Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2008-38

Bayesian Estimation of Random Effects Models for Multivariate Responses of Mixed Data

Helga Wagner and Regina Tüchler

November 2008
Revised: August 2009

Abstract

A random effects model is presented to estimate multivariate data of mixed data types. Such data typically appear in studies where different response variables are measured repeatedly for one subject. It is possible to relate normal, binary, multinomial and count data by our joint model. Further flexibility with respect to model specification is obtained by including modern variable selection techniques. Auxiliary mixture sampling leads to a Gibbs sampling type scheme which is easy to implement since no additional tuning is needed. The method is illustrated by transaction data of a costumer cohort acquired by an apparel retailer.

Keywords: Auxiliary mixture sampling; Generalized linear models; MCMC; Random effects model; Variable Selection

1 Introduction

In this paper we model multidimensional data which arise when different response variables are measured repeatedly for one subject. Usually these responses are not of the same type but are measured on different scales, yielding mixed data with continuous and discrete outcomes. Since measurements are taken repeatedly over time on each subject under study not only dependencies between the response components but also within-subject dependencies have to be taken into account. For repeated measurements of a single data type the usual approach is to use linear random effects models for normal or general random effects models for discrete data. However, combination of different data types to a joint model is a challenging problem. In the present paper we specify a random effects model which combines normal, binary, multinomial and count outcomes. We account for within subject dependencies by defining a random effects specification for the linear predictors of the single data types. These single response types are then linked by adding covariances between random effects of the different data types.

Such a general model for mixed data was not estimated in the literature before. This is mainly due to computational difficulties which arise when combining different data types. Clustered data of mixed type received attention in particular for a binary and a normal response component in the context of toxicity studies (Fitzmaurice and Laird, 1995; Catalano and Ryan, 1992; Regan and Catalano, 1999b,a). One approach is to model the joint distribution of both outcomes as the product of a marginal and a conditional distribution, see Cox and Wermuth (1992) for a discussion of different factorizations. Correlation of repeated measurements for one subject is taken into account in the marginal model as well as in the conditional model, estimation is accomplished by generalized estimation equations. The same type of approach is taken by Yang, Jian, and Zhang (2007) for bivariate longitudinal data where one component is continuous and the other is Poisson count. Within subject correlation is taken into account for each response type by assuming a compound symmetry covariance matrix for observations of one subject. A different modeling approach, taken in Regan and Catalano (1999b); Gueorguieva and Agresti (2001) and Faes, Aerts, Molenberghs, Geys, Teuns, and Bijnens (2008), is based on the interpretation of binary response as a dichotomization of an underlying normal variable and assuming a bivariate normal distribution for the normal response and the underlying normal variable. Correlation between the two responses and intra-cluster or within subject correlation can be taken into account by either explicit modeling of the covariance structure as in Regan and Catalano (1999b) or by a random effects specification where random effects and/or errors are assumed to follow a general bivariate normal distribution as in Gueorguieva and Agresti (2001). In principle this approach allows a full random effects specification for multivariate responses, however due to computational aspects so far researchers focused their work on simplified models. Faes et al. (2008) consider this problem in a classical setting and use pseudo-likelihood for joint estimation of all pairwise bivariate generalized linear mixed models.

In our paper we estimate a full random effects model. By using data augmentation we combine not only the normal responses but also the discrete ones to a linear model. The novel method of auxiliary mixture sampling then leads to a Gibbs sampling type scheme. Until recently Bayesian estimation of generalized linear models for categorical or count data was only possible if Metropolis-Hastings steps were included. Auxiliary mixture sampling for single data types was developed in Frühwirth-Schnatter and Wagner (2006) and Frühwirth-Schnatter, Frühwirth, Held, and Rue (2009) for Poisson counts, and in Frühwirth-Schnatter and Frühwirth (2007) for binomial and multinomial responses.

With many covariates at hand specification of random and fixed effects is a complicated problem. Recently variable selection tools are used to solve such model selection problems, see e.g. George and McCulloch (1997) for a description of the stochastic search variable approach, Smith and Kohn (2002) for covariance selection for normal data, and Frühwirth-Schnatter and Tüchler (2008) and Tüchler (2008) for covariance selection in normal and logistic random effects models, respectively. In our paper variable and covariance selection enable us to start with a very general model specification. All predictor variables at hand may be included and all effects may be specified as random effects. During the course of MCMC sampling those effects with zero means are detected and those effects which are fixed rather than random are restricted to fixed effects. Since the different data types are related through the variance-covariance matrix covariance selection also reveals whether such a relationship is present or not. If all covariances between effects of certain data types were selected as zero no relation between these data types would be present and the joint model would split into separate models.

The paper is structured as follows. In Section 2 we define the model. It is transformed into a Gaussian random effects model by auxiliary mixture sampling in Section 3.1 and variable and covariance selection is incorporated in Section 3.2. The prior and the simulation steps are described in Sections 3.3 and 3.4, respectively. The method is applied to simulated data in Section 4, and Section 5 gives a real-data example. Section 6 summarizes the results.

2 Random Effects Model for Mixed Data

Let $\mathbf{Y} = (Y^1, \dots, Y^K)'$ denote a multivariate response variable which is observed for $i = 1, \dots, N$ subjects on $t = 1, \dots, T_i$ occasions. The components $Y^k, k = 1, \dots, K$

may be either normal, binary, multinomial or Poisson counts. Let y_{it}^k denote the observation of the k-th component measured for subject i at time point t, let $\mathbf{y}_i^k$ denote the sequence of T_i observations for the k-th component of subject i, and let $\mathbf{y}_i$ summarize all $T_i\,K$ observations of subject i.

To relate the mean $\mu_{it}^k = E(y_{it}^k)$ to the linear predictor η_{it}^k we introduce a distinct link function $g_k(\mu_{it}^k) = \eta_{it}^k, k = 1, \dots, K$ for each component depending on the type of the k-th response component. For Poisson components we use the log-link-function

$$\mu_{it}^k = \exp(\eta_{it}^k),$$

for binary components we consider the logit link function

$$\mu_{it}^k = \frac{\exp(\eta_{it}^k)}{1 + \exp(\eta_{it}^k)},$$

whereas for normal components y_{it}^k we use the identical link

$$\mu_{it}^k = \eta_{it}^k$$

and assume a constant variance $y_{it}^k \sim \mathcal{N}\left(\mu_{it}^k, \sigma_k^2\right)$.

We consider the following random effects specification for the linear predictors $\boldsymbol{\eta}_i^k$ of the sequence $\mathbf{y}_i^k$:

$$\boldsymbol{\eta}_i^k = \mathbf{X}_i \boldsymbol{\beta}_i^k.$$

$\mathbf{X}_i$ is a design matrix of dimension $T_i \times d$, where d equals the number of covariates in the model. $\boldsymbol{\beta}_i^k$ are normally distributed random effects. We assume that the same covariates are used for each of the K response components, whereas the random effects are allowed to differ between components.

Dependency between repeated measurements is described by the random effects $\boldsymbol{\beta}_i^k$ shared for all measurements of one response component. To take into account dependency between the components we assume that the random effects $\boldsymbol{\beta}_i = ((\boldsymbol{\beta}_i^1)', \dots, (\boldsymbol{\beta}_i^K)')'$ of one subject follow a multivariate normal distribution

$$\boldsymbol{\beta}_i \sim \mathcal{N}_{dK}\left(\boldsymbol{\beta}, \mathbf{Q}\right),$$

with mean $\boldsymbol{\beta}$ and variance-covariance matrix $\mathbf{Q}$. Note that assuming pairwise independence between the random effects

$$Cov(\boldsymbol{\beta}_i^k, \boldsymbol{\beta}_i^{k'}) = \mathbf{0} \quad \text{for } k \neq k'; i = 1, \dots, N,$$

would correspond to separate modeling of each of the K components using linear random effects models for the normal and generalized random effects models for the discrete responses.

3 Inference Procedure

3.1 Data Augmentation

3.1.1 The Augmented Model

Bayesian estimation of the model defined in Section 2 can be performed by a simple Gibbs sampler as long as all response components are Gaussian. For discrete outcomes auxiliary mixture sampling leads to an augmented Gaussian model for which a Gibbs sampling scheme is available. Auxiliary mixture sampling for Poisson counts is developed in Frühwirth-Schnatter and Wagner (2006) and Frühwirth-Schnatter et al. (2009), whereas binary and multinomial logit models are estimated in Frühwirth-Schnatter and Frühwirth (2007).

Auxiliary mixture sampling is based on data augmentation and allows a model representation as a linear Gaussian random effects model in the auxiliary variables $\tilde{\mathbf{y}}_i^k$:

$$\tilde{\mathbf{y}}_i^k = \tilde{\mathbf{X}}_i^k \boldsymbol{\beta}_i^k + \tilde{\boldsymbol{\varepsilon}}_i^k, \quad k = 1, \dots K \tag{1}$$

where the error term $\tilde{\boldsymbol{\varepsilon}}_i^k$ is distributed as $\mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_i^k\right)$ with a diagonal matrix $\boldsymbol{\Sigma}_i^k$. The joint model for all outcome components is obtained by combining the separate models (1) to the following Gaussian random effects model:

$$\tilde{\mathbf{y}}_i = \begin{pmatrix} \tilde{\mathbf{y}}_i^1 \\ \vdots \\ \tilde{\mathbf{y}}_i^K \end{pmatrix} = \tilde{\mathbf{X}}_i \boldsymbol{\beta}_i + \tilde{\boldsymbol{\varepsilon}}_i, \quad \tilde{\boldsymbol{\varepsilon}}_i \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_i\right), \tag{2}$$

$$\boldsymbol{\beta}_i \sim \mathcal{N}_{dK}\left(\boldsymbol{\beta}, \mathbf{Q}\right), \tag{3}$$

where $\tilde{\mathbf{X}}_i$ and $\boldsymbol{\Sigma}_i$ are block diagonal matrices with entries $\tilde{\mathbf{X}}_i^1, \dots, \tilde{\mathbf{X}}_i^K$ and $\boldsymbol{\Sigma}_i^1, \dots, \boldsymbol{\Sigma}_i^K$, respectively. This linear random effects model implies the marginal model

$$\tilde{\mathbf{y}}_i \sim \mathcal{N}\left(\tilde{\mathbf{X}}_i \boldsymbol{\beta}, \tilde{\mathbf{X}}_i \mathbf{Q} \tilde{\mathbf{X}}_i' + \boldsymbol{\Sigma}_i\right).$$

We are now going to discuss the derivation of model (1) for the various data types. For ease of exposition we use superscripts n, b and c to indicate normal, binary and count components, respectively.

For a normal response Y^n no data augmentation is needed and hence $\tilde{\mathbf{y}}_i^n = \mathbf{y}_i^n, \tilde{\mathbf{X}}_i^n = \mathbf{X}_i$ and $\boldsymbol{\Sigma}_i^n = \sigma_n^2 \mathbf{I}$. Data augmentation is needed for the discrete data types and we describe details in the following two subsections.

3.1.2 Data augmentation for binary and multinomial components

For a binary component Y^b with outcomes $y_{it}^b \in \{0, 1\}$ auxiliary mixture sampling is based on the interpretation of logit models in terms of utilities as in McFadden (1974). Let $u_{it,0}$ be the utility of choosing category 0 and u_{it} be the utility of choosing category 1, which is modeled as

$$u_{it} = \mathbf{x}_{it} \boldsymbol{\beta}_i^b + \varepsilon_{it}^b. \tag{4}$$

Then $y_{it}^b = 1$, iff $u_{it} > u_{it,0}$, and $y_{it}^b = 0$ otherwise. The binary logit random effects model results as the marginal distribution of $\mathbf{y}_i^b$, if $u_{it,0}$ and ε_{it}^b follow a type I extreme value distribution, see Scott (2009). The latent utilities are introduced as missing variables in a first data augmentation step. The extreme value distribution of ε_{it}^b can be approximated very accurately by a mixture of ten normal components

$$p_\varepsilon(\varepsilon) = \exp(-\varepsilon - e^{-\varepsilon}) \approx \sum_{r=1}^{10} w_r f_N(\varepsilon; m_r, s_r^2),$$

where the weights w_r, the means m_r, and the variances s_r^2, $r = 1, \ldots, 10$, have been determined numerically by minimizing the Kullback-Leibler distance between the density of the type I extreme value distribution and the mixture approximation, see Frühwirth-Schnatter and Frühwirth (2007) for more details. As the weights, means and variances (w_r,m_r,s_r^2) are fixed numbers rather than unknown parameters, only the component indicators $r_{it}^b \in \{1, \ldots, 10\}$ have to be introduced for each utility u_{it} in the second data augmentation step to obtain the linear Gaussian model

$$u_{it} = \mathbf{x}_{it}\boldsymbol{\beta}_i^b + m_{r_{it}^b} + \tilde{\varepsilon}_{r_{it}^b}, \quad \tilde{\varepsilon}_{r_{it}^b} \sim \mathcal{N}\left(0, s_{r_{it}^b}^2\right).$$

Conditional on the component indicators we define the auxiliary variables $\tilde{y}_{it}^b = u_{it} - m_{r_{it}^b}$ and stack the elements $\tilde{y}_{it}^b$ for each subject to obtain the vector $\tilde{\mathbf{y}}_i^b$. The model for this auxiliary response vector is the linear random effects model

$$\tilde{\mathbf{y}}_i^b = \tilde{\mathbf{X}}_i^b\boldsymbol{\beta}_i^b + \tilde{\boldsymbol{\varepsilon}}_i^b,$$

where $\tilde{\boldsymbol{\varepsilon}}_i^b \sim \mathcal{N}_{T_i}\left(\mathbf{0}, \boldsymbol{\Sigma}_i^b\right)$, $\boldsymbol{\Sigma}_i^b$ is a diagonal matrix with elements $s_{r_{it}^b}^2$, and $\tilde{\mathbf{X}}_i^b = \mathbf{X}_i$.

Extension to a multinomial component Y^m where y_{it}^m takes a value in one of $L+1$ unordered categories is straightforward. For each observation y_{it}^m, however $L+1$ latent utilities $(u_{it,0}^m, u_{it,1}^m, \ldots, u_{it,L}^m)$ have to be introduced as missing data in the first data augmentation step, see Frühwirth-Schnatter and Frühwirth (2007).

3.1.3 Data augmentation for Poisson components

For a count response Y^c data augmentation is based on the interpretation of a Poisson count y_{it}^c as the number of jumps of an unobserved Poisson process with intensity μ_{it}^c in the time interval [0,1], see Frühwirth-Schnatter and Wagner (2006) and Frühwirth-Schnatter et al. (2009). In the first data augmentation step the *inter-arrival* time between the last jump before and the first jump after 1, denoted $\tau_{it,1}^c$, is introduced. For observations $y_{it}^c > 0$ it is required to add the *arrival* time of the last jump before 1, denoted by $\tau_{it,2}^c$, as a further latent variable.

As $\tau_{it,1}^c$ follows an exponential distribution $\mathcal{E}\left(\mu_{it}^c\right)$ and $\tau_{it,2}^c$ follows a Gamma distribution $\Gamma(y_{it}^c, \mu_{it}^c)$, the original Poisson regression model can be transformed into the linear model

$$-\log \tau_{it,1}^c = \mathbf{x}_{it}\boldsymbol{\beta}_i^c + \varepsilon_{it,1}^c, \tag{5}$$
$$-\log \tau_{it,2}^c = \mathbf{x}_{it}\boldsymbol{\beta}_i^c + \varepsilon_{it,2}^c, \tag{6}$$

where the distribution of $\varepsilon_{it,1}^c$ is a type I extreme value distribution and $\varepsilon_{it,2}^c$ is distributed as the negative logarithm of a Gamma random variable with integer shape parameter $\nu = y_{it}^c$. For $y_{it}^c = 0$ we are dealing only with $\tau_{it,1}^c$. The non-normal densities of $\varepsilon_{it,1}^c$ and $\varepsilon_{it,2}^c$ can be approximated by a mixture of normal components

$$p_\varepsilon(\varepsilon; \nu) = \frac{\exp(-\nu\varepsilon - e^{-\varepsilon})}{\Gamma(\nu)} \approx \sum_{r=1}^{R(\nu)} w_r(\nu) f_N(\varepsilon; m_r(\nu), s_r^2(\nu)).$$

The mixture approximation was derived numerically for integer values of ν, see Frühwirth-Schnatter et al. (2009) for details. The number of components $R(\nu)$

needed to obtain an accurate approximation depends on ν. The weights $w_r(\nu)$, means $m_r(\nu)$ and variances $s_r^2(\nu)$ depend on ν as well, and are fixed. Therefore only the component indicators $r_{it,j}^c$ have to be introduced for each auxiliary observation in the second data augmentation step. Conditional on the auxiliary variables, the Poisson model reduces to the linear Gaussian model

$$-\log \tau_{it,j}^c = \mathbf{x}_{it}\boldsymbol{\beta}_i^c + m_{r_{it,j}^c} + \tilde{\boldsymbol{\varepsilon}}_{r_{it,j}^c}, \quad \tilde{\boldsymbol{\varepsilon}}_{r_{it,j}^c} \sim N(0, s_{r_{it,j}^c}^2).$$

We stack the auxiliary responses $\tilde{y}_{it,j}^c = -\log \tau_{it,j}^c - m_{r_{it,j}^c}$ for each subject to obtain vector $\tilde{\mathbf{y}}_i^c$. The length of $\tilde{\mathbf{y}}_i^c$ is no longer T_i as for the original response $\mathbf{y}_i^c$ but $T_i + n_i$, where n_i is the number of nonzero counts for subject i. The conditional Gaussian random effects model is given by

$$\tilde{\mathbf{y}}_i^c = \tilde{\mathbf{X}}_i^c \boldsymbol{\beta}_i^c + \tilde{\boldsymbol{\varepsilon}}_i^c, \quad \tilde{\boldsymbol{\varepsilon}}_i^c \sim \mathcal{N}_{T_i+n_i}\left(\mathbf{0}, \boldsymbol{\Sigma}_i^c\right),$$

where $\tilde{\mathbf{X}}_i^c$ is chosen to match $\tilde{\mathbf{y}}_i^c$ and $\boldsymbol{\Sigma}_i^c$ is a diagonal matrix with elements equal to the variances $s_{r_{it,j}^c}^2$.

3.2 Model Selection

We are now going to include variable selection with respect to elements of the random effects mean $\boldsymbol{\beta}$ and covariance selection with respect to elements of the variance-covariance matrix $\mathbf{Q}$. A zero diagonal element in $\mathbf{Q}$ renders the corresponding effect a fixed effect and all off-diagonal elements are automatically identified as zero by the algorithm. A zero off-diagonal element indicates that no correlation is present between two random effects. Covariances between effects of different data types build the link between the different response types. Selecting solely zero covariances between different data types reduces the joint model to separate models each for one data type.

The random effects model (2), (3) is specified in the *centered* parameterization. The mean and variance-covariance matrix of the random effects appear in the latent equation. To carry out covariance selection we apply the Cholesky decomposition with lower-triangular Cholesky factors $\mathbf{C}$ to the variance-covariance matrix, $\mathbf{Q} = \mathbf{CC}'$, and rewrite the model in the equivalent *non-centered* parameterization:

$$\tilde{\mathbf{y}}_i = \begin{pmatrix} \tilde{\mathbf{y}}_i^1 \\ \vdots \\ \tilde{\mathbf{y}}_i^K \end{pmatrix} = \tilde{\mathbf{X}}_i \boldsymbol{\beta} + \tilde{\mathbf{X}}_i \mathbf{C} \boldsymbol{z}_i + \tilde{\varepsilon}_i, \quad \tilde{\varepsilon}_i \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_i\right), \tag{7}$$

$$\boldsymbol{z}_i \sim \mathcal{N}_{dK}\left(\mathbf{0}, \mathbf{I}\right), \tag{8}$$

where $\boldsymbol{\beta}_i = \mathbf{C}\boldsymbol{z}_i$, see Meng and van Dyk (1998).

To select variables in $\boldsymbol{\beta}$ we define for each element β_g of $\boldsymbol{\beta}$ an indicator δ_g, which takes the value 0 if element β_g is restricted to 0, whereas $\delta_g = 1$ indicates that β_g is unrestricted. The vector $\boldsymbol{\delta}$ consists of all dK indicators. We include only the unrestricted elements in vector $\boldsymbol{\beta}^{\boldsymbol{\delta}}$ and denote the corresponding design matrix $\tilde{\mathbf{X}}_i^{\boldsymbol{\delta}}$.

To carry out covariance-selection we follow the ideas of Frühwirth-Schnatter and Tüchler (2008). We stack the columns of the lower-triangular matrix $\mathbf{C}$ to obtain a vector of regression coefficients with design matrix $\mathbf{W}_i$. $\mathbf{W}_i$ is constructed by combining the individual effects $\boldsymbol{z}_i$ and the matrix $\tilde{\mathbf{X}}_i$, see the Appendix for details. We define an indicator vector $\boldsymbol{\gamma}$ of dimension $dK(dK+1)/2$ to select restricted and unrestricted elements in $\mathbf{C}$. We denote the vector of all unrestricted elements $\mathbf{C}^{\gamma}$ and its design matrix $\mathbf{W}_i^{\gamma}$.

The augmented model for variable and covariance selection reads:

$$\tilde{\mathbf{y}}_i = \begin{pmatrix} \tilde{\mathbf{y}}_i^1 \\ \vdots \\ \tilde{\mathbf{y}}_i^K \end{pmatrix} = \tilde{\mathbf{X}}_i^{\boldsymbol{\delta}} \boldsymbol{\beta}^{\boldsymbol{\delta}} + \mathbf{W}_i^{\gamma} \mathbf{C}^{\gamma} + \tilde{\boldsymbol{\varepsilon}}_i, \quad \tilde{\boldsymbol{\varepsilon}}_i \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_i\right). \tag{9}$$

The vector of all individual data vectors is denoted $\tilde{\mathbf{y}}$, and the vector of all individual effects is denoted $\boldsymbol{z}$. The diagonal matrix of all individual model error variances equals $\boldsymbol{\Sigma}$, and all individual design matrices are subsumed under $\tilde{\mathbf{X}}^{\boldsymbol{\delta}}$ and $\mathbf{W}^{\gamma}$, respectively.

The likelihood for the augmented model is given as

$$l(\tilde{\mathbf{y}}|\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \boldsymbol{\beta}, \mathbf{C}, \boldsymbol{z}, \mathbf{R}) \propto \prod_{i=1}^{N} \frac{1}{|\boldsymbol{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2} (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i^{\boldsymbol{\delta}} \boldsymbol{\beta}^{\boldsymbol{\delta}} - \mathbf{W}_i^{\gamma} \mathbf{C}^{\gamma})' \boldsymbol{\Sigma}_i^{-1} (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i^{\boldsymbol{\delta}} \boldsymbol{\beta}^{\boldsymbol{\delta}} - \mathbf{W}_i^{\gamma} \mathbf{C}^{\gamma}) \right).$$

3.3 Prior

The prior for the indicator vector $\boldsymbol{\delta}$ equals the following Beta function:

$$p(\boldsymbol{\delta}) = \text{Beta}(p_{\boldsymbol{\delta}} + 1, dK - p_{\boldsymbol{\delta}} + 1), \tag{10}$$

where $p_{\boldsymbol{\delta}}$ is the number of non-zero parameters in $\boldsymbol{\beta}$. This implies a prior dependence between the elements of the vector $\boldsymbol{\delta}$, see Smith and Kohn (2002).

The prior for vector $\boldsymbol{\gamma}$ is constructed in the same way and reads:

$$p(\boldsymbol{\gamma}) = \text{Beta}(p_{\boldsymbol{\gamma}} + 1, dK(dK+1)/2 - p_{\boldsymbol{\gamma}} + 1), \tag{11}$$

for $dK(dK+1)/2$ free elements and $p_{\boldsymbol{\gamma}}$ non-zero elements in the lower-triangular matrix $\mathbf{C}$.

Given the indicators $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ we specify a fractional prior with fraction $b = \sum_{i=1}^{N} T_i$ for the joint parameter vector of the unrestricted means and Cholesky factors, see Frühwirth-Schnatter and Tüchler (2008):

$$p(\boldsymbol{\beta}^{\boldsymbol{\delta}}, \mathbf{C}^{\gamma}|\boldsymbol{z}, \boldsymbol{\Sigma}, \tilde{\mathbf{y}}^b) = \mathcal{N}\left(\mathbf{a}_N, \mathbf{A}_N \frac{1}{b}\right), \tag{12}$$

where

$$\mathbf{A}_N^{-1} = [\mathbf{X}^{\boldsymbol{\delta}} \; \mathbf{W}^{\gamma}]' \boldsymbol{\Sigma}^{-1} [\mathbf{X}^{\boldsymbol{\delta}} \; \mathbf{W}^{\gamma}], \tag{13}$$

$$\mathbf{a}_N = \mathbf{A}_N [\mathbf{X}^{\boldsymbol{\delta}} \; \mathbf{W}^{\gamma}]' \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{y}}. \tag{14}$$

If the k-th component is normal, the prior of its specific error variance is the usual inverted Gamma prior:

$$\sigma_k^2 \sim \mathcal{G}^{-1}\left(c_{0k}/2, C_{0k}/2\right).$$

We use $\boldsymbol{\sigma}^2$ to address the collection of the error variances of all normal components.

3.4 MCMC Sampling scheme

An MCMC scheme to sample the model parameters $\boldsymbol{\beta}^{\boldsymbol{\delta}}$ and $\mathbf{C}^{\boldsymbol{\gamma}}$, the indicators $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$, the individual effects $\boldsymbol{z}$, the observation error variance for each normal component σ_k^2, and the augmented data $\tilde{\mathbf{y}}$ and $\mathbf{R} = (\mathbf{R}_1, \ldots, \mathbf{R}_N)$ (where $\mathbf{R}_i = (r_{it}^k), t = 1, \ldots, T_i; k = 1, \ldots, K$) is easy to implement as it requires only draws from standard densities. We give details of these steps in the Appendix.

(i) Perform the data augmentation steps described in Section 3.1 for binary and count components of the response vector to obtain $\tilde{\mathbf{y}}$ and $\mathbf{R}$.

(ii) Sample each element δ_g of the indicator vector $\boldsymbol{\delta}$ separately conditional on $\boldsymbol{\delta}_{\setminus g}$ (all other elements of $\boldsymbol{\delta}$), $\boldsymbol{\gamma}$, $\boldsymbol{z}$, $\boldsymbol{\sigma}^2$, $\mathbf{R}$ and $\tilde{\mathbf{y}}$.

(iii) Sample each element γ_h of the indicator vector $\boldsymbol{\gamma}$ separately conditional on $\boldsymbol{\gamma}_{\setminus h}$ (all other elements of $\boldsymbol{\gamma}$), $\boldsymbol{\delta}$, $\boldsymbol{z}$, $\boldsymbol{\sigma}^2$, $\mathbf{R}$ and $\tilde{\mathbf{y}}$.

(iv) Sample the non-zero elements $\boldsymbol{\beta}^{\boldsymbol{\delta}}$ and $\mathbf{C}^{\boldsymbol{\gamma}}$ together in one block conditional on $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$, $\boldsymbol{z}$, $\boldsymbol{\sigma}^2$, $\mathbf{R}$ and $\tilde{\mathbf{y}}$ from a multivariate normal distribution.

(v) Sample the individual effects $\boldsymbol{z}$ conditional on $\boldsymbol{\beta}^{\boldsymbol{\delta}}$, $\mathbf{C}^{\boldsymbol{\gamma}}$, $\boldsymbol{\sigma}^2$, $\mathbf{R}$ and $\tilde{\mathbf{y}}$ from multivariate normal distributions.

(vi) Sample the model error variance σ_k^2 for each normal response conditional on $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$, $\boldsymbol{z}$ and $\mathbf{y}^k$ from an inverted Gamma distribution.

4 Simulation example

We generated data for a normal and a Poisson component for N=300 subjects at T=6 time-points and 2 covariates x_1 and x_2, where $x_1 \sim U[-1, 1]$ and x_2 is a binary variable with $P(x_2 = 1) = 0.5$. The vector of regression coefficients, which includes also an intercept for the normal as well as the Poisson component was set to $\boldsymbol{\beta} = (3, 0, -0.3, 0.5, 0.2, 0)$ and $\sigma^2 = 0.5$. For the random effects variance-covariance matrix we used

$$\mathbf{Q} = \begin{pmatrix} 0.16 & 0 & -0.12 & 0.12 & 0 & -0.12 \\ 0 & 0.16 & 0 & 0 & 0.12 & 0 \\ -0.12 & 0 & 0.25 & -0.09 & 0 & 0.17 \\ 0.12 & 0 & -0.09 & 0.09 & 0 & -0.09 \\ 0 & 0.12 & 0 & 0 & 0.09 & 0 \\ -0.12 & 0 & 0.17 & -0.09 & 0 & 0.13 \end{pmatrix}.$$

$\mathbf{Q}$ has a sparse structure with high correlations between each random effect for the normal component and the respective random effect for the Poisson component. This is a situation where joint modeling of the two components in combination with model selection is particularly useful. As pointed out in Section 2 a block-diagonal structure of $\mathbf{Q}$ with no correlation between normal and Poisson component would result in separate random effects models for each component. Jointly modeling both components should allow to reveal the correct correlation structure in this situation.

Using the fractional prior for model selection only the parameters for the inverted Gamma prior for the model error variance of the normal component have to be chosen. We use the uninformative prior $\mathcal{G}^{-1}(0,0)$, i.e. $p(\sigma^2) \propto 1/\sigma^2$. Implementation of the MCMC scheme described was carried out in MATLAB (Version 7.2.0). The sampler was run for 20,000 iterations after a burn-in of 10,000. The first 1,000 draws of the burnin were drawn from the unrestricted model. 5 chains with different starting values were run and found to converge very quickly.

In Table 1 we give posterior estimates and HPD-regions of the mean parameters $\boldsymbol{\beta}$ for the normal and for the count part. The last column equals the posterior probabilities of the mean parameters to be unrestricted. These probabilities are obtained as posterior means of the indicators δ_g.

Table 1: Mixed Model: estimates, 95% HPD intervals and probabilities to be unrestricted for $\boldsymbol{\beta}$.

variable	$\hat{\boldsymbol{\beta}}$	HPD($\hat{\boldsymbol{\beta}}$)	$\Pr(\delta_g = 1)$
	normal response		
intercept	2.93	(2.87, 3.00)	*1.00*
x_1	0.00	(-0.01, 0.05)	*0.11*
x_2	-0.24	(-0.33, -0.16)	*1.00*
σ^2	0.47	(0.43, 0.51)	
	count response		
intercept	0.47	(0.41, 0.52)	*1.00*
x_1	0.12	(0.05, 0.20)	*0.98*
x_2	-0.01	(-0.09, 0.03)	*0.23*

Posterior estimates of the Cholesky factor $\mathbf{C}$ and the posterior probabilities of each element of $\mathbf{C}$ to be unrestricted are reported in Table 2. We follow Frühwirth-Schnatter and Tüchler (2008) and interpret the pattern of restricted elements in $\mathbf{C}$: The rank of $\mathbf{Q}$ equals three on average. All effects are random and rank reduction is caused by linear dependence among these random effects.

The variance-covariance matrix may be derived from the Cholesky factor $\mathbf{C}$ and the identity $\mathbf{Q} = \mathbf{C}\mathbf{C}'$. Its estimates together with posterior probabilities to be unrestricted are given in Table 3. We observe that the pattern of our data generating $\mathbf{Q}$ is reproduced very well. The nonzero elements of $\mathbf{Q}$ are unrestricted with probability 1, whereas the zero elements have a probability smaller than 0.5 to be included in the model.

To evaluate the sensitivity with respect to the parameters of inverse Gamma prior for σ^2, we repeated the analysis with parameters $c_0 = 0.001, C_0 = 0.001$, $c_0 = 0.01, C_0 = 0.01$, $c_0 = 0.1, C_0 = 0.1$ and $c_0 = 0.5, C_0 = 2$ and found essentially the same results. To illustrate the benefit of a joint model we fitted separate models for the normal and count response. Results are given in Table 4 and Table 5, respectively.

The mean structure is similar to that of the joint model, except for the effect of x_1 on the count response which is included in the model only in 41.4% of the draws. Substantial discrepancy however can be seen for the covariance structure. Whereas

Table 2: Mixed model: Estimates (upper row) and posterior probabilities to be unrestricted (lower row, in italic) for **C**.

0.43					
1.00					
0.00	-0.51				
0.05	*1.00*				
-0.32	-0.02	0.37			
1.00	*0.23*	*1.00*			
0.37	0.00	0.01	0.00		
1.00	*0.11*	*0.20*	*0.19*		
-0.01	-0.23	0.00	0.00	-0.00	
0.20	*1.00*	*0.09*	*0.02*	*0.18*	
-0.36	-0.00	0.17	-0.00	-0.00	-0.00
1.00	*0.09*	*0.95*	*0.03*	*0.02*	*0.24*

in the mixed model the correlation structure is estimated correctly, this is not the case for the separate models: for the normal response a correlation between x_1 and x_2 is falsely included with inclusion probability 0.58, and for count response only the intercept is selected as random effect.

5 Application

Our data come from an apparel retailer who collected information about the buying behaviour of costumers. The data set comprises monthly data over five years for 2,157 costumers. We included a continuous response about the costumers' monthly profitability contributions and a count response about the number of different items purchased in the respective time periods. The two response variables were related to three covariates measuring marketing activities. These variables are the fraction of spendings the costumer made on weekends (*weekend*), the fraction of shopping trips the costumer made with coupon redemption (*coupon*) and the number of mailings the costumer received in the time period (*mail*).

For MCMC estimation of the random effects model we specified an inverted Gamma prior for the model error variance of the normal component, $\mathcal{G}^{-1}(0,0)$. The Gibbs sampler was run for 40,000 iterations after a burn-in of 10,000. The first 1,000 draws of the burnin were drawn from the unrestricted model. Convergence was checked by running 4 chains from different starting values.

In Table 6 we give posterior estimates and HPD-regions of the mean parameters $\boldsymbol{\beta}$ for the normal and for the count part. The last column equals the posterior probabilities of the mean parameters to be unrestricted. These probabilities are obtained as posterior means of the indicators δ_g. Consistent with marketing managerial expectations we obtain a high effect of the number the coupon redemptions on the profitability contribution as well as on the number of items purchased. Interestingly the effects of *weekend* and *mail* are rather small and for the count response part they have only a probability of 0.55 and 0.22 to be included. In Figure 1 we give

Table 3: Mixed model: Estimates (upper row) and posterior probabilities to be unrestricted (lower row, in italic) for **Q**.

		Normal response			Count response		
		intercept	x_1	x_2	intercept	x_1	x_2
Normal response	intercept	0.18					
		1.00					
	x_1	0.00	0.26				
		0.05	*1.00*				
	x_2	-0.14	0.01	0.24			
		1.00	*0.27*	*1.00*			
Count response	intercept	0.16	0.00	-0.11	0.14		
		1.00	*0.16*	*1.00*	*1.00*		
	x_1	-0.01	0.12	0.01	-0.00	0.06	
		0.20	*1.00*	*0.42*	*0.31*	*1.00*	
	x_2	-0.15	0.00	0.18	-0.13	0.01	0.17
		1.00	*0.13*	*1.00*	*1.00*	*0.34*	*1.00*

Table 4: Separate Models: estimates, 95% HPD intervals and probabilities to be unrestricted for $\boldsymbol{\beta}$.

variable	$\hat{\boldsymbol{\beta}}$	HPD($\hat{\boldsymbol{\beta}}$)	Pr($\delta_g = 1$)
	normal response		
intercept	2.94	(2.87, 3.01)	*1.00*
x_1	0.00	(-0.01, 0.05)	*0.10*
x_2	-0.25	(-0.34, -0.16)	*1.00*
σ^2	0.47	(0.43, 0.51)	
intercept	0.33	(0.26, 0.39)	*1.00*
x_1	0.03	(0.00, 0.10)	*0.41*
x_2	-0.01	(-0.07, 0.00)	*0.17*

paths for the mean of *coupon* for the normal component and of *mail* for the count component. We find a very stable behaviour for the *coupon* variable whereas the uncertainty about the *mail* variable is also reflected in the sample path. Its mean is unrestricted only for 22 percent of the iterations.

In Table 7 we give the posterior estimates of the Cholesky factor **C** and the posterior probabilities of each element of **C** to be unrestricted. From the pattern of restricted elements in **C** we can conclude, that the rank of **Q** equals four on average. All effects are random and rank reduction is caused by linear dependence among these random effects.

Results for the elements of the variance-covariance matrix $\mathbf{Q} = \mathbf{CC}'$ together with posterior probabilities to be unrestricted are given in Table 8. We see that many elements of **Q** are unrestricted with probability 1, whereas the others have a very low probability to be included in the model. It is easy to obtain posterior

Table 5: Separate models: Estimates (upper row) and posterior probabilities to be unrestricted (lower row, in italic) for **Q**.

		Normal response			Count response		
		intercept	x_1	x_2	intercept	x_1	x_2
Normal response	intercept	0.19					
		1.00					
	x_1	0.00	0.26				
		0.16	*1.00*				
	x_2	-0.14	0.03	0.25			
		1.00	*0.58*	*1.00*			
Count response	intercept				0.18		
					1.00		
	x_1				0.00	0.00	
					0.03	*0.08*	
	x_2				-0.02	-0.00	0.01
					0.41	*0.02*	*0.45*

correlations from the covariance matrices in each sample step. We derive pairwise high correlations with coefficients of almost 1 between the random effects of the count and the normal part for the *intercept*, the *weekend*, the *coupon* and *mail* effect. We expect an increased profit value as the number of items purchased increases.

The correlation between the random effects for *weekend* and *mail* is positive. People who like to go shopping on weekends also tend to have a positive reaction to mailings. Interestingly both variables are negatively correlated with the subject specific *intercepts*. This has important managerial implications. High mailing frequencies might lead to diminishing returns on profit values and might reduce the number of items purchased for high-profit-costumers, whereas an increased mailing activity might stimulate low-profit-costumers' interest. Similar conclusions might be drawn for the effect of the fractions of spendings made on weekends. As we will see below this correlation between subject specific effects of *mail* and *weekend* does not become obvious if we estimate only separate models.

We are now going to compare the mixed model with separate models for the normal and count response, respectively. We estimate both models and give results in Table 9 and Table 10, respectively. When comparing the mixed model with the two separate models we do not find much difference in the mean structure but substantial differences in the covariance structure. For the normal as well as for the count model the effects of *mail* are fixed now, and for the count part the *mail* variable is totally deleted during 62 percent of the iterations. For *weekend* the probability to be a random effect is 0.53 and only 0.04, respectively. The two separate models would suggest that the number of mailings as well as the fraction of spendings made on weekend have a small positive effect on the profit value as well as on the number of items purchased. Only from the mixed model it becomes obvious that the employment of the marketing action *mail* or a high fraction of spendings on weekends might even lead to reduced average profit values and numbers of purchased

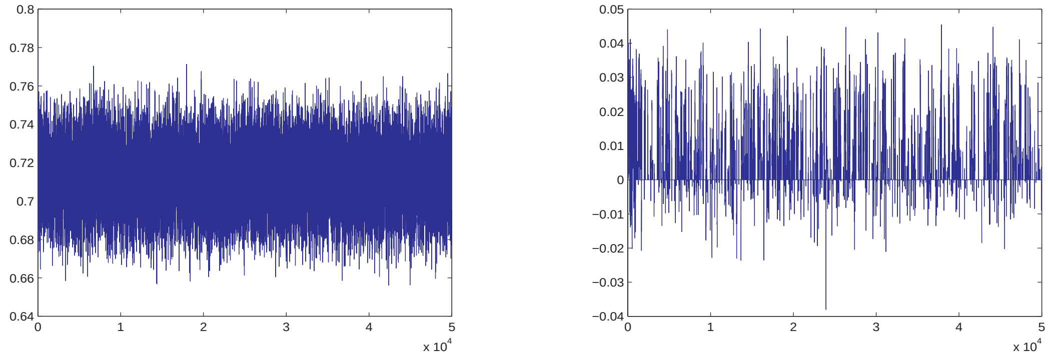

Figure 1: Sample paths of the *coupon* effect for the normal response (left) and the *mail* effect for the count response (right)

Table 6: Mixed Model: estimates, 95% HPD intervals and probabilities to be unrestricted for $\boldsymbol{\beta}$.

variable	$\hat{\boldsymbol{\beta}}$	HPD($\hat{\boldsymbol{\beta}}$)	$\Pr(\delta_g = 1)$
		normal response	
intercept	4.65	(4.62, 4.68)	*1.00*
weekend	0.06	(0.03, 0.09)	*1.00*
coupon	0.71	(0.68, 0.74)	*1.00*
mail	0.05	(0.02, 0.08)	0.96
σ^2	0.62	(0.61, 0.63)	
		count response	
intercept	0.50	(0.47, 0.52)	*1.00*
weekend	0.01	(0.00, 0.04)	*0.55*
coupon	0.37	(0.34, 0.39)	*1.00*
mail	0.00	(-0.00, 0.03)	*0.22*

items for some customers.

6 Conclusions

We present a random effects model for jointly estimating panel data of mixed response types. We may incorporate normal, binary, multinomial and Poisson count data into our model. Contrary to existing methods our estimation procedure is very easy to implement and works for very general model specifications. Auxiliary mixture sampling leads to an MCMC scheme which involves solely standard densities. Variable and/or covariance selection may be added to the estimation procedure and enables us to specify all variables at hand as random effects from the start. If necessary these variables are restricted to fixed effects or even excluded during the course of the selection procedure.

Table 7: Mixed model: Estimates (upper row) and posterior probabilities to be unrestricted (lower row, in italic) for **C**.

-0.34							
1.00							
0.13	0.28						
1.00	*1.00*						
0.00	-0.00	0.29					
0.09	*0.03*	*1.00*					
0.14	-0.01	0.00	0.23				
1.00	*0.21*	*0.00*	*1.00*				
-0.21	0.00	-0.00	-0.00	0.00			
1.00	*0.02*	*0.04*	*0.02*	*0.02*			
0.08	0.18	-0.00	-0.00	0.00	-0.00		
1.00	*1.00*	*0.02*	*0.04*	*0.00*	*0.02*		
0.00	0.00	0.19	0.00	-0.00	0.00	-0.00	
0.02	*0.04*	*1.00*	*0.02*	*0.00*	*0.00*	*0.02*	
0.10	-0.01	-0.00	0.18	-0.00	-0.00	0.00	0.00
1.00	*0.21*	*0.02*	*1.00*	*0.00*	*0.00*	*0.00*	*0.02*

Table 8: Mixed model: Estimates (upper row) and posterior probabilities to be unrestricted (lower row, in italic) for **Q**.

		Normal response				Count response			
		intercept	weekend	coupon	mail	intercept	weekend	coupon	mail
Normal response	intercept	0.11							
		1.00							
	weekend	-0.04	0.09						
		1.00	*1.00*						
	coupon	-0.00	0.00	0.08					
		0.09	*0.12*	*1.00*					
	mail	-0.05	0.01	0.00	0.08				
		1.00	*1.00*	*0.10*	*1.00*				
Count response	intercept	0.07	-0.03	-0.00	-0.03	0.04			
		1.00	*1.00*	*0.13*	*1.00*	*1.00*			
	weekend	-0.03	0.06	0.00	0.01	-0.02	0.04		
		1.00	*1.00*	*0.14*	*1.00*	*1.00*	*1.00*		
	coupon	-0.00	0.00	0.06	0.00	-0.00	0.00	0.04	
		0.02	*0.06*	*1.00*	*0.05*	*0.07*	*0.08*	*1.00*	
	mail	-0.03	0.01	0.00	0.06	-0.02	0.01	-0.00	0.04
		1.00	*1.00*	*0.12*	*1.00*	*1.00*	*1.00*	*0.06*	*1.00*

7 Acknowledgements

We thank Thomas Reutterer for providing the data and for many illuminating discussions of the marketing issues. Thanks to Christoph Pamminger for carefully

Table 9: Separate Models: estimates, 95% HPD intervals and probabilities to be unrestricted for $\boldsymbol{\beta}$.

variable	$\hat{\boldsymbol{\beta}}$	HPD($\hat{\boldsymbol{\beta}}$)	$\Pr(\delta_g = 1)$
normal response			
intercept	4.64	(4.61, 4.66)	*1.00*
weekend	0.07	(0.04, 0.09)	*1.00*
coupon	0.73	(0.70, 0.75)	*1.00*
mail	0.05	(0.03, 0.07)	*1.00*
σ^2	0.68	(0.67, 0.70)	
count response			
intercept	0.49	(0.47, 0.52)	*1.00*
weekend	0.03	(0.00, 0.05)	*0.80*
coupon	0.38	(0.35, 0.40)	*1.00*
mail	0.01	(0.00, 0.03)	0.38

Table 10: Separate models: Estimates (upper row) and posterior probabilities to be unrestricted (lower row, in italic) for $\mathbf{Q}$.

		Normal response				Count response			
		intercept	weekend	coupon	mail	intercept	weekend	coupon	mail
Normal response	intercept	0.05 *1.00*							
	weekend	0.00 *0.01*	0.01 *0.53*						
	coupon	0.02 *0.92*	0.00 *0.02*	0.02 *1.00*					
	mail	-0.00 *0.03*	-0.00 *0.01*	-0.00 *0.03*	0.00 *0.06*				
Count response	intercept					0.01 *1.00*			
	weekend					0.00 *0.01*	0.00 *0.04*		
	coupon					0.01 *1.00*	0.00 *0.01*	0.02 *1.00*	
	mail					0.00 *0.01*	-0.00 *0.00*	0.00 *0.01*	0.00 *0.08*

reading the manuscript and for his competent comments.

8 Appendix

8.1 Constructing $\mathbf{W}_i$

Let $\boldsymbol{z}_i = (z_{i,1}, \ldots, z_{i,dK})'$ denote the individual effects for subject i. Conditional on $\boldsymbol{z}_i$ the design matrix for the first column of $\mathbf{C}$ is constructed from all dK columns of $\tilde{\mathbf{X}}_i$ and the first element of $\boldsymbol{z}_i$ and equals $\tilde{\mathbf{X}}_{i(1:dK)} \cdot z_{i,1}$. To construct the design matrix for the lower triangular part of the second column of $\mathbf{C}$ we have to combine only the last $dK-1$ columns of $\tilde{\mathbf{X}}_i$ with the second element of $\boldsymbol{z}_i$: $\tilde{\mathbf{X}}_{i(2:dK)} \cdot z_{i,2}$. We proceed in that way until the design matrices for all lower triangular columns of $\mathbf{C}$ are constructed. Finally we stack all these matrices and obtain the new design matrix

$$\mathbf{W}_i = [\tilde{\mathbf{X}}_{it(1:dK)} z_{i,1} \; \tilde{\mathbf{X}}_{i(2:dK)} z_{i,2} \; \ldots \; \tilde{\mathbf{X}}_{i(dK)} z_{i,dK}].$$

The vector of the regression coefficients which belongs to $\mathbf{W}_i$ has dimension $dK(dK+1)/2$ and consists of the lower triangular elements of $\mathbf{C}$ stacked columnwise.

8.2 Sampling the utilities for binary data

The data augmentation step is based on model (4) and is valid for any specification of a linear predictor, like for example for a random effects specification $\boldsymbol{\eta}^b_{it} = \mathbf{x}_{it}\boldsymbol{\beta}^b_i$. As the errors in (4) follow a type I extreme value distribution the latent utilities u_{it} are derived from exponential distributions

$$\begin{aligned} \exp(-u_{it}) &\sim \mathcal{E}\left(\exp(\eta_{it})+1\right) && \text{if} \quad y_{it}=1, \\ \exp(-u_{it}) &\sim \mathcal{E}\left(\exp(\eta_{it})+1\right) + \mathcal{E}\left(\exp(\eta_{it})\right) && \text{if} \quad y_{it}=0. \end{aligned} \tag{15}$$

and it is easy to generate the utilities from

$$u_{it} = -\log\left(-\frac{\log(U_{it})}{1+\exp(\eta_{it})} - \frac{\log(U^*_{it})}{\exp(\eta_{it})} I_{\{y_{it}=0\}}\right), \tag{16}$$

where U_{it} and U^*_{it} are uniform random variables and $I_{\{\cdot\}}$ denotes the indicator function.

8.3 Sampling the arrival and inter-arrival times for count data

For count data the data augmentation step based on model (5), (6) requires sampling of the auxiliary variables $\tau^c_{it,1}$ and $\tau^c_{it,2}$.

If $y^c_{it} = 0$, $\tau^c_{it,1}$ is the waiting time for the first jump of the corresponding Poisson process with intensity μ^c_{it}, which is known to occur after $t = 1$. The residual inter-arrival time ξ_{it} after $t = 1$ follows the exponential distribution, $\xi_{it} \sim \mathcal{E}\left(\mu^c_{it}\right)$, and therefore

$$\tau^c_{it,1} = 1 + \xi_{it}.$$

For $y^c_{it} > 0$, $\tau^c_{it,1}$ is the inter-arrival time between the last jump before and the first jump after $t = 1$. In this case a further auxiliary variable $\tau^c_{it,2}$, the arrival time of the y^c_{it}-th jump, is required and $\tau^c_{it,1}$ is given as

$$\tau^c_{it,1} = 1 - \tau^c_{it,2} + \xi_{it}.$$

From the properties of the Poisson process it follows that $\tau^c_{it,2}$ is the maximum of y^c_{it} uniform random variables and hence has a Beta distribution

$$\tau^c_{it,2} \sim \mathcal{B}\left(y^c_{it}, 1\right).$$

8.4 Sampling the indicators R

Sampling the indicators $\mathbf{R}$ amounts to sampling the component indicators of the finite normal mixture with $R(\nu)$ components and fixed parameters $m_r(\nu), s^2_r(\nu), w_r(\nu); r = 1, \ldots, R(\nu)$ given in Frühwirth-Schnatter et al. (2009). For each binary observation ν equals 1 and the component indicator r^b_{it} is sampled conditional on the latent utility u_{it} and the linear predictor η^b_{it} from the discrete distribution

$$\Pr(r^b_{it} = r^*|u_{it}, \eta^b_{it}) \propto w_{r^*}(1)\, \varphi(u_{it} - \eta^b_{it}; m_{r^*}(1), s^2_{r^*}(1)), \quad r^* = 1, \ldots, R(1).$$

Here $\varphi(x; \mu, \sigma^2)$ denotes the probability density function of the $\mathcal{N}(\mu, \sigma^2)$ distribution at x. For a count observation the component indicator $r^c_{it,1}$ is sampled from

$$\Pr(r^c_{it,1} = r^*|\tau^c_{it,1}, \eta^c_{it}) \propto w_{r^*}(1)\, \varphi(-\log \tau^c_{it,1} - \eta^c_{it}; m_{r^*}(1), s^2_{r^*}(1)), \quad r^* = 1, \ldots, R(1).$$

For observations $y^c_{it} > 0$ a second component indicator $r^c_{it,2}$ has to be sampled from

$$\Pr(r^c_{it,2} = r^*|\tau^c_{it,2}, \eta^c_{it}) \propto w_{r^*}(y^c_{it})\, \varphi(-\log \tau^c_{it,2} - \eta^c_{it}; m_{r^*}(y^c_{it}), s^2_{r^*}(y^c_{it})) \quad r^* = 1, \ldots, R(y^c_{it}).$$

8.5 Variable and Covariance Selection

Selection of unrestricted elements in $\boldsymbol{\beta}$ and $\mathbf{C}$ amounts to sampling of the indicators $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$, respectively.

To generate a draw from $\delta_g|\boldsymbol{\delta}_{\backslash g}, \boldsymbol{\gamma}, \boldsymbol{z}, \mathbf{R}, \tilde{\mathbf{y}}$ we specify the following conditional prior, where $p_{\boldsymbol{\delta}}$ is equal to the number of non-zero elements in $\boldsymbol{\delta}$, before sampling the new value of δ_g. If the old value $\delta^{old}_g = 1$, then we obtain

$$p(\delta_g = 0|\boldsymbol{\delta}_{\backslash g}) = (dK - p_{\boldsymbol{\delta}} + 1)/(dK + 1), \quad p(\delta_g = 1|\boldsymbol{\delta}_{\backslash g}) = p_{\boldsymbol{\delta}}/(dK + 1).$$

If $\delta^{old}_g = 0$, then

$$p(\delta_g = 0|\boldsymbol{\delta}_{\backslash g}) = (dK - p_{\boldsymbol{\delta}})/(dK + 1), \quad p(\delta_g = 1|\boldsymbol{\delta}_{\backslash g}) = (p_{\boldsymbol{\delta}} + 1)/(dK + 1).$$

The conditional priors for $p(\gamma_h|\boldsymbol{\gamma}_{\backslash h})$ may be derived in the same way with the total number of free elements being $dK(dK + 1)/2$.

To sample the indicators marginally with respect to $\boldsymbol{\beta}^{\boldsymbol{\delta}}$ and $\mathbf{C}^{\gamma}$ we combine (12) with the remaining $(1 - b)$ proportion of the likelihood $p(\tilde{\mathbf{y}}|\boldsymbol{\beta}^{\boldsymbol{\delta}}, \mathbf{C}^{\gamma}, \boldsymbol{z}, \sigma^2_k, \mathbf{R})^{(1-b)}$. Integration with respect to $\boldsymbol{\theta}^{\boldsymbol{\delta}}$ and $\mathbf{C}^{\gamma}$ yields the conditional distribution

$$p(\tilde{\mathbf{y}}|\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{z}, \sigma^2_k, \mathbf{R}) = b^{(p_{\boldsymbol{\delta}}+p_{\gamma})/2} \left(\frac{1}{2\pi}\right)^{(1-b)/2 \sum_{i=1}^N T_i} |\boldsymbol{\Sigma}|^{-(1-b)/2} \exp\left(-\frac{(1-b)}{2} \mathbf{S}\right) \tag{17}$$

where

$$\mathbf{S} = (\tilde{\mathbf{y}} - [\tilde{\mathbf{X}}^{\boldsymbol{\delta}}\ \mathbf{W}^{\gamma}]\, \mathbf{a}_N)' \cdot \boldsymbol{\Sigma}^{-1} \cdot (\tilde{\mathbf{y}} - [\tilde{\mathbf{X}}^{\boldsymbol{\delta}}\ \mathbf{W}^{\gamma}]\, \mathbf{a}_N), \tag{18}$$

and $\mathbf{a}_N$ is given in (14). To sample the indicators $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ a Gibbs sampler or the fast sampling scheme by Smith and Kohn (2002) may be used. When sampling $\boldsymbol{\gamma}$ we have to ensure identification of the Cholesky factors $\mathbf{C}$ as in Frühwirth-Schnatter and Tüchler (2008).

8.6 Sampling the parameters of the random effects model

8.6.1 Sampling $\boldsymbol{\beta}^{\boldsymbol{\delta}}$, $\mathbf{C}^{\gamma}$

We sample the non-zero elements $\boldsymbol{\beta}^{\boldsymbol{\delta}}$ and $\mathbf{C}^{\gamma}$ together in one block. By combining prior (12) with the remaining $(1-b)$ proportion of the conditional likelihood we obtain the normally distributed joint posterior:

$$p(\boldsymbol{\beta}^{\boldsymbol{\delta}}, \mathbf{C}^{\gamma}|\boldsymbol{z}, \sigma_k^2, \mathbf{R}, \tilde{\mathbf{y}}) \sim \mathcal{N}\left(\mathbf{a}_N, \mathbf{A}_N\right),$$

with the posterior moments given in (13) and (14).

8.6.2 Sampling $\boldsymbol{z}$

The individual effects $\boldsymbol{z}_i$ are conditionally independent for subjects $i = 1, \ldots, N$ and are generated from a multivariate normal distribution:

$$\begin{aligned} p(\boldsymbol{z}_i|\boldsymbol{\beta}^{\boldsymbol{\delta}}, \mathbf{C}^{\gamma}, \boldsymbol{\Sigma}_i, \mathbf{R}_i, \tilde{\mathbf{y}}_i) &\sim \mathcal{N}\left(\mathbf{p}_i, \mathbf{P}_i\right), \\ \mathbf{P}_i^{-1} &= (\tilde{\mathbf{X}}_i\mathbf{C})'\boldsymbol{\Sigma}_i^{-1}(\tilde{\mathbf{X}}_i\mathbf{C}) + \mathbf{I}, \\ \mathbf{p}_i &= \mathbf{P}_i(\tilde{\mathbf{X}}_i\mathbf{C})'\boldsymbol{\Sigma}_i^{-1}(\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i^{\boldsymbol{\delta}}\boldsymbol{\beta}^{\boldsymbol{\delta}}). \end{aligned}$$

8.6.3 Sampling σ_k^2

For each normal component k we sample σ_k^2 from the inverted Gamma posterior $\mathcal{G}^{-1}\left(c_{Nk}/2, C_{Nk}/2\right)$ with $c_{Nk} = \sum_{i=1}^{N} T_i + c_{0k}$ and $C_{Nk} = C_{0k} + \sum_{i=1}^{N}\sum_{t=1}^{T_i}(y_{it} - [\mathbf{x}_{it}^{\boldsymbol{\delta}}\,\mathbf{w}_{it}^{\gamma}]\mathbf{a}_N)^2$.

References

Catalano, P. J. and L. M. Ryan (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association 87*, 651–658.

Cox, D. R. and N. Wermuth (1992). Response models for mixed binary and quantitative variables. *Biometrika 79*, 441–461.

Faes, C., M. Aerts, G. Molenberghs, H. Geys, G. Teuns and L. Bijnens (2008). A high-dimensional joint model for longitudinal outcomes of different nature. *Statistics in Medicine 27*, 4408–4427.

Fitzmaurice, G. and N. M. Laird (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American Statistical Association 90*, 845–852.

Frühwirth-Schnatter, S. and R. Frühwirth (2007). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics and Data analysis 51*, 3509–3528.

Frühwirth-Schnatter, S., R. Frühwirth, L. Held, and H. Rue (2009). Improved auxiliary mixture sampling for hierarchical models of non-gaussian data. *Statistics and Computing, forthcoming.*

Frühwirth-Schnatter, S. and R. Tüchler (2008). Bayesian parsimonious covariance estimation for hierarchical linear mixed models. *Statistics and Computing 18*, 1–13.

Frühwirth-Schnatter, S. and H. Wagner (2006). Auxiliary mixture sampling for parameter-driven models of time series of small counts with applications to state space modelling. *Biometrika 93*, 827–841.

George, E. I. and R. E. McCulloch (1997). Approaches for Bayesian variable selection. *Statistica Sinica 7*, 339–373.

Gueorguieva, R. V. and A. Agresti (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association 96*, 1102–1112.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers of Econometrics*, pp. 105–142. New York: Academic.

Meng, X.-L. and D. A. van Dyk (1998). Fast EM-type implementations for mixed effects models. *Journal of the Royal Statistical Society, Series B 60*, 559–578.

Regan, M. M. and P. J. Catalano (1999a). Bivariate dose-reponse modeling and risk estimation in developmental toxicology. *Journal of Agricultural, Biological and Environmental Statistics 4*, 217–237.

Regan, M. M. and P. J. Catalano (1999b). Likelihood models for clustered binary an continuous outcomes: Application to toxicology. *Biometrics 55*, 760–768.

Scott, S. L. (2009). Data augmentation, frequentistic estimation, and the Bayesian analysis of multinomial logit models. *Statistical Papers*, in press.

Smith, M. and R. Kohn (2002). Parsimonious covariance matrix estimation for longitudinal data. *Journal of the American Statistical Association 97*, 1141–1153.

Tüchler, R. (2008). Bayesian variable selection for logistic models using auxiliary mixture sampling. *Journal of Computational and Graphical Statistics 17*, 76–94.

Yang, Y., K. Jian and J. Zhang (2007). Regression models for mixed Poisson and continuous longitudinal data. *Statistics in Medicine 26*, 3782–3800.