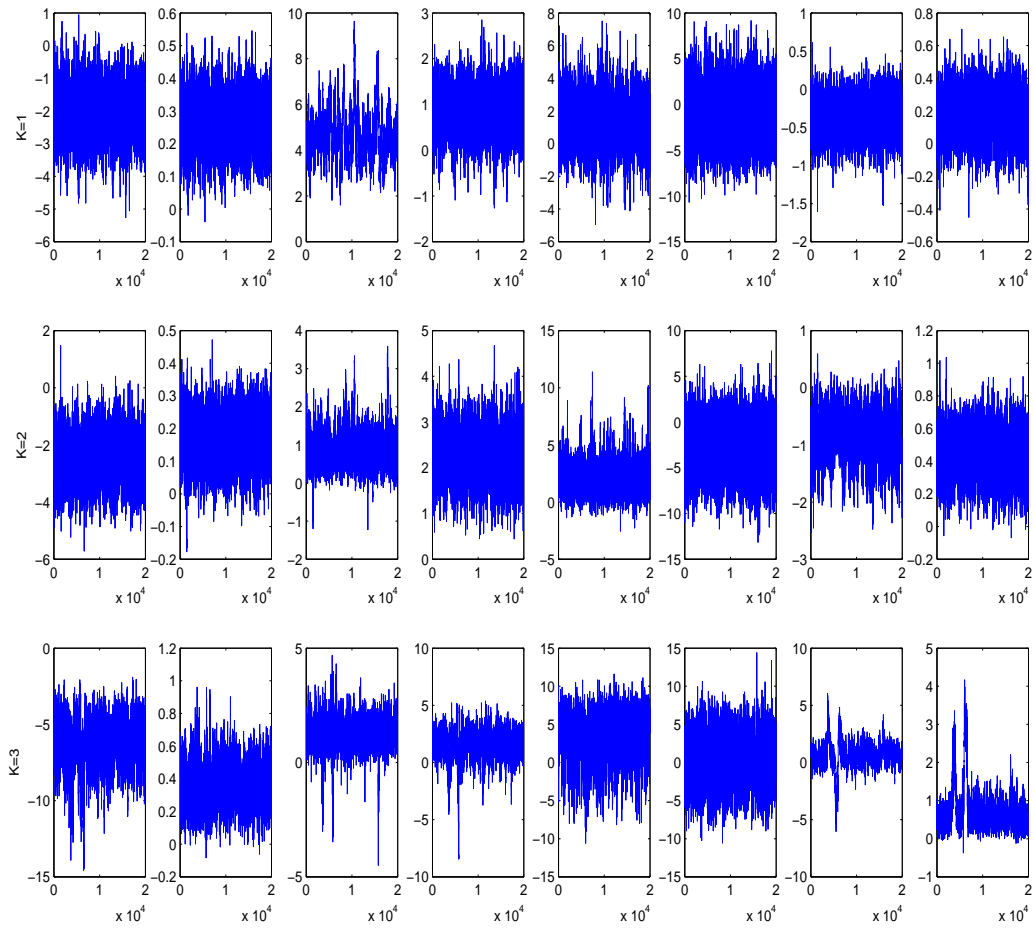
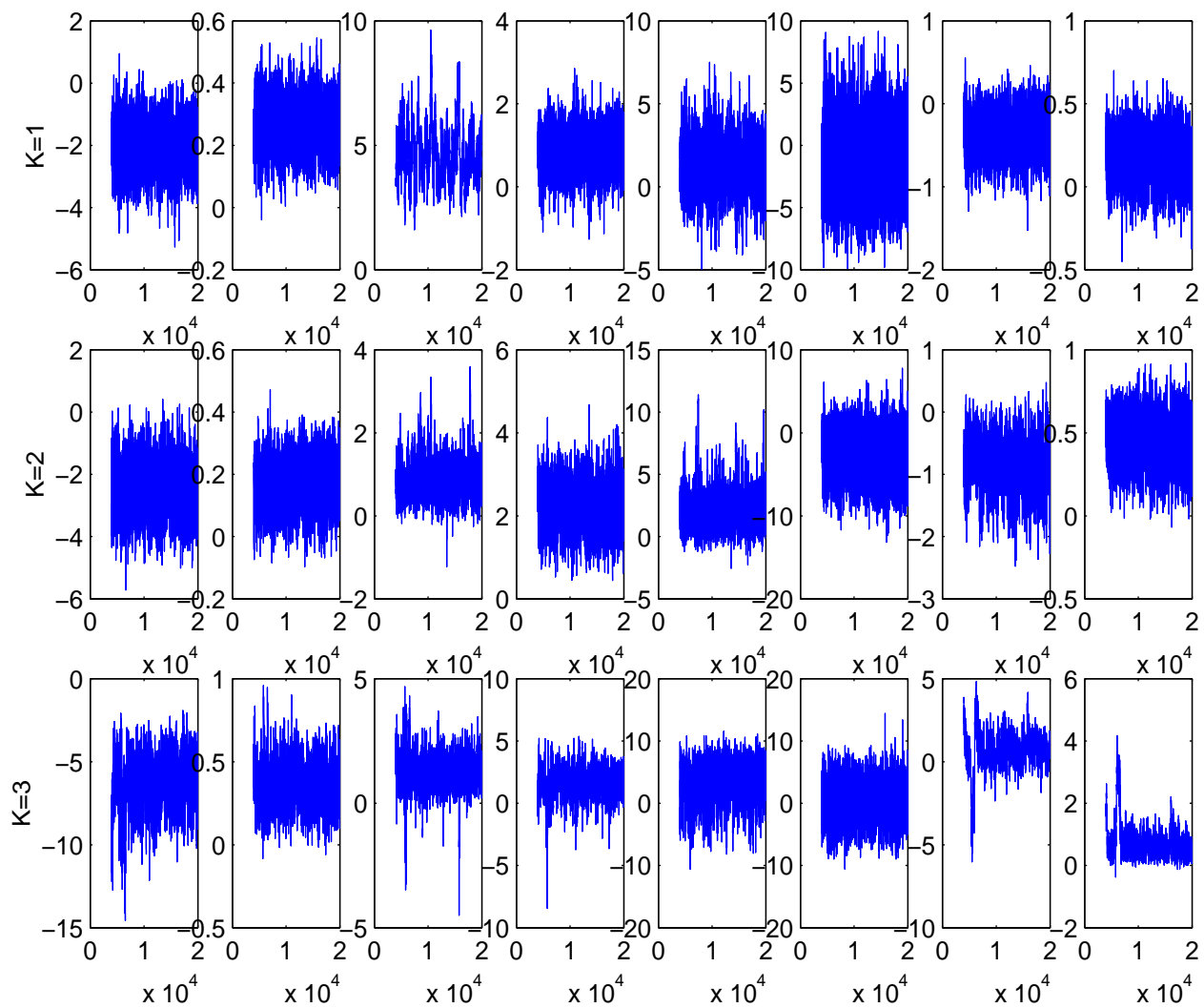Figure 1: MCMC output before relabeling - 3 clusters
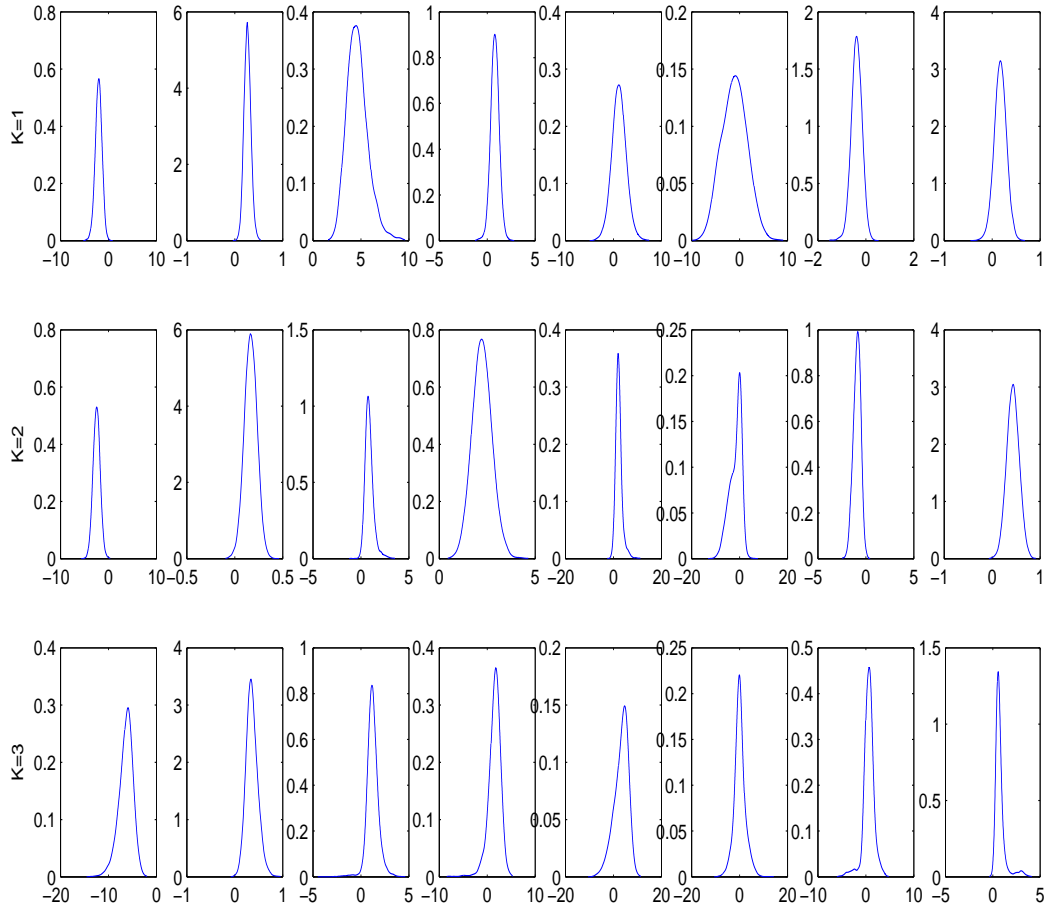
Figure 2: MCMC output after relabeling - 3 clusters

Figure 3: MCMC output after relabeling - 3 clusters