A Comparison of Bayesian Mixed Data Models for Austrian SILC Data

Helga Wagner
Johannes Kepler University

Regina Tüchler
Austrian Federal Economic Chamber

Abstract

In many applications multidimensional outcome variables measured on different scales are of interest. In this paper we consider regression modelling of a bivariate response with a normal and a binary component. We use three different approaches to model dependence: a joint logit-normal model for the two responses, a factorization model with linear dependence and a factorization model with flexible non-linear dependence. We apply these approaches to Austrian SILC data to analyse material deprivation and household income.

Keywords: joint modelling, logit-normal, factorization model, data augmentation, material deprivation, living conditions.

1. Introduction

In the past years the topic of well-being of societies became increasingly important in European politics. It is commonly agreed that the GDP does not sufficiently measure this concept and that complementary indicators are necessary to get a more comprehensive picture of living conditions. Initiatives that deal with this subject are the "GDP and beyond" initiative (http://www.beyond-gdp.eu/), the Stiglitz-Sen-Fitoussi Commission (Stiglitz, Sen, and Fitoussi 2009) and the Sponsorship Group on Measuring Progress, Well-being and Sustainable Development (Eurostat 2011). Currently scoreboards of indicators are being developed in European statistics. These scoreboards consist e.g. of economic indicators, social indicators, or environmental indicators. Since such scoreboards include many different measures we face an increasing need for analyses of dependencies between them and of driving factors. From the methodological point of view the analysis of such dependencies requires models which are able to deal with multidimensional data of mixed type. Our paper meets these needs. We develop mixed data models which incorporate continuous and binary data in regression type models. In our application we focus on social measures coming from the European survey on income and living conditions (EU-SILC).

In this paper we investigate the dependence between a monetary and a material aspect of living conditions in Austria. The monetary aspect is captured via the household income - the money each household has to make its living from, whereas the material aspect is represented by the so-called material deprivation indicator. According to the definition a household faces material deprivation if the members are not capable to meet certain predefined needs like e.g. TV, phone, holiday away from home.

We combine the continuous outcome variable household income and the binary outcome variable material deprivation in mixed data models and analyse their dependence on socio-demographic factors like e.g. the age or activity status of the main-income earner, the household type and migration status. We derive the importance of these explanatory variables by introducing variable selection.

We consider different modelling approaches to deal with multidimensional data of mixed type. We define the models for a continuous and a binary outcome, but they may easily be extended to other data types. In all joint models we use a linear regression model for the continuous and a logistic regression model for the binary outcome. Based on a representation of the binary outcome through a latent continuous utility, the joint bivariate distribution of the error terms is specified as a mixture of bivariate normal distributions in the first modelling approach, whereas the other models use a factorization of the joint bivariate error distribution.

Our paper is organized as follows: Section 2 introduces the different models and corresponding priors for the Bayesian analysis. MCMC estimation is explained in Section 3. In Section 4 we analyse Austrian SILC data and compare results from the three different models. We summarize the findings of the paper in Section 5.

2. Model specification

2.1. Regression model

Let $\mathbf{y}_i = (y_i^b, y_i^n)'$ denote a bivariate response observed for subjects $i = 1, \ldots, n$, where y_i^b is a binary and y_i^n a normal component. $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{id})$ denotes a $1 \times (d+1)$ vector of covariates. To specify a joint regression model for $\mathbf{y}_i$, we assume an underlying latent continuous variable u_i, which determines the value of the binary response: $y_i^b = 1$ if $u_i > 0$ and $y_i^b = 0$ otherwise, and model the bivariate response $(u_i, y_i^n)'$. In all models considered in this paper we use a regression specification for the mean of the bivariate response and model dependence via the error terms.

The joint regression model of (u_i, y_i^n) thus is a SUR (seemingly unrelated regressions) model, given as

$$\begin{pmatrix} u_i \\ y_i^n \end{pmatrix} = \begin{pmatrix} \mathbf{x}_i \boldsymbol{\beta}^b \\ \mathbf{x}_i \boldsymbol{\beta}^n \end{pmatrix} + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_i^b \\ \varepsilon_i^n \end{pmatrix}, \tag{1}$$

where $\boldsymbol{\beta}^n = (\beta_0^n, \beta_1^n, \ldots, \beta_d^n)'$ and $\boldsymbol{\beta}^b = (\beta_0^b, \beta_1^b, \ldots, \beta_d^b)'$ denote the regression coefficients including the intercept for the two responses. Further we assume that $\varepsilon_i^n \sim \mathcal{N}\left(0, \sigma^2\right)$ in all models, i.e. we specify a standard normal regression model for y_i^n,

$$y_i^n \sim \mathcal{N}\left(\mathbf{x}_i \boldsymbol{\beta}^n, \sigma^2\right) \tag{2}$$

and specify either the marginal distribution of ε_i^b or the conditional distribution of $\varepsilon_i^b | \varepsilon_i^n$ as an (approximate) standard logistic distribution. In all models we use the representation of the standard logistic distribution $\mathcal{L}og(0,1)$ as a scale mixture of six normal components derived by Mohanan and Stefanski (1992). Table 1 gives the fixed variances s_r^2 and weights w_r of this approximation, which was shown to be very accurate in Frühwirth-Schnatter and Frühwirth (2010).

Table 1: Variances and weights of the normal components in the finite mixture approximation of the standard logistic distribution.

r	1	2	3	4	5	6
s_r^2	0.68159	1.2419	2.2388	4.0724	7.4371	13.772
$100\, w_r$	1.8446	17.268	37.393	31.697	10.89	0.90745

Using this approximation the logistic regression model with linear predictor η_i^b can be represented as

$$u_i = \eta_i^b + \epsilon_i, \qquad p(\epsilon_i) = \sum_{r=1}^{6} w_r \varphi(\epsilon_i; 0, s_r^2), \tag{3}$$

$$y_i^b = I_{(0,\infty)}(u_i), \tag{4}$$

where $\varphi(y_i; \mu, \sigma^2)$ denotes the pdf of the $\mathcal{N}\left(\mu, \sigma^2\right)$-distribution.

In our first model we specify the bivariate distribution of the error term $\boldsymbol{\varepsilon}_i$ in equation (1) as a finite scale mixture of bivariate normal distributions,

$$p(\boldsymbol{\varepsilon}) = \sum_{r=1}^{6} w_r \varphi_2(\boldsymbol{\varepsilon};\, \mathbf{0}, \boldsymbol{\Sigma}_r), \quad \boldsymbol{\Sigma}_r = \begin{pmatrix} s_r^2 & s_r \sigma \rho \\ s_r \sigma \rho & \sigma^2 \end{pmatrix}. \tag{5}$$

Here $\varphi_2(\mathbf{y};\, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the pdf of the bivariate Normal distribution with moments $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The mixture components share the same correlation ρ, but the variances s_r^2 are component specific. We call this model the *logit-normal model* (LN), as the marginal distribution of u_i is essentially logistic with mean $\eta_i^b = \mathbf{x}_i \boldsymbol{\beta}^b$, yielding a logit regression model for the binary response y_i^b.

As a further modelling approach we consider factorization models with a conditional logit model for y_i^b. In the *conditional linear model* (CL) we specify the predictor η_i^b as a linear function of the covariates $\mathbf{x}_i$ and the standardized error of the normal model, i.e.

$$\eta_i^b = \mathbf{x}_i \boldsymbol{\beta}^b + \psi \, \frac{y_i^n - \mathbf{x}_i \boldsymbol{\beta}^n}{\sigma}.$$

The composite error of the latent utility is given as

$$\varepsilon_i^b = u_i - \mathbf{x}_i \boldsymbol{\beta}^b = \psi \frac{\varepsilon_i^n}{\sigma} + \epsilon_i,$$

where the error ϵ_i has a standard logistic distribution and hence the conditional distribution of the composite error is $\varepsilon_i^b | \varepsilon_i^n \sim \mathcal{L}og(\psi \, \varepsilon_i^n / \sigma, 1)$. Note, that marginally ε_i^b has not a standard logistic distribution but a location mixture of logistic components with normal mixing distribution,

$$p(\varepsilon_i^b) = \int_{\mathcal{R}} p(\varepsilon_i^b | \mathcal{L}og(\psi \, \varepsilon_i^n / \sigma, 1)) \, \varphi(\varepsilon_i^n; 0, 1) \, d\varepsilon_i^n,$$

and the conditional linear model therefore is not equivalent to the logit-normal model.

To highlight the difference between both models we make use of the finite mixture approximation of the standard logistic distribution: the joint error distribution in the conditional linear model is

$$p(\varepsilon^b, \varepsilon^n) = \sum_r w_r \varphi(\varepsilon^b; \psi \, \varepsilon^n / \sigma, s_r^2) \varphi(\varepsilon^n / \sigma) = \sum_r w_r \varphi_2((\varepsilon^b, \varepsilon^n / \sigma)';\, \mathbf{0}, \boldsymbol{\Sigma}_r^*),$$

where $\boldsymbol{\Sigma}_r^* = \begin{pmatrix} s_r^2 + \psi^2 & \psi \sigma \\ \psi \sigma & \sigma^2 \end{pmatrix}$. This is a mixture of bivariate normal distributions, where correlations $\rho_r = \psi / \sqrt{s_r^2 + \psi^2}$ and variances $s_r^2 + \psi^2$ are component specific. An implication of this model is that dependence will be smaller for mixture components with higher variance. Note, that in contrast the correlation ρ is constant for all mixture components in the logit-normal model.

The third model we consider is an extension of the conditional linear model, where the linear term $\mathbf{x}_i \boldsymbol{\beta}^b$ is combined with a smooth function of the standardized residuals ε_i^n / σ to

$$\eta_i^b = \mathbf{x}_i \boldsymbol{\beta}^b + f\Big(\frac{y_i^n - \mathbf{x}_i \boldsymbol{\beta}^n}{\sigma}\Big). \tag{6}$$

The smooth function $f(z)$ is represented as a linear combination of B-spline basis functions B_j, see Lang and Brezger (2004),

$$f(z) = \sum_{j=1}^{J} \gamma_j B_j(z),$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_J)$ denotes the coefficients of the B-spline basis functions. As this model allows to capture non-linear dependence between the two error terms ε^n and ε^b it is more flexible than the conditional linear model, and we will call it the *conditional flexible model* (CF). We note here that also in the flexible model the marginal error distribution of ε^b is no longer logistic.

2.2. Prior distributions

Bayesian model specification is completed by assigning prior distributions to the model parameters. We consider a prior of the structure $p(\boldsymbol{\beta}^n, \boldsymbol{\beta}^b, \sigma^2, \boldsymbol{\vartheta}) = p(\boldsymbol{\beta}^n)p(\boldsymbol{\beta}^b)p(\sigma^2)p(\boldsymbol{\vartheta})$, where $\boldsymbol{\vartheta}$ denotes the model specific parameters.

Priors for the regression coefficients $\boldsymbol{\beta}^c$, $c = n, b$ could be specified as multivariate normal distributions. As we intend to perform variable selection we will however use spike and slab prior distributions for all regression effects. Spike and slab priors are mixtures of a spike component at zero, which allows to shrink small effects to zero and a rather flat slab component, see e.g. George and McCulloch (1997); Ishwaran and Rao (2005); Malsiner-Walli and Wagner (2011) for different variants of spike and slab priors. We introduce a vector of binary indicators $\boldsymbol{\delta}^c = (\delta_1^c, \ldots, \delta_d^c)'$ for $c = n, b$, with elements taking the value 1 if the corresponding coefficient is unrestricted and 0 otherwise. The prior for $\boldsymbol{\beta}^c$ can be specified hierarchically as

$$p(\boldsymbol{\beta}^c|\boldsymbol{\delta}^c) = p_{\text{slab}}(\beta_0^c) \prod_{j:\delta_j^c=1} p_{\text{slab}}(\beta_j^c) \prod_{j:\delta_j^c=0} p_{\text{spike}}(\beta_j^c), \tag{7}$$

$$p(\boldsymbol{\delta}^c) = \prod_{j=1}^{d}(\omega^c)^{\delta_j^c}(1-\omega^c)^{1-\delta_j^c}, \qquad \omega^c \sim \mathcal{B}\left(a_{0,c}, b_{0,c}\right), \tag{8}$$

where $\mathcal{B}()$ denotes the Beta distribution. Note, that for both components the intercept is not subject to selection in this specification, and hence is assigned a slab prior.

We will use independent normal slabs $p_{\text{slab}}(\beta_j^c) = \varphi(\beta_j^c; 0, B_0^c)$ and either Dirac spikes, i.e. a point mass at zero or continuous spikes specified as $\mathcal{N}\left(0, \alpha B_0^c\right)$ with $\alpha << 1$. Our choice of the spike component is dictated by convenience of MCMC sampling: Dirac spikes require computation of the marginal likelihood, which can be determined analytically only for normal and conditionally normal regression models. With the normal mixture approximation of the standard logistic distribution, marginal likelihoods are analytically available for all parameters in the joint logit-normal model and the conditional linear model and we use a Dirac spike for $\boldsymbol{\beta}^n$ and $\boldsymbol{\beta}^b$ in these models. In the conditional flexible model a Dirac spike is assigned to $\boldsymbol{\beta}^b$ but we use a continuous spike for the regression effects $\boldsymbol{\beta}^n$ of the normal response.

For the remaining model parameters we assign priors which are standard for Bayesian analysis. In the logit-normal model we assume prior independence of σ and ρ with $\theta = \ln \sigma \sim \mathcal{N}\left(d_0, D_0\right)$ and a standard normal prior truncated to $[-1, 1]$ for ρ. In both factorization models a $\mathcal{G}^{-1}\left(s_0, S_0\right)$-prior is specified for the error variance σ^2. We chose a normal prior, $\mathcal{N}\left(0, P_0\right)$ for ψ in the conditional linear model. As an alternative a spike and slab prior distribution could be specified to determine whether there is linear dependence between the error terms in both models. Finally, in the conditional flexible model we assume the standard second order random walk prior for the spline coefficients $\boldsymbol{\gamma}$ with variance $\tau^2 \sim \mathcal{G}^{-1}\left(a, b\right)$, see Lang and Brezger (2004).

3. Bayesian inference

Bayesian inference for the models specified in Section 2 is feasible by sampling from the posterior distribution using data augmentation and MCMC methods. As we use the latent utility representation for the binary response and the finite mixture approximation of the standard logistic distribution, the latent utilities $\mathbf{u} = (u_1, \ldots, u_n)'$ as well as component indicators $\mathbf{r} = (r_1, \ldots, r_n)'$ have to be sampled additionally to the model parameters. We will use $\boldsymbol{\Theta}$ to denote the collection of all model parameters and the indices $\text{LN}, \text{CL}, \text{CF}$ to address a specific model. The different modelling approaches

lend themselves to different MCMC schemes, which are convenient for posterior estimation. These are detailed below.

3.1. MCMC for the logit-normal model

Conditional on the auxiliary variables $\mathbf{u}$ and $\mathbf{r}$ the logit-normal model is a bivariate linear Gaussian regression model with regression coefficients $\boldsymbol{\beta} = \left((\boldsymbol{\beta}^n)', (\boldsymbol{\beta}^b)' \right)'$, where joint sampling of the binary indicators $\boldsymbol{\delta} = \left((\boldsymbol{\delta}^n)', (\boldsymbol{\delta}^b)' \right)'$ and the regression coefficients $\boldsymbol{\beta}$ is feasible. Posterior inference for all model parameters can be accomplished by the following sampling scheme:

(I) Sample the component indicators $\mathbf{r}$ from $p(\mathbf{r}|\boldsymbol{\beta}, \boldsymbol{\delta}, \rho, \theta, \mathbf{u}, \mathbf{y}) \propto \prod_{i=1}^{n} p(u_i|\boldsymbol{\beta}, \boldsymbol{\delta}, \rho, \theta, r_i, y_i^n) p(r_i)$.

(II) Sample (ρ, θ) and the latent utilities $\mathbf{u}$:

    (IIa) Sample ρ and θ together from the posterior $p(\rho, \theta|\boldsymbol{\beta}, \mathbf{r}, \mathbf{y}) = \prod_{i=1}^{n} p(\mathbf{y}_i|\boldsymbol{\beta}, \rho, \theta, r_i) p(\theta) p(\rho)$ using an MH-step.

    (IIb) For $i = 1, \ldots, n$ sample the latent utilities $\mathbf{u}$ from the posterior $p(u_i|\boldsymbol{\beta}, \rho, \theta, r_i, \mathbf{y}_i)$.

(III) Sample the indicator variables and the regression coefficients $(\boldsymbol{\delta}, \boldsymbol{\beta})$ from the full conditional posterior $p(\boldsymbol{\delta}, \boldsymbol{\beta}|\rho, \theta, \omega^n, \omega^b, \mathbf{u}, \mathbf{r}, \mathbf{y})$.

(IV) For $c = n, b$ sample ω^c from the Beta posterior, $\mathcal{B}\left(\sum \delta_j^c + a_{0,c}, d - \sum \delta_j^c + b_{0,c} \right)$.

All sampling steps with the exception of step (IIa) are simple Gibbs steps. The component indicators in step (I) are sampled independently from the discrete distributions

$$P(r_i = r) \propto \phi\left(\frac{u_i - m_{i,r}}{s_{i,r}} \right) \pi_r, \quad r = 1, \ldots, 6,$$

where $m_{i,r}$ and $s_{i,r}$ are the parameters of the conditional normal distribution $u_i|y_i^n \sim \mathcal{N}\left(m_{i,r}, s_{i,r}^2 \right)$, given as

$$m_{i,r} = \mathbf{x}_i \boldsymbol{\beta}^b + s_r \rho \frac{\varepsilon_i^n}{\sigma}, \tag{9}$$

$$s_{i,r} = s_r \sqrt{1 - \rho^2}. \tag{10}$$

As proposal for the MH-step (IIa) we use a bivariate Student t-distribution with 10 degrees of freedom, where the mean is the ML estimate of the likelihood $p(\mathbf{y}|\boldsymbol{\beta}, \theta, \rho, \mathbf{r})$ after a few maximising iterations and the variance-covariance parameter is the inverse Hessian at this point. The full conditionals for the latent utilities u_i in step (IIb) are the normal distributions $\mathcal{N}\left(m_{i,r}, s_{i,r}^2 \right)$ truncated to $(-\infty, 0)$ if $y_i^b = 0$ and to $(0, \infty)$ if $y_i^b = 1$.

Details on sampling step (III), which is a standard step for variable selection in Gaussian regression models, are given in Appendix A.

3.2. MCMC for the conditional linear model

In the conditional linear model the joint likelihood of normal observations and latent utilities, conditional on the component indicators is given as

$$\prod_{i=1}^{n} p(u_i, y_i^n|r_i, \boldsymbol{\Theta}_{\mathrm{CL}}) = \prod_{i=1}^{n} p(y_i^n|\boldsymbol{\beta}^n, \sigma^2) p(u_i|\boldsymbol{\Theta}_{\mathrm{CL}}, s_{r_i}, y_i^n), \tag{11}$$

where

$$u_i|\boldsymbol{\Theta}_{\mathrm{CL}}, s_{r_i}, y_i^n = \mathbf{x}_i \boldsymbol{\beta}^b + \psi \frac{y_i - \mathbf{x}_i \boldsymbol{\beta}^n}{\sigma} + \tilde{\epsilon}_i, \quad \tilde{\epsilon}_i \sim \mathcal{N}\left(0, s_{r_i}^2 \right). \tag{12}$$

This suggests sampling $(\boldsymbol{\delta}^b, \boldsymbol{\beta}^b, \psi)$ and $(\boldsymbol{\delta}^n, \boldsymbol{\beta}^n)$ separately, as the full conditional posterior distribution of $(\boldsymbol{\delta}^b, \boldsymbol{\beta}^b, \psi)$ involves only the linear regression model (12).

Hence we use the following sampling scheme for the conditional linear model:

(I) Sample the indicators and regression coefficients $(\boldsymbol{\delta}^n, \boldsymbol{\beta}^n)$ jointly from $p(\boldsymbol{\delta}^n, \boldsymbol{\beta}^n | \sigma^2, \boldsymbol{\beta}^b, \psi, \mathbf{u}, \mathbf{r}, \mathbf{y})$.

(II) Sample the error variance σ^2 from its conditional posterior $\mathcal{G}^{-1}\left(s_0 + n/2, S_0 + \sum (y_i^n - \mathbf{x}_i \boldsymbol{\beta}^n)^2\right)$.

(III) Sample the auxiliary variables $\mathbf{u}$ and $\mathbf{r}$ from the full conditional $p(\mathbf{u}, \mathbf{r} | \sigma^2, \boldsymbol{\beta}^n, \boldsymbol{\beta}^b, \psi, \mathbf{y})$.

(IV) Sample the indicators and regression coefficients in the conditional logit model $(\boldsymbol{\delta}^b, \boldsymbol{\beta}^b, \psi)$ from the full conditional
$$p(\boldsymbol{\delta}^b, \boldsymbol{\beta}^b, \psi | \boldsymbol{\beta}^n, \sigma^2, \mathbf{u}, \mathbf{r}, \mathbf{y}).$$

(V) Sample $\omega^c, c = n, b$ from the $\mathcal{B}\left(\sum \delta_j^c + a_{0,c}, d - \sum \delta_j^c + b_{0,c}\right)$.

Sampling steps (III), (IV) and sampling of ω^b in step (V) are standard steps for Bayesian variable selection in a logit regression model, see e.g. Wagner and Duller (2012) for full details. If $\psi = 0$ the joint model decomposes into a linear and a logit regression model and the remaining steps (I), (II) and sampling of ω^n in step (V) perform Bayesian estimation with variable selection in the normal regression model. For $\psi \neq 0$, $\boldsymbol{\beta}^n$ is a regression parameter the heterogeneous linear regression model

$$y_i^n = \mathbf{x}_i \boldsymbol{\beta}^n + \varepsilon_i, \qquad\qquad \varepsilon_i \sim \mathcal{N}\left(0, \sigma^2\right),$$
$$w_i = \frac{\psi}{\sigma} \mathbf{x}_i \boldsymbol{\beta}^n + \tilde{\epsilon}_i, \qquad\qquad \tilde{\epsilon}_i \sim \mathcal{N}\left(0, s_{r_i}^2\right),$$

with working observations defined as $w_i = \psi \frac{y_i^n}{\sigma} - (u_i - \mathbf{x}_i \boldsymbol{\beta}^b)$. For this model, the indicators $\boldsymbol{\delta}^n$ and the regression coefficients $\boldsymbol{\beta}^n$ can be sampled jointly in one Gibbs step, see in Appendix B for details.

3.3. MCMC for the conditional flexible model

In the flexible specification of the conditional model given as

$$u_i | y_i^n = \mathbf{x}_i \boldsymbol{\beta}^b + \sum_{j=1}^{J} B_j \left(\frac{y_i^n - \mathbf{x}_i \boldsymbol{\beta}^n}{\sigma}\right) \gamma_j + \tilde{\epsilon}_i, \quad \tilde{\epsilon}_i \sim \mathcal{N}\left(0, s_{r_i}^2\right) \tag{13}$$

the errors of the normal regression model enter nonlinearly. Hence the posterior of $\boldsymbol{\delta}^n$, marginalised over the regression effects $\boldsymbol{\beta}^n$ is not available in closed form. This is the reason why we choose a spike and slab prior with continuous spike for the regression effect $\boldsymbol{\beta}^n$ which allows to sample the indicators $\boldsymbol{\delta}^n$ conditional on $\boldsymbol{\beta}^n$ (see Malsiner-Walli and Wagner 2011, for more details). As the full conditional posteriors of $\boldsymbol{\beta}^n$ and σ^2 are not of closed form we sample these parameters using an MH-step.

The sampling scheme for posterior inference in the conditional flexible model consists of the following steps:

(I) Sample $\boldsymbol{\beta}^n$ and $\boldsymbol{\delta}^n$.

(Ia) Sample the regression coefficients $\boldsymbol{\beta}^n$ of the normal model from the full conditional $p(\boldsymbol{\beta}^n | \boldsymbol{\delta}^n, \boldsymbol{\beta}^b, \boldsymbol{\gamma}, \sigma^2, \mathbf{u}, \mathbf{r}, \mathbf{y})$ using an MH-step.

(Ib) Sample the indicators $\boldsymbol{\delta}^n$ from the full conditional $p(\boldsymbol{\delta}^n | \boldsymbol{\beta}^n, \omega^n) = \prod_{j=1}^{d} p(\delta_j^n = 1 | \beta_j^n, \omega^n)$, where
$$p(\delta_j^n = 1) = \frac{1}{1 + \frac{1-\omega^n}{\omega^n} L_j}, \qquad L_j = \frac{p_{\text{spike}}(\beta_j^n)}{p_{\text{slab}}(\beta_j^n)}.$$

(II) Sample the error variance σ^2 from its conditional posterior $p(\sigma^2 | \boldsymbol{\beta}^n, \boldsymbol{\beta}^b, \boldsymbol{\gamma}, \mathbf{u}, \mathbf{r}, \mathbf{y})$ using an MH-step.

(III) Sample the auxiliary variables $\mathbf{u}$ and $\mathbf{r}$ from the full conditional $p(\mathbf{u}, \mathbf{r} | \sigma^2, \boldsymbol{\beta}^n, \boldsymbol{\beta}^b, \boldsymbol{\gamma}, \mathbf{y})$.

(IV) Sample the indicators and regression coefficients in the conditional logit model $(\boldsymbol{\delta}^b, \boldsymbol{\beta}^b)$ from the full conditional
$$p(\boldsymbol{\delta}^b, \boldsymbol{\beta}^b | \boldsymbol{\beta}^n, \boldsymbol{\gamma}, \sigma^2, \mathbf{y}, \mathbf{u}, \mathbf{r}).$$

(V) Sample the spline coefficients $\boldsymbol{\gamma}$ from the model

$$u_i - \mathbf{x}_i\boldsymbol{\beta}^b = \sum_{j=1}^{J} B_j(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}^n}{\sigma})\gamma_j + \tilde{\epsilon}_i$$

and the hyper-parameter τ^2 from $p(\tau^2|\boldsymbol{\gamma})$.

Sampling steps (III) and (IV) are described in Wagner and Duller (2012) and details on step (V) are given in Lang and Brezger (2004). For sampling $\boldsymbol{\beta}^n$ and σ^2 in steps (Ia) and (II) respectively, we use the posterior distributions resulting from the marginal regression model (2) as proposals. As these proposals ignore only the information on $\boldsymbol{\beta}^n$ and σ^2 contained in the binary observations, this strategy works well.

4. Analysis of household income and material deprivation

4.1. Data

Our data come from the European household survey EU-SILC, which focuses on income and living conditions but also includes questions about socio-demographic attributes. We combine the logarithm of household income and the material deprivation indicator to the bivariate mixed response, which is analysed applying the different models. According to European guidelines a person is hit by material deprivation if at least four out of the following nine criteria are fulfilled: (1) arrears on mortgage or rent payments, utility bills, hire purchase instalments or other loan payments; (2) household cannot afford paying for one week's annual holiday away from home; (3) household cannot afford a meal with meat, chicken, fish (or vegetarian equivalent) every second day; (4) household cannot bear unexpected financial expenses of an amount which varies for different countries and is about 900 Euros for Austria; (5) household cannot afford a telephone (including mobile phone); (6) household cannot afford a colour TV; (7) household cannot afford a washing machine; (8) household cannot afford a car and (9) household is not able to pay for keeping its home adequately warm.

Our data set contains 3694 households from the EU-SILC 2009 survey in Austria BMASK (2011). Following Fusco, Guio, and Marlier (2010) we consider only those data sets where the main-income-earner of the household, i.e. the person with the highest income, is not retired and at least one adult person is less than 60 years old.

We include several covariates which may have an influence on the responses material deprivation and household income, respectively. Some of these covariates are associated with the main-income earner whereas other covariates are household variables. The variables of the main-income-earner are gender, age, activity status (with categories full-time work, part-time work, unemployed and out-of-labour-force), education (with categories lower education, medium education, higher education and university) and migration background. A person has migration background if he or she either now has or once had a non-EU/EFTA citizenship. To allow for a deviation from a pure linear relationship between the two responses and age we add the logarithm of age as predictor. The household variables are the type of household (with six categories: single, two adults/no children, single-parent household, two adults/one or two children, two adults/more than two children and other household), the type of building (categorized in single-family house, house with two families, multi-family house with three to nine households, multifamily house with more than nine households and other) and the population density (with categories high, medium and low).

4.2. Variable Selection

As a first step of data analysis we performed Bayesian variable selection to identify important regressors for both response variables. We use a uniform prior for the inclusion probabilities $\omega^c \sim \mathcal{B}(1,1)$, normal slabs with variance $B_0 = 5$ and set $\alpha = 0.005^2$ for the continuous spike. We use a standard

normal prior for θ in the logit-normal model and an improper $\mathcal{G}^{-1}(0,0)$-prior for σ^2 in both factorization models. Finally $\psi \sim \mathcal{N}(0,5)$ in the conditional linear model and in the conditional flexible model we use cubic P-splines with 41 inner knots on the interval $[-8,8]$ and $\tau^2 \sim \mathcal{G}^{-1}(0.001,0.001)$.

To estimate posterior inclusion probabilities MCMC was run for 100 000 iterations after a burn-in of 10 000 draws where the first 5 000 draws are from the unrestricted model. Convergence was checked by running several chains. Integrated autocorrelation times are highest for the conditional flexible model, where they range from 12-82 for the regression coefficients of the binary and from 5-54 for the coefficients of the normal response. Posterior means are estimated by the means of all draws after burn-in.

The variable selection results were similar for the normal response log income in all three models, which differ only with respect to the specification for the binary response. Table 2 gives detailed results on estimated inclusion probabilities. Based on posterior inclusion probabilities larger 0.5 the following variables were selected: linear age effect, dummy variables for all categories of activity status and education, migration background, dummy variables for all household categories except for households with 2 adults and 1-2 children, the category "multifamily house with more than nine households" for the type of building variable and the category "low population density". For material deprivation there are slight differences in the three specifications. In all three models dummy variables for all categories of activity status and education, migration background, as well as two dummies for the type of building variable (3-9 families, multifamily house with more than nine households) were selected. The posterior inclusion probability for a logarithmic effect of age and the dummy for households with two adults and no children are close to 0.5 in all three models. The dummy variables for single parent households and for low population density have a posterior inclusion probability larger than 0.5 in the conditional linear model but slightly below 0.5 in the two other models.

To check for sensitivity with respect to the number of P-spline knots, the analysis was repeated with only 21 inner knots, yielding essentially the same results.

4.3. Model Selection

We compare model adequacy of different specifications of the three models by the DIC (Spiegelhalter, Best, Carlin, and Van der Linde 2002), defined as

$$DIC = \overline{D(\mathbf{\Theta})} + p_D(\mathbf{y}, \bar{\mathbf{\Theta}}(\mathbf{y})).$$

Here $\overline{D(\mathbf{\Theta})}$ is the posterior mean of the deviance

$$D(\mathbf{\Theta}) = -2\log p(\mathbf{y}|\mathbf{\Theta}) + 2\log f(\mathbf{y}),$$

where $f(\mathbf{y})$ is some fully specified standardizing term. $p_D(\mathbf{y}, \bar{\mathbf{\Theta}}(\mathbf{y}))$ is a measure of complexity,

$$p_D(\mathbf{y}, \bar{\mathbf{\Theta}}(\mathbf{y})) = E_{\mathbf{\Theta}|\mathbf{y}}(-2\log p(\mathbf{y}|\mathbf{\Theta})) + 2\log p(\mathbf{y}|\bar{\mathbf{\Theta}}(\mathbf{y})),$$

where $\bar{\mathbf{\Theta}}(\mathbf{y})$ is the posterior mean of $\mathbf{\Theta}$. The model with the smallest DIC is to be preferred. DIC is very popular for Bayesian model comparison, as it can readily be computed from the MCMC output. We follow here Celeux, Forbes, Robert, and Titterington (2006) and set $f(\mathbf{y}) = 1$.

For all models we compute the likelihood based on the factorization

$$\log p(\mathbf{y}|\mathbf{\Theta}) = \sum_{i=1}^{n} \big(\log p(y_i^n|\mathbf{\Theta}) + \log p(y_i^b|\mathbf{\Theta}, y_i^n)\big).$$

For the logit-normal model the conditional distribution of y_i^b is given as

$$p(y_i^b|\mathbf{\Theta}, y_i^n) = \begin{cases} 1 - \sum \Phi(\frac{m_{i,r}}{s_{i,r}})\pi_r & \text{if } y_i^b = 0 \\ \sum \Phi(\frac{m_{i,r}}{s_{i,r}})\pi_r & \text{if } y_i^b = 1, \end{cases} \tag{14}$$

where conditional mean $m_{i,r}$ and standard deviation $s_{i,r}$ are given in equations (9) and (10).

Table 2: Posterior inclusion probabilities for the regression effects.

inclusion probability	household income			material deprivation		
	NL	CL	CF	NL	CL	CF
gender (base: male)	0.20	0.23	0.15	0.30	0.36	0.23
age (centered at 15)						
linear (in 10 years)	**1.00**	**1.00**	**0.99**	0.23	0.26	0.23
log	0.06	0.07	0.07	0.49	**0.50**	**0.51**
activity status (base: full time)						
part-time	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
unemployed	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
out-of-labour	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
education (base: lower)						
medium	**1.00**	**1.00**	**1.00**	**0.65**	**0.76**	**0.64**
higher	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
university	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
migration (base: no migration)	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
type of household (base: single)						
2 adults/no children	**1.00**	**1.00**	**1.00**	0.49	**0.56**	**0.51**
single-parent	**0.84**	**0.89**	**0.74**	0.41	**0.54**	0.42
2 adults/1-2 children	0.02	0.03	0.03	0.18	0.22	0.20
2 adults/3+ children	**1.00**	**1.00**	**1.00**	0.19	0.26	0.20
other	**1.00**	**1.00**	**0.99**	0.21	0.27	0.22
type of building (base: single-family)						
2 families	0.02	0.03	0.02	0.23	0.27	0.22
3-9 families	0.45	0.47	0.22	**0.95**	**0.92**	**0.92**
10+ families	**1.00**	**1.00**	**0.98**	**1.00**	**0.99**	**0.99**
other	0.15	0.18	0.13	0.40	0.45	0.40
population density (base: high)						
medium	0.04	0.06	0.03	0.17	0.22	0.17
low	**0.88**	**0.86**	**0.51**	0.47	**0.62**	0.49

As the marginal models for the continuous outcomes are essentially the same in all three specifications, we further focus on a comparison of the conditional models for the binary outcome based on the posterior predictive distribution.

For model comparison we use the Brier score Brier (1950) for the binary outcome y_i^b,

$$S = \sum_{i=1}^{n}(p(Y_i^b = 1) - y_i^b)^2,$$

which takes the value 0 for a perfect forecast and the maximum value 1 for the worst forecast. An estimate of S is obtained by replacing $p(Y_i^b = 1)$ by its posterior mean, given as

$$\hat{p}(Y_i^b = 1|\mathbf{\Theta}, y_i^n) = \frac{1}{m}\sum_{i=1}^{m} p(Y_i^b = 1|\mathbf{\Theta}^{(m)}, y_i^n).$$

We compare four specifications of the three models, which differ with regard to the included covariates. The full model includes all covariates for both responses. Models M1-M3 include only those regressors with inclusion probability larger than 0.5 for the normal response and differ with respect to the covariates included in the logit model. The sparsest model is M3, which includes only those regressors selected in all specifications (i.e. all dummies for activity status, education, migration, dummies for type of building 3-9 families and 10 and more families). M2 includes the regressors selected in at least two models (regressors in M3 plus log(age) and dummy for households with 2 adults/no children) and finally M1 includes all regressors selected in at least one specification (regressors in M2 and dummies for single-parent households and low population density). For each set of covariates the DIC is lowest for the conditional flexible model with the lowest overall value obtained for model M1 and and also the Brier Score is lowest for this model.

Table 3: Model comparison: DIC and Brier Score.

	DIC			Brier Score		
model	NL	CL	CF	NL	CL	CF
full	5233.5	5238.3	**5222.6**	0.0414	0.0415	**0.0408**
M1	5222.9	5228.1	**5212.2**	0.0416	0.0416	**0.0408**
M2	5225.4	5231.6	**5213.8**	0.0418	0.0418	**0.0410**
M3	5229.6	5235.4	**5217.7**	0.0420	0.0420	**0.0412**
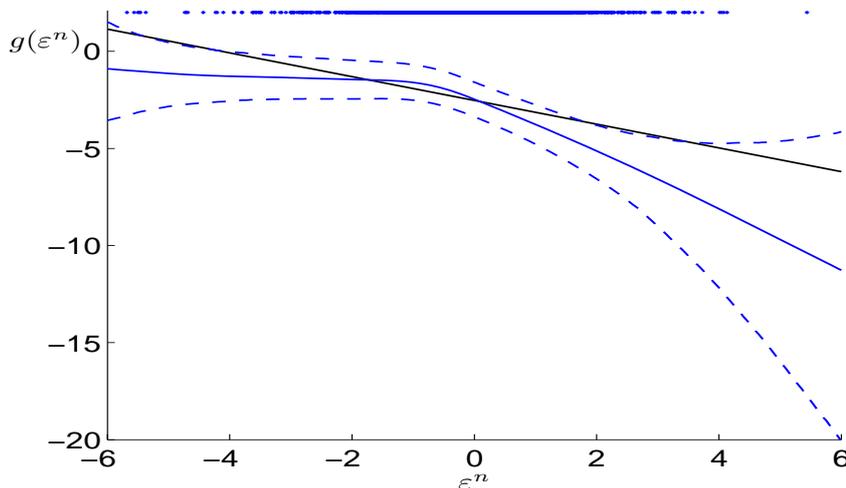
Figure 1: Flexible (full curved line; pointwise 90% HPD-intervals dashed) and linear dependence (full straight line).

Figure 1 compares conditional linear and flexible dependence in model M1. The plot shows the posterior mean of $g(\varepsilon^n) = \beta_0^b + f(\varepsilon^n/\sigma)$, where β_0^b is the intercept in the conditional logit model, together with pointwise 90%-credible intervals. The estimated smooth function has a kink with almost zero slope before and a negative slope after the breakpoint. Hence the risk of material deprivation changes only little before the breakpoint but decreases quickly afterwards. Also shown (in black) is the estimated linear function $\beta_0^b + \psi(\varepsilon^n/\sigma)$ in the conditional linear model. The dots at the top of the figure indicate the residuals from the normal model $\hat{\varepsilon}_i = y_i^n - \mathbf{x}_i \hat{\boldsymbol{\beta}}^n$.

4.4. Results

In Section 4.3, based on DIC and Brier score, the conditional flexible model was selected and we report estimates for the regression effects on log(income) and on material deprivation in Table 4.

From Table 4 we see that age has a positive effect on the household-income. The activity status plays an important role. Naturally, full-time jobs yield the highest household-income and the smallest risk of material deprivation, whereas households with the main-income-earner working only on a part-time basis have less income and a higher risk of material deprivation. These effects are even stronger for households with a main-income-earner who is unemployed or out-of-labour-force. It is well-known that education has an important influence on the living conditions of households. This is also confirmed by our study. The higher the level of education the bigger is the estimated effect on the income and the smaller on the material deprivation response. The importance of the migration status for the economic situation of households is also revealed in our study as households with a main-income-earner who currently has or once had a non-EU/EFTA citizenship have a smaller income and a higher risk of material deprivation. The type of the household has an influence on the income variable. The income of households with two adults and no children is higher than the income of the baseline category single-household, whereas the income of single-parent households and of households with more than two children is smaller. Households living in a building with many flats have less income and are more likely in a situation of material deprivation. Households living in an area with low

Table 4: Posterior mean estimates (std.) of the regression effects.

variable	log(income)	material deprivation
Intercept	9.659 (0.033)	-2.468 (0.542)
gender (base: male)	.	.
age (centered at 15)		
linear (in 10 years)	0.071 (0.007)	.
log	.	-0.326 (0.151)
activity status (base: full-time)		
part-time	-0.243 (0.025)	1.066 (0.257)
unemployed	-0.396 (0.031)	2.396 (0.231)
out-of-labour	-0.550 (0.037)	1.917 (0.322)
education (base: lower)		
medium	0.128 (0.025)	-0.481 (0.215)
higher	0.276 (0.028)	-1.746 (0.302)
university	0.419 (0.029)	-1.831 (0.335)
migration (base: no migration)	-0.285 (0.033)	1.535 (0.250)
type of household (base: single)		
2 adults/no children	0.181 (0.018)	-0.405 (0.251)
single-parent	-0.110 (0.028)	0.444 (0.241)
2 adults/1-2 children	.	.
2 adults/3+ children	-0.206 (0.029)	.
other	0.109 (0.021)	.
type of building (base: single-family)		
2 families	.	.
3-9 families	.	0.643 (0.254)
10+ families	-0.091 (0.017)	0.958 (0.237)
other	.	.
population density (base: high)		
medium	.	.
low	-0.054 (0.015)	-0.395 (0.231)

population density have less income than households in areas with high or medium density.

The proportion of materially deprived estimated in this model is 0.0531 (std. dev. 0.0038) and corresponds exactly to the proportion of materially deprived in the sample.

In this application interest is also on the probability of material deprivation as a function of income. Though not of primary interest in joint regression modelling (where dependence is modelled in the error distribution) the probability of the binary response taking the value 1 as a function of the continuous response is available for given covariate and parameter values in the joint models considered here. For the conditional flexible model it can be estimated by the posterior mean of $p(Y^b = 1|y^n, \mathbf{x})$, i.e.

$$\hat{p}(Y^b = 1|y^n, \mathbf{x}) = \int p(Y^b = 1|y^n, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma) p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma|\mathbf{y}) d\boldsymbol{\beta} d\boldsymbol{\gamma} d\sigma.$$

From the MCMC draws we compute the estimate

$$\bar{p}(Y^b = 1|y^n) = \sum_{m=1}^{M} \frac{\exp(\eta^{(m)})}{1 + \exp(\eta^{(m)})},$$

where

$$\eta^{(m)} = \mathbf{x}\boldsymbol{\beta}^{b,(m)} + \sum_{j=1}^{J} B_j \left(\frac{y^n - \mathbf{x}\boldsymbol{\beta}^{n,(m)}}{\sigma^{(m)}} \right) \gamma_j^{(m)}.$$

Figure 2 shows the estimated probability of material deprivation for different households. In all plots the reference household with a main income earner of median age 42 and baseline values in all other covariates is compared to a household differing in only one covariate value. The estimated probability
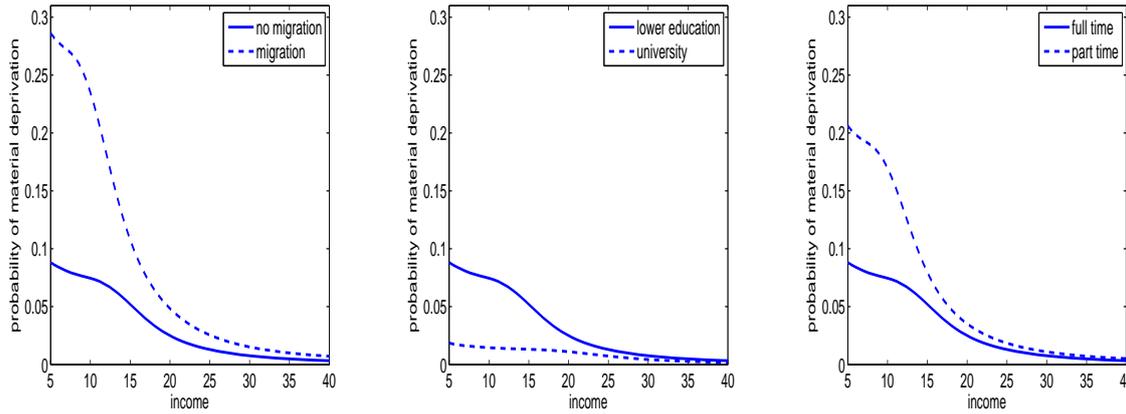
Figure 2: Probability of material deprivation for different households conditional on covariates and income (in 1 000 Euro). Main income earner of age 42 years, baseline values for all other covariates.

of material deprivation is shown for households with/without migration background of the main-income earner in the left panel, for households with lower education/university degree of the main-income-earner in the middle and for households with full-time/part-time employed main-income-earner in the right panel.

For a household income of 21 219 Euro, which is the median income in our sample, the probability of material deprivation is estimated as 2.1% for the reference household but almost doubles to 4.09% when the main-income-earner has migration background. The risk of material deprivation for an otherwise reference household, is estimated as 1.02% if the main-income-earner has a university degree and as 2.97% if the main-income-earner works only part-time.

Due to the non-linear dependence between the error terms each of the curves in Figure 2 has a kink with risk of material deprivation decreasing more pronouncedly afterwards. The income corresponding to this kink is around 10 000 Euro but differs with covariate values.

5. Conclusion

We presented three different approaches for joint regression modelling of a bivariate response with a normal and a binary component. In these models we use the latent utility specification for the binary response. The first two models, the normal-logit and the conditional linear model assume linear dependence between the latent utility and the normal response and differ with respect to the marginal error distribution of the latent utility whereas the third model, the conditional flexible model, allows for non-linear dependence between the error terms of the normal response and the latent utility. For each model Bayesian estimation and variable selection is feasible by straightforward MCMC sampling.

In joint regression modelling of log household income and material deprivation, based on the DIC and the Brier score, the conditional flexible model turned out to be preferred to the other models for each of the mean specifications we considered.

Extensions of the flexible model, where the normal distribution of the continuous response is replaced by a scale mixture of normal distributions with fatter tails, e.g. a t-distribution or a normal-gamma distribution Griffin and Brown (2010) is straightforward and requires only an additional sampling step to draw the scale parameters. Also, the logit model could be easily replaced by a probit or a robit model where the latent utility follows a normal or a t-distribution.

Finally we emphasize that our focus was on joint regression modelling of a binary and a normal outcome. All models considered here share the property that the error term and not the observed normal response enters as a regressor in the conditional model for the binary component. This implies that the effect of the continuous on the binary outcome is heterogeneous, i.e. it depends on

the regressors included in the linear predictor of the continuous response, as shown in Figure 2. A non-differential, smooth nonlinear effect of the endogenous continuous variable could be estimated in a model, where the continuous variable enters as endogenous regressor with smooth nonlinear effect in the logit model. Bayesian estimation of models allowing for a smooth effect of an endogenous regressor is considered in Chib, Greenberg, and Jeliazkov (2009) for continuous response with normal and in Wiesenfarth, Hisgen, Kneib, and Cadarso-Suarez (2012) for a flexible error distribution. As noted in Chib *et al.* (2009) extension to a binary response requires a further data augmentation step, in which the unobserved latent utility is sampled. A comparison of our analysis to this latter modelling approach would be an interesting task for future research.

Appendix

A Sampling indicator variables and regression coefficients in the logit-normal model

Let $\tilde{\mathbf{y}}_i$ denote the bivariate variables $\tilde{\mathbf{y}}_i = (u_i, y_i^n)'$ and $\tilde{\mathbf{y}}$ the stacked vector of all $\tilde{\mathbf{y}}_i$. By $\tilde{\mathbf{X}}$ we denote the corresponding regressor matrix in the joint regression model

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

and $\boldsymbol{\Sigma}$ is the block diagonal matrix with blocks $\boldsymbol{\Sigma}_{r_i}$ given in equation (5).

We sample the indicator variables one at a time from the posterior marginalised over $\boldsymbol{\beta}$, which is given as

$$p(\delta_j = 1|\boldsymbol{\delta}_{\backslash j}, \rho, \theta, \tilde{\mathbf{y}}) = \frac{1}{1 + \frac{p(0, \boldsymbol{\delta}_{\backslash j})}{p(1, \boldsymbol{\delta}_{\backslash j})}R_j}, \quad R_j = \frac{p(\tilde{\mathbf{y}}|0, \boldsymbol{\delta}_{\backslash j}, \rho, \theta)}{p(\tilde{\mathbf{y}}|1, \boldsymbol{\delta}_{\backslash j}, \rho, \theta)},$$

where $\boldsymbol{\delta}_{\backslash j}$ includes all indicators but δ_j. The posterior for δ_j involves the conditional marginal likelihoods of two heteroscedastic linear regression models with design matrices differing only by inclusion/exclusion of the j-th column of the matrix $\tilde{\mathbf{X}}$. The conditional marginal likelihood of a linear regression model is available in closed form as

$$p(\tilde{\mathbf{y}}|\boldsymbol{\delta}, \rho, \theta) \propto \frac{|\mathbf{B}_{\boldsymbol{\delta}}|^{1/2}}{|\mathbf{B}_{0,\boldsymbol{\delta}}|^{1/2}} \exp\left(-\frac{1}{2}(\tilde{\mathbf{y}}'\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{y}} - \mathbf{b}_{\boldsymbol{\delta}}'\mathbf{B}_{\boldsymbol{\delta}}^{-1}\mathbf{b}_{\boldsymbol{\delta}} + \mathbf{b}_{0,\boldsymbol{\delta}}'\mathbf{B}_{0,\boldsymbol{\delta}}^{-1}\mathbf{b}_{0,\boldsymbol{\delta}})\right),$$

where $\mathbf{B}_{\boldsymbol{\delta}}$ and $\mathbf{b}_{\boldsymbol{\delta}}$ are the moments of the normal posterior

$$\mathbf{B}_{\boldsymbol{\delta}} = (\tilde{\mathbf{X}}_{\boldsymbol{\delta}}'\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{X}}_{\boldsymbol{\delta}} + \mathbf{B}_0^{-1})^{-1},$$
$$\mathbf{b}_{\boldsymbol{\delta}} = \mathbf{B}_{\boldsymbol{\delta}}\tilde{\mathbf{X}}_{\boldsymbol{\delta}}'\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{y}},$$

and $\mathbf{X}_{\boldsymbol{\delta}}$ is the appropriate design matrix, including those regressors, for which the corresponding indicator variable takes the value 1.

Regression coefficients β_j for which the corresponding indicator $\delta_j = 0$ are set to zero and the remaining elements $\boldsymbol{\beta}_{\boldsymbol{\delta}}$ are sampled from the normal posterior $\mathcal{N}(\mathbf{b}_{\boldsymbol{\delta}}, \mathbf{B}_{\boldsymbol{\delta}})$.

B Details on posterior sampling the conditional linear model

We give details on sampling the indicators and regression coefficients for the normal response in the conditional linear model. As noted in Section 3.2 we deal with the heterogeneous linear regression model

$$\mathbf{y}^n = \mathbf{X}\boldsymbol{\beta}^n + \boldsymbol{\varepsilon}, \qquad\qquad \boldsymbol{\varepsilon} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}\right), \qquad\qquad (15)$$
$$\mathbf{w} = \mathbf{V}\boldsymbol{\beta}^n + \tilde{\boldsymbol{\epsilon}}, \qquad\qquad \tilde{\boldsymbol{\epsilon}} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{S}\right), \qquad\qquad (16)$$

where $\mathbf{y}^n = (y_1^n, \ldots, y_N^n)'$, and $\mathbf{w} = (w_1, \ldots, w_N)'$ is the column vectors of the (working) responses. $\mathbf{X}$ is the regressor matrix for the normal response, $\mathbf{V}$ is the matrix with rows $\mathbf{v}_i = \frac{\psi}{\sigma}\mathbf{x}_i$ and $\mathbf{S} =$

$\mathrm{diag}(s^2_{r_i})$ is the diagonal matrix of the error variances. The conditional posterior inclusion probability of regression coefficient β^n_k in this model is given as

$$p(\delta^n_k = 1 | \boldsymbol{\delta}^n_{\setminus k}, \omega^n, \sigma^2, \mathbf{S}, \mathbf{y}^n, \mathbf{w}) = \frac{1}{1 + \frac{1-\omega^n}{\omega^n} \frac{p(\mathbf{y}^n, \mathbf{w} | \delta^n_k = 0, \boldsymbol{\delta}^n_{\setminus k}, \sigma^2, \mathbf{S})}{p(\mathbf{y}^n, \mathbf{w} | \delta^n_k = 1, \boldsymbol{\delta}^n_{\setminus k}, \sigma^2, \mathbf{S})}}.$$

Denoting by $\mathbf{X}_{\boldsymbol{\delta}}$ and $\mathbf{V}_{\boldsymbol{\delta}}$ the submatrices of $\mathbf{X}$ and $\mathbf{V}$ consisting of those columns for which the corresponding elements of the indicator vector $\boldsymbol{\delta}$ is equal to 1, and by $\mathbf{B}_{0,\boldsymbol{\delta}}$ the prior variance matrix for the corresponding elements of $\boldsymbol{\beta}^n$, the marginal likelihood $p(\mathbf{y}^n, \mathbf{w} | \boldsymbol{\delta}, \sigma^2, \mathbf{S})$ is explicitly available as

$$-2 \log p(\mathbf{y}^n, \mathbf{w} | \sigma^2, \boldsymbol{\delta}) = (\log |\mathbf{B}_{N,\boldsymbol{\delta}}| - \log |\mathbf{B}_{0,\boldsymbol{\delta}}|) + \Big(\sum_{i=1}^n \big(\frac{(y^n_i)^2}{\sigma^2} + \frac{w^2_i}{s^2_{r_i}} \big) - \mathbf{b}'_{N,\boldsymbol{\delta}} (\mathbf{B}_{N,\boldsymbol{\delta}})^{-1} \mathbf{b}_{N,\boldsymbol{\delta}} \Big) + N \log(2\pi\sigma^2),$$

where

$$\mathbf{B}^{-1}_{N,\boldsymbol{\delta}} = \frac{1}{\sigma^2} \mathbf{X}'_{\boldsymbol{\delta}} \mathbf{X}_{\boldsymbol{\delta}} + \mathbf{V}'_{\boldsymbol{\delta}} \mathbf{S}^{-1} \mathbf{V}_{\boldsymbol{\delta}} + \mathbf{B}^{-1}_{0,\boldsymbol{\delta}},$$

$$\mathbf{b}_{N,\boldsymbol{\delta}} = \mathbf{B}_{N,\boldsymbol{\delta}} (\mathbf{X}'_{\boldsymbol{\delta}} \frac{\mathbf{y}^n}{\sigma^2} + \mathbf{V}'_{\boldsymbol{\delta}} \mathbf{S}^{-1} \mathbf{w}).$$

References

BMASK (2011). "Armutsgefährdung und Lebensbedingungen in Österreich, Ergebnisse aus EU-SILC 2009. Studie der Statistik Austria im Auftrag des BMASK." *Sozialpolitische Studienreihe*, **5**.

Brier GW (1950). "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review*, **78**, 1–3.

Celeux G, Forbes F, Robert C, Titterington DM (2006). "Deviance Information Criteria for Missing Data Models." *Bayesian Analysis*, **4**, 651–674.

Chib S, Greenberg E, Jeliazkov I (2009). "Estimation of Semiparametric Models in the Presence of Endogeneity and Sample Selection." *Journal of Computational and Graphical Statistics*, **18**, 321–348.

Eurostat (2011). "Sponsorship Group on Measuring Progress, Well-being and Sustainable Development, Final report." *EEA ESSC 2011/11/05/EN*.

Frühwirth-Schnatter S, Frühwirth R (2010). "Data Augmentation and MCMC for Binary and Multinomial Logit Models." In T Kneib, G Tutz (eds.), *Statistical Modelling and Regression Structures – Festschrift in Honour of Ludwig Fahrmeir*, pp. 111–132. Physica-Verlag, Heidelberg.

Fusco A, Guio AC, Marlier E (2010). "Income Poverty and Material Deprivation in European Countries." *Eurostat - Methodologies Working papers*.

George EI, McCulloch R (1997). "Approaches for Bayesian Variable Selection." *Statistica Sinica*, **7**, 339–373.

Griffin J, Brown PJ (2010). "Inference with Normal-Gamma Prior Distributions in Regression Problems." *Bayesian Analysis*, **5**, 171–188.

Ishwaran H, Rao SJ (2005). "Spike and Slab Variable Selection: Frequentist and Bayesian Strategies." *Annals of Statistics*, **33**, 730–773.

Lang S, Brezger A (2004). "Bayesian P-Splines." *Journal of Computational and Graphical Statistics*, **13**, 183–212.

Malsiner-Walli G, Wagner H (2011). "Comparing Spike and Slab Priors for Bayesian Variable Selection." *Austrian Journal of Statistics*, **40**, 241–264.

Mohanan J, Stefanski LA (1992). "Normal Scale Mixture Approximations to $F^\star(z)$ and Computation of the Logistics Normal Integral." In N Balakrishnan (ed.), *Handbook of the Logistic Distribution*, pp. 529–549. Marcel Dekker, New York.

Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A (2002). "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal Statistical Society, Series B, Methodological*, **64**(4), 583–616.

Stiglitz J, Sen A, Fitoussi J (2009). "Report by the Commission on the Measurement of Economic Performance and Social Progress." URL www.stiglitz-sen-fitoussi.fr.

Wagner H, Duller C (2012). "Bayesian Model Selection for Logistic Regression Models with Random Intercept." *Computational Statistics and Data Analysis*, **56**.

Wiesenfarth M, Hisgen CM, Kneib T, Cadarso-Suarez C (2012). "Bayesian Nonparametric Instrumental Variable Regression based on Penalized Splines and Dirichlet Process Mixtures." *Technical report*.

Affiliation:

Helga Wagner
Department of Applied Statistics and Econometrics
Johannes Kepler Universität Linz
A-4040 Linz, Austria
E-mail: helga.wagner@jku.at
URL: http://www.jku.at/ifas

Regina Tüchler
Department of Statistics
Austrian Federal Economic Chamber
A-1040 Vienna, Austria E-mail: regina.tuechler@wko.at
URL: https://www.wko.at/Content.Node/Mitarbeiterkontaktseite.html?rollenid=2279681