University of Innsbruck

Bayesian structured additive distributional regression for multivariate responses

**Nadja Klein, Thomas Kneib,
Stephan Klasen, Stefan Lang**

Working Papers in Economics and Statistics

2013-35

Research Platform eeecon
Empirical and Experimental Economics
University of Innsbruck
http://eeecon.uibk.ac.at/

University of Innsbruck
Working Papers in Economics and Statistics

The series is jointly edited and published by

- Department of Economics
- Department of Public Finance
- Department of Statistics

Contact Address:
University of Innsbruck
Department of Public Finance
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 7171
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

The most recent version of all working papers can be downloaded at
http://eeecon.uibk.ac.at/wopec/

For a list of recent papers see the backpages of this paper.

Bayesian Structured Additive Distributional Regression for Multivariate Responses

Nadja Klein, Thomas Kneib
Chair of Statistics
Georg-August-University Göttingen

Stephan Klasen
Chair of Development Economics
Georg-August-University Göttingen

Stefan Lang
Department of Statistics
University of Innsbruck

Abstract

In this paper, we propose a unified Bayesian approach for multivariate structured additive distributional regression analysis where inference is applicable to a huge class of multivariate response distributions, comprising continuous, discrete and latent models, and where each parameter of these potentially complex distributions is modelled by a structured additive predictor. The latter is an additive composition of different types of covariate effects e.g. nonlinear effects of continuous variables, random effects, spatial variations, or interaction effects. Inference is realised by a generic, efficient Markov chain Monte Carlo algorithm based on iteratively weighted least squares approximations and with multivariate Gaussian priors to enforce specific properties of functional effects. Examples will be given by illustrations on analysing the joint model of risk factors for chronic and acute childhood malnutrition in India and on ecological regression for German election results.

Key words: correlated responses; iteratively weighted least squares proposal; Markov chain Monte Carlo simulation; penalised splines; semiparametric regression; Dirichlet regression; seemingly unrelated regression.

1 Introduction

Papers on regression analysis for multivariate responses usually focus on one specific distribution such as the multivariate normal distribution in seemingly unrelated regression (SUR) models [Zellner, 1962, Greene, 2011] or distributions for categorical outcomes and count data, see for example Winkelmann [2008] or Tutz [2011] for recent overviews. Bayesian inference for categorical outcomes has been treated extensively in the last decade [see for example Chen and Dey, 2000, Albert and Chib, 1993, Imai and van Dyk, 2005, Frühwirth-Schnatter et al., 2009]. Most of these approaches focus exclusively on linear predictors following the classical framework of generalised linear models [McCullagh and Nelder, 1989, Fahrmeir et al., 2013]. However, the restriction to a parametric predictor does not capture the flexibility of modelling (possibly more realistic) nonlinear impacts of covariates or spatial variation within the data. Only a few papers are available dealing with multivariate responses and nonparametric predictors. For instance, Fahrmeir and Lang [2001] proposed multicategorical regression models in the spirit of generalised additive models [GAM, Hastie and Tibshirani, 1990, Ruppert et al., 2003, Wood, 2006, Fahrmeir et al., 2013] from a Bayesian point of view. Bayesian SUR models with semiparametric predictors are developed in Lang et al. [2003]. However, in all these models only the mean of the components of the response is related to covariates, neglecting the potential dependence of higher moments or correlations of the response vector on covariates. Therefore, covariate effects that are indeed straightforward to estimate and easy to interpret may lead to misspecification of the model and invalid conclusions drawn from it. Smith and Kohn [2000] for example showed in simulation studies that estimates can become inefficient and that nonlinear effects can be biased when applying univariate regressions instead of a multivariate model.

Accordingly, it is of great interest to provide a framework that is flexible enough to model more than just the mean while remaining interpretable and reliable at the same time. The aforementioned problems can be solved by the framework of generalised additive models for location, scale and shape [GAMLSS, Rigby and Stasinopoulos, 2005]) where potentially complex parametric distributions can be assumed for the response variable. Additionally, each parameter of the distribution i.e. variances and further moments can be modelled in terms of covariates since they are related to additive regression predictors. Estimations for a large number of different types of distributions are obtained from Newton-Raphson or Fisher scoring type algorithms used to maximise the (penalised) likelihood. However, the framework of GAMLSS is currently restricted to univariate responses.

As a consequence, we extend Bayesian distributional regression for univariate responses, recently developed by Klein et al. [2013] to a generic approach for multivariate responses in the spirit of GAMLSS. Inference is realised by a Markov chain Monte Carlo simulation algorithm based on distribution-specific iteratively weighted least squares approximations to the full conditionals. The approach is implemented in the free software package BayesX (www.bayesx.org).

The notion of multivariate distributional regression is more general than multivariate GAMLSS since the parameters of the response distributions do not always relate directly to location, scale or shape but instead functions of several parameters usually lead to these characteristics. In our second application on German elections, the response vector in this study consists of fractions of electoral votes for five different parties where the remaining votes are pooled in an additional component of the response. Hence, a restriction is given by the fact that the sum of all proportions equals one and the positive density can be represented by a five-dimensional open simplex. A natural candidate for analysing such fractions is therefore the Dirichlet distribution with six positive parameters where none of these parameters is directly linked to location, scale or shape but ratios or the sum of several parameters can be interpreted more meaningfully.

The normal distribution is one of a few exceptions where the parameters are directly interpretable since they represent expectation and variance of the response. This favourable property is preserved even in the bivariate case where in addition to first and second moment of the components of the response vector, the correlation parameter can be explained by various covariate effects. We will use the bivariate normal distribution in our application on childhood malnutrition in India in Section 3 in order to study the joint model of two different Z-scores, where one of them represents chronic malnutrition and the other measures an acute poor nutritional status. We also consider an extension based on the bivariate t-distribution to contrast the bivariate normal distribution as an alternative with heavier tails. The performance of Bayesian inference in bivariate normal and bivariate t-models is demonstrated in a simulation study (see the supplement Section A). In the appendix A, we present an extension of the multivariate normal model to the multivariate probit model as an example of binary multivariate regression which is often employed in economic and biostatistical research.

The rest of the paper is structured as follows: In Section 2, we first introduce the representation of multivariate regression models and present a generic formulation for inference in Bayesian structured additive distributional regression for multivariate responses. Section 3 demonstrates a bivariate example based on a high-dimensional geoadditive regression model for a Indian childhood malnutrition data while Section 4 contains results of an ecological regression for elections of the Germany's federal parliament (Bundestag) in 2009. The final Section 5 concludes and provides comments on directions of future research.

2 Bayesian Multivariate Distributional Regression

2.1 Observation Models

Let $f_i(y_{i1}, \dots, y_{iD} | \vartheta_{i1}, \dots, \vartheta_{iK})$, $i = 1, \dots, n$, be the conditional K-parametric densities of D-dimensional random variables $(y_{i1}, \dots, y_{iD})'$ given some covariate information $\boldsymbol{\nu}_i$. With the help of monotone, twice differentiable response functions h_k, the idea of distributional regression is to link each parameter ϑ_{ik} to a semiparametric structured additive predictor η_{ik} formed of the covariates, such that $\vartheta_{ik} = h_k(\eta_{ik})$ and $\eta_{ik} = h_k^{-1}(\vartheta_{ik})$. The response function is usually chosen to maintain restrictions on the parameter space, like the exponential function $\vartheta_{ik} = \exp(\eta_{ik})$ to ensure a parameter with values on the positive real line, or the identity function if the parameter space is unrestricted. If $\vartheta_{ik} \in [-1, 1]$, which is e.g. the case for the correlation between to variables, the transformation $\vartheta_{ik} = \frac{\eta_{ik}}{\sqrt{1+\eta_{ik}^2}}$ is suitable.

2.1.1 Examples of Multivariate Response Distributions

In the following, we describe examples of multivariate distributions that play important and useful roles in applied research. Note, however, that more parametric distributions may be added by transferring the inferential procedure introduced in Section 2.2.

Multivariate Continuous Distributions For a D-dimensional multivariate normal random vector $\boldsymbol{y} = (y_1, \ldots, y_D)'$ we write

$$\boldsymbol{y} \sim \mathrm{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

with expectation $\boldsymbol{\mu} = (\mathrm{E}(y_1), \ldots, \mathrm{E}(y_D))' \in \mathbb{R}^D$ and positive semi definite covariance matrix $\boldsymbol{\Sigma} = \mathrm{Cov}(y_i, y_j) \in \mathbb{R}^{D \times D}$ for $i, j = 1, \ldots, D$. Since $\boldsymbol{\Sigma}$ may not have full column rank (and in this case no proper density exists), we assume that $\boldsymbol{\Sigma}$ is positive definite, implying the density

$$f(y_1, \ldots, y_D) = \frac{1}{\sqrt{(2\pi)^D \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y} - \boldsymbol{\mu}\right)\right).$$

The multivariate normal distribution has several practically and theoretically attractive properties, see for example Kotz et al. [2005].

A special case is the bivariate formulation ($D = 2$) where $\boldsymbol{\mu} = (\mathrm{E}(y_1), \mathrm{E}(y_2))' = (\mu_1, \mu_2)'$ and $\boldsymbol{\Sigma}$ becomes

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

with $\sigma_1^2 = \mathrm{Var}(y_1)$, $\sigma_2^2 = \mathrm{Var}(y_2)$ and $\rho = \mathrm{Cor}(y_1, y_2)$. In distributional regression, the expectations and standard errors of the marginal distributions as well as the correlation parameter can be estimated as functions of covariates, i.e.

$$\eta_i^{\mu_1} = \mu_{1,i}, \qquad \eta_i^{\mu_2} = \mu_{2,i},$$

$$\eta_i^{\sigma_1} = \log(\sigma_{1,i}), \qquad \eta_i^{\sigma_2} = \log(\sigma_{2,i}),$$

$$\eta_i^{\rho} = \frac{\rho_i}{\sqrt{1 - \rho_i^2}}.$$

We will apply the bivariate normal distribution in our application to childhood malnutrition in India (Section 3) and compare it to the bivariate t-distribution, an alternative to the bivariate normal distribution with fatter tails that in general is considered to be less sensitive with respect to extreme observations. The multivariate t-distribution [Kotz et al., 2005] is a multidimensional formulation of the univariate t-distribution. A D-dimensional random variable $\boldsymbol{y} = (y_1, \ldots, y_D)'$ is said to follow a D-dimensional t-distribution, i.e. $\boldsymbol{y} \sim \mathrm{t}\left(n_{df}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ with parameters $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_D)'$, $\frac{n_{df}}{n_{df}-2}\boldsymbol{\Sigma}$ ($\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as defined above) and degrees of freedom $n_{df} > 0$ if the density of

$\boldsymbol{y}$ is given by

$$f(y_1,\ldots,y_D) = \frac{\Gamma\left(\frac{n_{df}+D}{2}\right)}{\Gamma\left(\frac{n_{df}}{2}\right)\left(n_{df}\pi\right)^{D/2}} \left(\det\left(\boldsymbol{\Sigma}\right)\right)^{-\frac{1}{2}} \left[1 + \left(\boldsymbol{y}-\boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{y}-\boldsymbol{\mu}\right)\right]^{-\frac{n_{df}+D}{2}}.$$

Because of $\boldsymbol{y} \xrightarrow{d} \mathrm{N}\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$ for $n_{df} \to \infty$, the multivariate t-distribution can be seen as an approximation of the multivariate normal distribution getting better with increasing n_{df}. Compared to multivariate normal regression we obtain an additional predictor

$$\eta_i^{n_{df}} = \log\left(n_{df}\right).$$

Multivariate Binary Distribution The multivariate probit model is a generalisation of the univariate probit model which can be used to estimate several correlated binary outcomes jointly. This model is of particular interest to researchers since it allows for the estimation of the treatment effect that a binary endogenous variable has on a binary outcome in the presence of unobservables [Heckman, 1978, Maddala, 1983, Woolridge, 2002]. From a Bayesian point of view, inference can be performed based on a latent model representation, that allows to estimate a complex correlation structure on the components of the response: Assume a D-variate probit model with dependent binary variable $\boldsymbol{y} = (y_1,\ldots,y_D)'$ and corresponding unobservable latent variable $\boldsymbol{y}^* = (y_1^*,\ldots,y_D^*)'$. Similar to the univariate case we assume a multivariate normal distribution for $\boldsymbol{y}^*$ and write

$$\boldsymbol{y}^* = \boldsymbol{\eta}^{\mu} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathrm{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}\right)$$

with $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_D)' = (\mathrm{E}(y_1^*),\ldots,\mathrm{E}(y_D^*))'$ and $\boldsymbol{\Sigma} = \mathrm{Cor}(y_{d_1}^*, y_{d_2}^*)$ if $d_1 \neq d_2$ and one otherwise for $d_1, d_2 = 1,\ldots,D$. Then, $\boldsymbol{y}$ is an indicator for whether the latent variable $\boldsymbol{y}^*$ is positive i.e.

$$y_{id} = 1 \iff y_{id}^* > 0, \quad i = 1,\ldots,n \text{ and } d = 1,\ldots,D.$$

Following the ideas of Albert and Chib [1993] we show in Appendix A that Bayesian inference in the multivariate probit model can be realised with the same quantities as for the multivariate normal distribution such that no further computations are necessary apart from the imputation of the latent responses $\boldsymbol{y}^*$.

Dirichlet Distribution In our second application on elections, the quantity of interest is given by a vector of votes or their proportions for different parties. The sum of this vector equals the number of all electorates (or one in case of proportions). Let $D - 1$ be the number of important parties and the remaining ones be grouped in one additional category. Then the response can be described by a D-dimensional non-negative random variable $\boldsymbol{y} = (y_1, \ldots, y_D)'$, $D \geq 2$ that follows a Dirichlet distribution. More specifically, $\boldsymbol{y}$ is said to be Dirichlet distributed with parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D) \in \mathbb{R}^D_{>0}$,

$$\boldsymbol{y} \sim \mathrm{Dir}\,(\boldsymbol{\alpha})$$

if the following two conditions are satisfied:

$$\begin{aligned} f(y_1, \ldots, y_{D-1}) &= \frac{1}{\mathrm{B}(\boldsymbol{\alpha})} \prod_{d=1}^{D} y_d^{\alpha_d - 1} \\ \sum_{d=1}^{D} y_d &= 1, \qquad y_d \geq 0 \end{aligned}$$

where the normalising constant is the multinomial beta function of $\boldsymbol{\alpha}$:

$$\mathrm{B}\,(\boldsymbol{\alpha}) = \frac{\prod_{d=1}^{D} \Gamma(\alpha_d)}{\Gamma\left(\sum_{d=1}^{D} \alpha_d\right)}.$$

We want to give some important properties of the Dirichlet distribution:

- There is a complete symmetry between the D couples (y_d, α_d) due to $y_D = 1 - \sum_{d=1}^{D-1} y_d$ and f could be defined on any subset of size $D - 1$.
- The Dirichlet distribution is a generalisation of the beta distribution since in the cases of $D = 2$ it can be seen easily that y_1 is beta distributed i.e. $y_1 \sim \mathrm{Be}(\alpha_1, \alpha_2)$.
- When all α_d are equal to one, the Dirichlet distribution reduces to the uniform distribution on the simplex defined by $\sum_{d=1}^{D} y_d = 1$, $y_d > 0$ into $\mathbb{R}^D$.
- f is zero outside the open $D - 1$ dimensional simplex.
- The univariate marginal distributions of y_d are beta distributions with parameters α_d and $-\alpha_d + \sum_{k=1}^{D} \alpha_k$.

- The $y_d's$ can be seen as the probability of an event to fall in category d with expectation

$$\mathrm{E}(y_d) = \frac{\alpha_d}{\alpha_0},$$

where $\alpha_0 = \sum_{d=1}^{D} \alpha_d$ can be interpreted as a precision parameter.

- In our approach to Dirichlet regression each parameter α_d is linked to an structured additive predictor i.e $\eta_i^{\alpha_d} = \log(\alpha_{d,i})$.

2.2 Generic Regression Formulation

For any parameter of the multivariate distributions discussed in the previous section, the semiparametric predictor has the general form

$$\eta_i^{\vartheta_k} = \sum_{j=1}^{J_k} f_j^{\vartheta_k}(\boldsymbol{\nu}_i)$$

comprising various functions $f_j^{\vartheta_k}(\boldsymbol{\nu}_i)$ defined on the complete covariate information $\boldsymbol{\nu}_i$. Specific components may for instance be given by

- linear functions $f_j^{\vartheta_k}(\boldsymbol{\nu}_i) = \boldsymbol{x}_i' \beta_j^{\vartheta_k}$, including the overall level $\beta_{0j}^{\vartheta_k}$ of the predictor and $\boldsymbol{x}_i$ is a subvector of $\boldsymbol{\nu}_i$,

- continuous functions $f_j^{\vartheta_k}(\boldsymbol{\nu}_i) = f_j^{\vartheta_k}(x_i)$, where x_i is a single element of $\boldsymbol{\nu}_i$ and f is an appropriate smooth function to represent the effect of x_i on ϑ_{ik},

- spatial variations $f_j^{\vartheta_k}(\boldsymbol{\nu}_i) = f_j^{\vartheta_k}(s_i)$, where s_i is a spatial unit, or

- random effects $f_j^{\vartheta_k}(\boldsymbol{\nu}_i) = \beta_{j,g_i}^{\vartheta_k}$, where g_i is a cluster variable that groups the observations.

The predictors can then always be written in the generic matrix notation

$$\boldsymbol{\eta}^{\vartheta_k} = \sum_{j=1}^{J_k} \boldsymbol{Z}_j^{\vartheta_k} \boldsymbol{\beta}_j^{\vartheta_k}$$

where the design matrices $\boldsymbol{Z}_j^{\vartheta_k}$ are obtained by appropriate basis function expansions and $\boldsymbol{\beta}_j^{\vartheta_k}$ are the vectors of regression coefficients to be estimated.

Specific properties of the basis coefficients such as smoothness are regularised by assuming possibly improper Gaussian priors

$$\boldsymbol{\beta}_j^{\vartheta_k} \sim \left(\frac{1}{(\tau_j^{\vartheta_k})^2}\right)^{\mathrm{rk}\left(\boldsymbol{K}_j^{\vartheta_k}\right)/2} \exp\left(-\frac{1}{2(\tau_j^{\vartheta_k})^2}\left(\boldsymbol{\beta}_j^{\vartheta_k}\right)' \boldsymbol{K}_j^{\vartheta_k}\boldsymbol{\beta}_j^{\vartheta_k}\right) \tag{1}$$

where $\boldsymbol{K}_j^{\vartheta_k}$ is a prior precision matrix and $(\tau_j^{\vartheta_k})^2$ are smoothing variances with inverse gamma hyperpriors i.e. $(\tau_j^{\vartheta_k})^2 \sim \mathrm{IG}(a_j^{\vartheta_k}, b_j^{\vartheta_k})$ in order to obtain a data-driven amount of smoothness with $a_j^{\vartheta_k}, b_j^{\vartheta_k} = 0.001$ as default values for practical analyses.

As a result, each term $\boldsymbol{f}_j^{\vartheta_k} = (f_j^{\vartheta_k}(\boldsymbol{\nu}_i), \ldots, f_j^{\vartheta_k}(\boldsymbol{\nu}_i))' = \boldsymbol{Z}_j^{\vartheta_k}\boldsymbol{\beta}_j^{\vartheta_k}$ is determined by a design matrix $\boldsymbol{Z}_j^{\vartheta_k}$ and a prior precision or penalty matrix $\boldsymbol{K}_j^{\vartheta_k}$. We give specific examples in the following:

Continuous Covariates To approximate potentially nonlinear effects, we use Bayesian P-splines, compare Eilers and Marx [1996] and Brezger and Lang [2006] for detailed explanations. The $n \times D$ design matrix $\boldsymbol{Z}$ in this setting is composed of D B-spline basis functions evaluated at observed covariates x_i. Assuming a first or second order random walk for $\boldsymbol{\beta}$ i.e.

$$\beta_d|\beta_{d-1}, \tau^2 \sim \mathrm{N}\left(\beta_{d-1}, \tau^2\right), \quad d = 2, \ldots, D$$

or

$$\beta_d|\beta_{d-1}, \beta_{d-2}, \tau^2 \sim \mathrm{N}\left(2\beta_{d-1} - \beta_{d-2}, \tau^2\right), \quad d = 3, \ldots, D$$

as smoothness prior with noninformative priors for initial values yields the penalty matrix $\boldsymbol{K} = \boldsymbol{D}'\boldsymbol{D}$ where $\boldsymbol{D}$ is a difference matrix of first or second order.

Spatial Effects For discrete spatial effects observed on a lattice or regions, we consider Markov random fields, see Rue and Held [2005]. Let $s_i \in \{1, \ldots, S\}$ denote the index or region observation i belongs to. Then $f(s_i) = \beta_{s_i}$ is assumed such that we estimate separate parameters $\beta_1, \ldots, \beta_S$ for each region. As a consequence, the $n \times S$ design matrix is an incidence matrix i.e. $\boldsymbol{Z}[i, s] = 1$ if observation i belongs to location s and zero otherwise. The simplest Markov random field prior for the coefficients β_s is defined by

$$\beta_s|\beta_r,\ r \neq s,\ \tau^2 \sim \mathrm{N}\left(\sum_{r \in \partial_s} \frac{1}{N_s}\beta_r, \frac{\tau^2}{N_s}\right),$$

where N_s is the number of regions in ∂_s and ∂_s denotes the neighbours of region s. The penalty matrix is given by

$$\boldsymbol{K}[s,r] = \begin{cases} -1 & s \neq r, \quad r \in \partial_s \\ 0 & s \neq r, \quad r \notin \partial_s \\ N_s & s = r. \end{cases}$$

For detailed explanations on structured additive regression with further examples we refer the reader to Fahrmeir et al. [2013].

Generic Algorithm Depending on the response distribution, the full conditionals $\log\left(p\left(\boldsymbol{\beta}_j^{\vartheta_k}|\cdot\right)\right)$ for the coefficient vectors of several distribution parameters $\boldsymbol{\vartheta}_k$ may not be written in a closed form. In contrast, in the bivariate normal distribution for instance, it can be shown that the Gaussian priors yield a conjugate model for the parameters μ_d, $d = 1, 2$, such that the full conditionals for the regression coefficients corresponding to the expectation parameters are again Gaussian. In the following, we however describe the situation where the full conditionals are not analytically accessible and note that in cases where the full conditionals can be obtained explicitly the resulting Metropolis-Hastings updates are reduced to Gibbs samplers from multivariate normal distributions with the same parameters that we will propose in the approximations to the full conditionals in more complicated situations.

A quadratic Taylor expansion of the log-likelihood function around the mode leads to iteratively weighted least square (IWLS) proposals [Gamerman, 1997] with distribution-and parameter-specific expectation $\boldsymbol{\mu}_j^{\vartheta_k}$ and precision matrix $\boldsymbol{P}_j^{\vartheta_k}$

$$\boldsymbol{\mu}_j^{\vartheta_k} = \left(\boldsymbol{P}_j^{\vartheta_k}\right)^{-1}\left(\boldsymbol{Z}_j^{\vartheta_k}\right)'\boldsymbol{W}^{\vartheta_k}\left(\boldsymbol{z} - \left(\boldsymbol{\eta}_{-j}\right)^{\vartheta_k}\right) \qquad \boldsymbol{P}_j^{\vartheta_k} = \left(\boldsymbol{Z}_j^{\vartheta_k}\right)'\boldsymbol{W}^{\vartheta_k}\boldsymbol{Z}_j^{\vartheta_k} + \frac{1}{\left(\tau_j^{\vartheta_k}\right)^2}\boldsymbol{K}_j^{\vartheta_k} \tag{2}$$

for the regression vectors, which are again multivariate normal distributions i.e.

$$\boldsymbol{\beta}_j^{\vartheta_k} \sim \mathrm{N}\left(\boldsymbol{\mu}_j^{\vartheta_k}, \left(\boldsymbol{P}_j^{\vartheta_k}\right)^{-1}\right).$$

Here, $\boldsymbol{z} = \boldsymbol{\eta}^{\vartheta_k} + \left(\boldsymbol{W}^{\vartheta_k}\right)^{-1}\boldsymbol{v}^{\vartheta_k}$ are the working observations, $\left(\boldsymbol{\eta}_{-j}\right)^{\vartheta_k}$ denotes the predictor without the j-th component, $\boldsymbol{v}^{\vartheta_k} = \partial l/\partial\boldsymbol{\eta}^{\vartheta_k}$ are the score vectors of the log-likelihood $l \equiv l\left(\boldsymbol{\eta}^{\vartheta_1}, \ldots, \boldsymbol{\eta}^{\vartheta_K}\right)$ and $\boldsymbol{W}^{\vartheta_k}$ are working weight matrices,

$w_i^{\vartheta_k} = \mathrm{E}\left(-\partial^2 l/\partial\left(\eta_i^{\vartheta_k}\right)^2\right)$ on the diagonals and zero otherwise. The modularity of MCMC allows to represent the sampler in a unified framework where in each iteration $t = 1, \dots, T$ the final Metropolis-Hastings algorithm loops over the regression coefficients of different effects and over all distribution parameters. In summary, the proposed procedure can therefore be seen as a multivariate extension of the one given Klein et al. [2013] for the univariate case:

1. For $t = 1, \dots, T$ go to 2.
2. For $k = 1, \dots, K$ go to 3.
3. For $j = 1, \dots, J_k$ propose $\boldsymbol{\beta}_j^p$ from the density $q\left(\left(\boldsymbol{\beta}_j^{\vartheta_k}\right)^{[t]}, \boldsymbol{\beta}_j^p\right) = \mathrm{N}\left(\left(\boldsymbol{\mu}_j^{\vartheta_k}\right)^{[t]}, \left(\left(\boldsymbol{P}_j^{\vartheta_k}\right)^{[t]}\right)^{-1}\right)$ with expectation $\boldsymbol{\mu}_j^{\vartheta_k}$ and precision matrix $\boldsymbol{P}_j^{\vartheta_k}$ given in (2) and accept $\boldsymbol{\beta}_j^p$ as a new state of $\left(\boldsymbol{\beta}_j^{\vartheta_k}\right)^{[t]}$ with acceptance probability

$$\alpha\left(\left(\boldsymbol{\beta}_j^{\vartheta_k}\right)^{[t]}, \boldsymbol{\beta}_j^p\right) = \min\left\{\frac{p\left(\boldsymbol{\beta}_j^p|\cdot\right) q\left(\boldsymbol{\beta}_j^p, \left(\boldsymbol{\beta}_j^{\vartheta_k}\right)^{[t]}\right)}{p\left(\left(\boldsymbol{\beta}_j^{\vartheta_k}\right)^{[t]}\middle|\cdot\right) q\left(\left(\boldsymbol{\beta}_j^{\vartheta_k}\right)^{[t]}, \boldsymbol{\beta}_j^p\right)}, 1\right\}.$$

3 Measuring Correlations between Anthropometric Characteristics of Childhood Malnutrition

Childhood malnutrition is one of the most urgent public health challenges in developing and transition countries since it is not only related to the growth of children but also has severe long term impacts. A rich database with information on fertility, family planning, maternal and child health, as well as child survival, HIV/AIDS, malaria, and nutrition is provided by Demographic and Health Surveys (DHS, www.measuredhs.com) consisting of more than 300 surveys conducted in 90 countries. Usually, childhood malnutrition is measured by a Z-score that compares the nutritional status of children in the population of interest with the nutritional status in a reference population. Consequently the Z-score is defined as

$$Z_i = \frac{AC_i - \mu_{AC}}{\sigma_{AC}}$$

where AC_i denotes the anthropometric characteristic for child i, while μ_{AC} and σ_{AC} correspond to median and standard deviation in the reference population (stratified with respect to age and gender). Depending on the choice of the anthropometric indicator, different aspects of malnutrition can be assessed. Insufficient height for age is an indicator for chronic malnutrition (*stunting*) whereas insufficient weight for height captures acute malnutrition (*wasting*). Here, we focus on the joint model with a special interest in the correlation between both risk factors. Our analysis is based on the 1998/99 survey for India, containing information on 24,316 children. A detailed description and a pre-selection of the large number of all potential covariates provided in the data set is given in Belitz and Lang [2008] and Fahrmeir and Kneib [2011]. A boosting and quantile regression based analysis of a similar data set from 2006 India DHS without spatial information can be found in Fenske et al. [2011].

As possible response distributions we assume a bivariate normal and t-distribution.

A complete presentation and discussion of results would go beyond the scope of the paper but additional results can be found in the supplement Section B. Here, we primarily focus on the results for the correlation coefficient between *stunting* and *wasting*. The predictors of all parameters $\mu_{stunting}, \mu_{wasting}, \sigma_{stunting}, \sigma_{wasting}, \rho$ (and in case of the t-distribution of n_{df}) are of the form

$$\begin{aligned}\eta_i &= \boldsymbol{x}_i'\beta + f_1(cage_i) + f_2(breastfeeding_i) + f_3(mage_i) + f_4(mbmi_i) \\ &\quad + f_5(medu_i) + f_6(edupartner_i) + f_{spat}(dist_i) + \beta_{dist_i}\end{aligned}$$

where f_1 to f_6 are smooth functions of the covariates age of the child ($cage$) in months, lactation ($breastfeeding$) in months, age of the mother ($mage$), body mass index of the mother ($mbmi$), education years of the mother ($medu$), and education years of the mother's partner ($edupartner$). We apply Bayesian P-splines of degree three with twenty inner knots and a second-order random walk penalty for the nonlinear smooth terms of continuous covariates. The vector $\boldsymbol{x}_i$ contains a constant comprising the overall level of the predictors and several linear effects (e.g. binary and categorial variables) which are not discussed here. The spatial function f_{spat} and the district-specific random effect β_{dist} represent the complete spatial effect of the district the child belongs to. While f_{spat} captures spatial correlations, β_{dist} catches local and small scale variations. The former one is based on a Markov random field prior and the latter is assigned an i.i.d. Gaussian prior, therefore corresponding to a random intercept based on clusters defined by regions. Lang and Fahrmeir [2001] have shown through simulations that the two components of the complete spatial effect can in general not be separated and only the sum of both effects can be estimated satisfactorily. Therefore results for spatial effects will always contain the sum of the two parts.

The predictive ability of the bivariate normal and the bivariate t-models are compared in terms of the deviance information criterion [DIC, Spiegelhalter et al., 2002] and proper scoring rules, proposed by Gneiting and Raftery [2007]. More precisely we consider the quadratic score $S(f_r, \boldsymbol{y}_r) = 2f_r(y_{r1}, y_{r2}) - ||f_r(y_{r1}, y_{r2})||_2^2$, the spherical score $S(f_r, \boldsymbol{y}_r) = f_r(y_{r1}, y_{r2})/||f_r(y_{r1}, y_{r2})||_2$ and the logarithmic score $S(f_r, \boldsymbol{y}_r) = \log(f_r(y_{r1}, y_{r2}))$ under the Lebesque measure on the measurable space $\left(\mathbb{R}^2, \mathcal{B}\left(\mathbb{R}^2\right)\right)$ and with $||f_r(x,y)||_2 = \left(\int\int f_r(x_1, x_2)^2 \mathrm{d}x_1 \mathrm{d}x_2\right)^{1/2}$. Here, $\boldsymbol{y}_r = (y_{r1}, y_{r2})'$ is an observation from a hold out sample $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_R$ and f_r is either the density of a bivariate normal or a bivariate t-distribution with plugged in predicted parameters which are obtained by cross validation. We therefore split the data set in ten equal parts, use nine parts for estimation and predict the parameters of the remaining part each. The predictive ability of the two models can then be compared by the aggregated average score $S_R = \frac{1}{R}\sum_{r=1}^{R} S(F_r, \boldsymbol{y}_r)$ with predictive distributions $F_r(y_{r1}, y_{r2}) = \int_{-\infty}^{y_{r1}} \int_{-\infty}^{y_{r2}} f_r(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2$. Note that higher scores represent better probabilistic forecasts. Unfortunately, the concept of quantile residuals [Dunn and Smyth, 1996] used by Klein et al. [2013] in the context of univariate Bayesian structured additive distributional regression as a graphical device to check the fit to the data under different response distributions is not expendable to the multivariate framework since the cumulative distribution functions $F_r : U \subset \mathbb{R}^D \to [0,1]$ can not be inverted. In the bivariate normal and t-model it would however be possible to look at the residuals of the marginal distributions.

The scores together with the DIC can be found in Table 1 while the residuals of the marginal distributions are given in Figure 1. Obviously, all goodness of fit measures are slightly in favour of the t-distribution. In addition, the estimated effects based on the bivariate normal and t-distribution are visually close to each other. Therefore, we restrict our presentation primarily to the results based on the bivariate t-distribution.

Figure 2 visualises posterior mean point estimates and posterior probabilities for the complete spatial effect. Figure 3 shows posterior mean estimates for the nonpara-

Distribution	DIC	Quadratic Score	Logarithmic Score	Spherical Score
Bivariate normal	157,601	0.058	-3.34	0.242
Bivariate t	**156,460**	**0.060**	**-3.29**	**0.245**

Table 1: Childhood malnutrition. Comparison of DIC achieved of estimates based on the whole data set and average score contributions obtained from ten-fold cross validations.

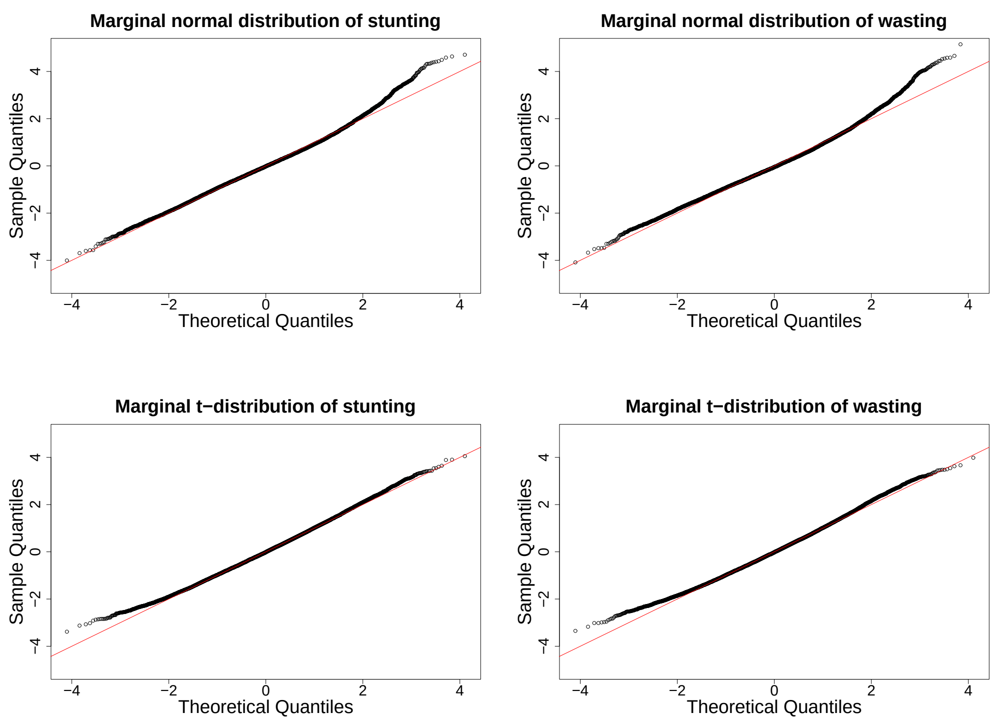

Figure 1: Childhood malnutrition. Quantile-quantile plots of residuals of the marginal distributions in the bivariate normal model (first row) and the bivariate t-model (second row).

metric effects together with 80% and 95% pointwise credible intervals. In Figures 4 (normal distribution) and 5 (t-distribution) contour lines of estimated densities for four different ages are depicted. The other effects are kept constant at the estimated functions evaluated at mean covariate values each.

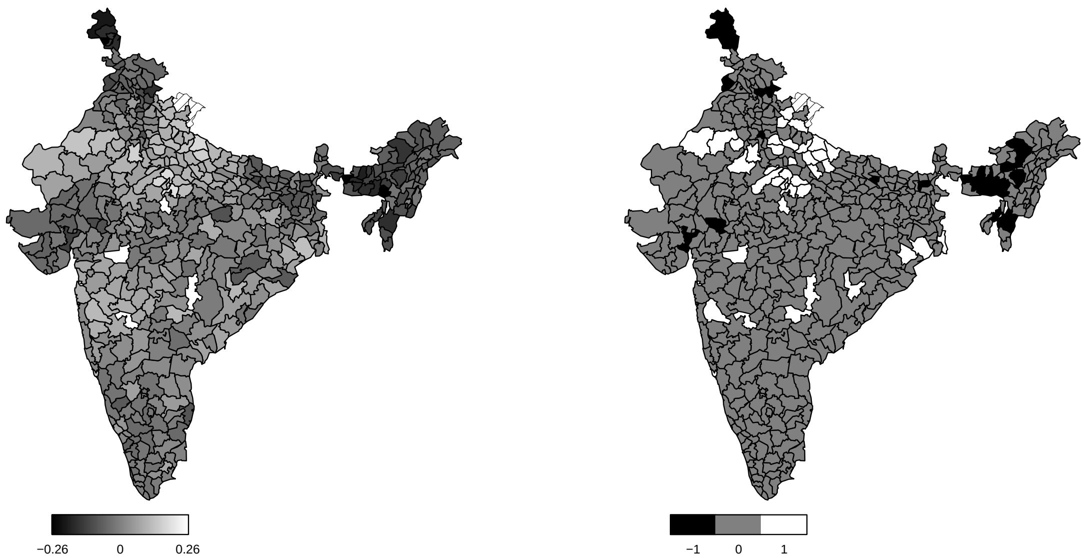

Figure 2: Childhood malnutrition, bivariate t-distribution. Posterior mean estimates of the complete spatial effects on ρ (left) and 80% posterior probabilities (right), centred around zero each.

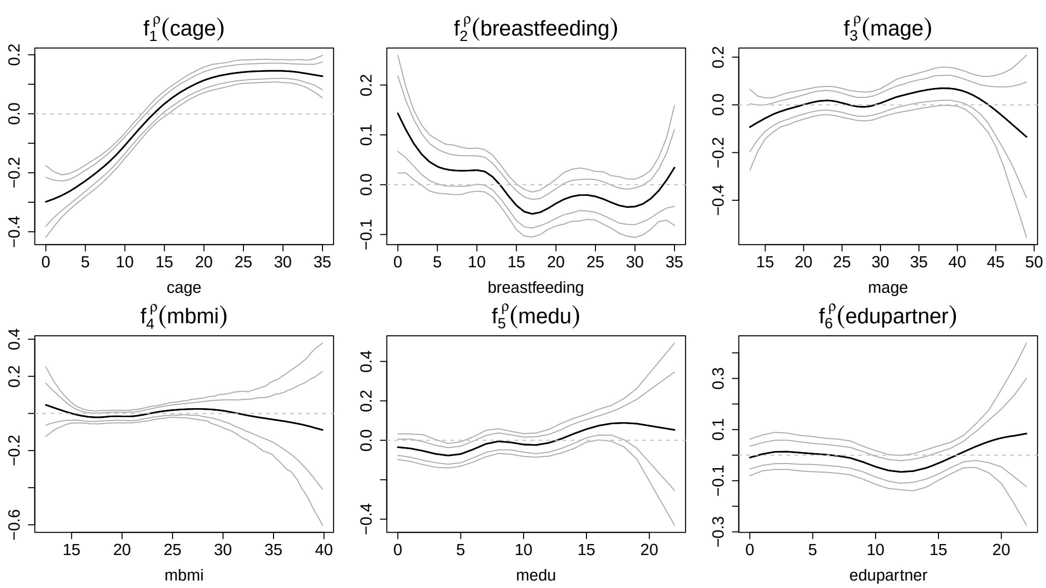

Figure 3: Childhood malnutrition, bivariate t-distribution. Posterior mean estimates of nonparametric effects on ρ together with 80% and 95% pointwise credible intervals, centred around zero each.

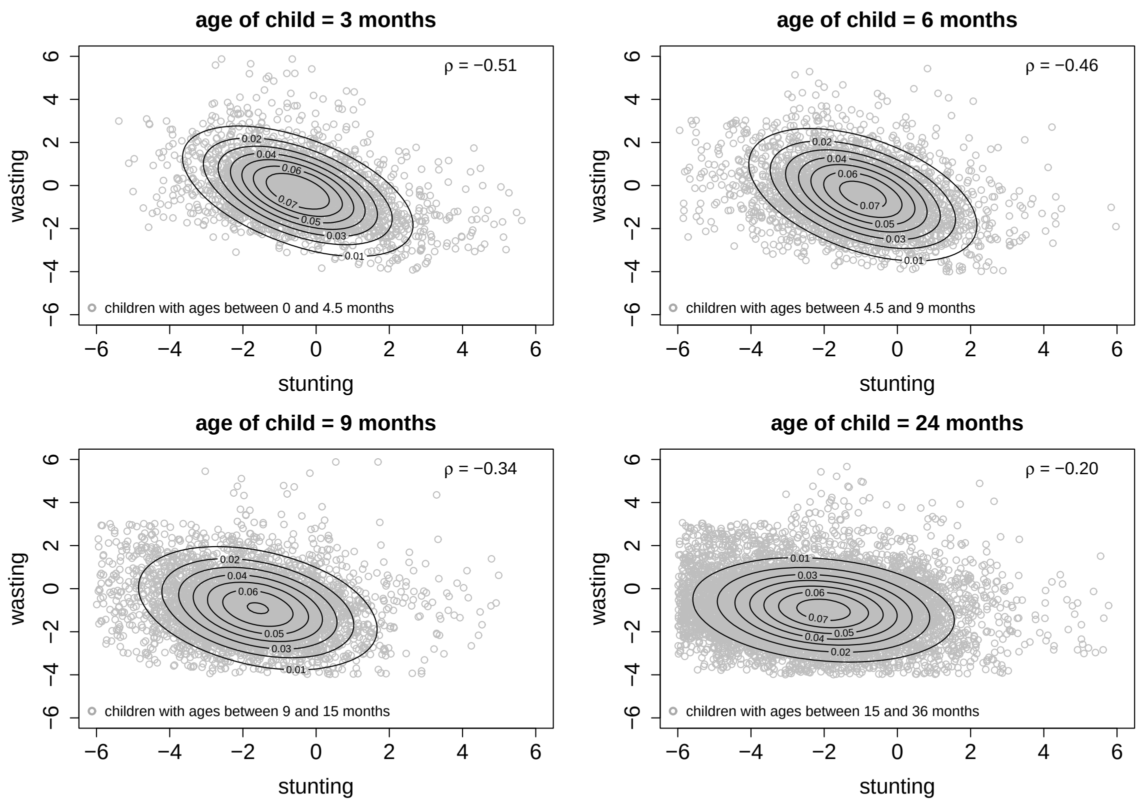

Figure 4: Childhood malnutrition, bivariate normal distribution. Contour lines of densities for four different ages, the remaining effects are kept constant at estimated effects evaluated at mean covariate values each.

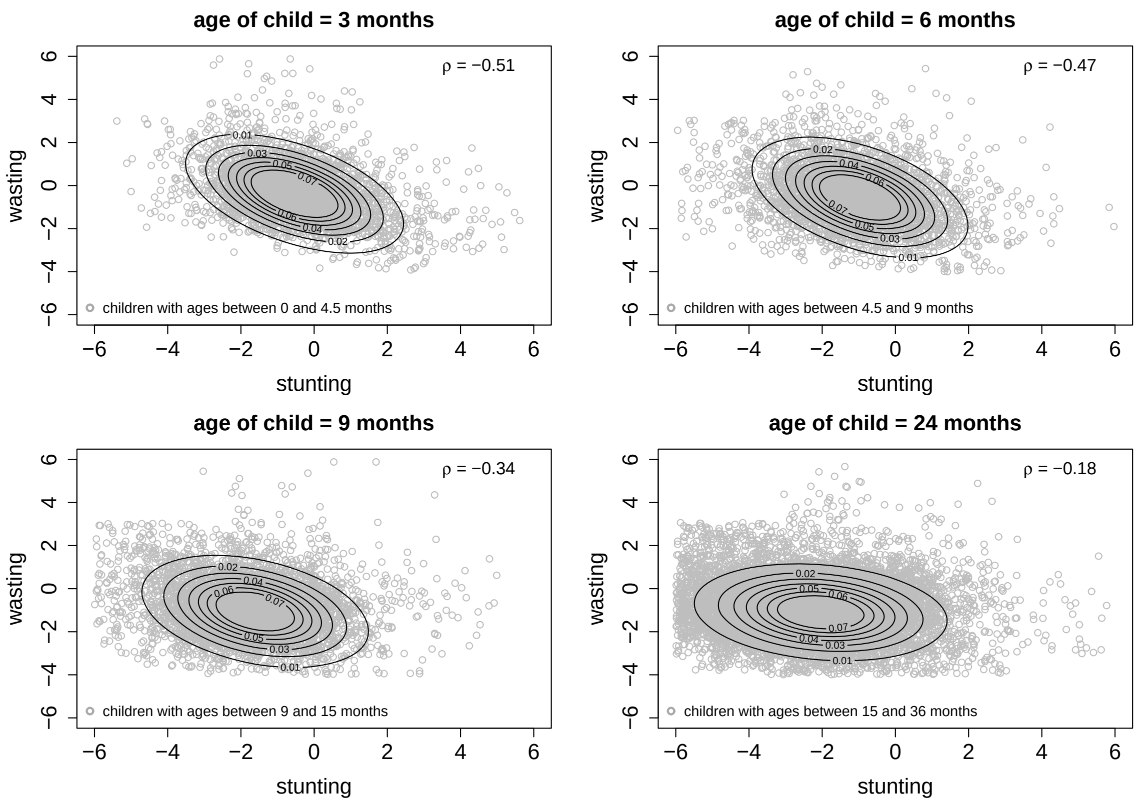

Figure 5: Childhood malnutrition, bivariate t-distribution. Contour lines of densities for four different ages, the remaining effects are kept constant at estimated effects evaluated at mean covariate values each.

Especially in the north of India and at the border to China there are several districts expected to have a negative correlation whereas in some districts in the inner country we estimate a small positive correlation, compare Figure 2. Figure 3 shows that in particular, none of the effects related to the mother's characteristics is estimated to be significant in the sense that posterior credible intervals of all these effects contain the zero line (visualised as a grey dashed line in the figure). In contrast, the effect of age of the child turns out to be nonlinear with a steady increase until an age of about two years. Afterwards the effect stays approximately constant. One explanation is a more dynamic process of malnutrition at the very early months after birth. The effect of duration of breastfeeding which shows a decline up to two years could be explained by the fact that a shorter lactation increases the risk of both, acute and chronic malnutrition.

Figures 4 and 5 indicate that the correlation between *wasting* and *stunting* decreases with the age of the child. Furthermore, the variance of *stunting* increases while the one of *wasting* becomes smaller for older children. This can be related to the adoption of the body to reduced nutrient intake. Children stop growing (become stunted) and this helps in improving their wasting score as they need fewer calories to nourish the smaller body [see Wiesenfarth et al., 2012, for further discussion] The densities obtained from a t-distribution are more flat and with heavier tails as expected.

In a nutshell, estimates indicate that characteristics of the child are more relevant than the mother's abilities to explain the joint model of chronic and acute malnutrition. Especially the age of the child plays an important role since for very young children *wasting* and *stunting* are estimated to have the highest correlation. It furthermore revealed several local pattern with extremer correlations compared to the rest of India.

4 Selected Socio-Demographic Factors on Germany's Federal elections

In this section, we present an analysis on Germany's federal election 2009 as an example of Dirichlet regression. The data are provided by the DSTATIS (Statistisches Bundesamt, www.destatis.de) and contain proportions of the electorate voting (response variable) on 5 parties for each of the 413 districts (Landkreise) in Germany.

The proportion of votes for the Christian Democratic Union/Christian Social Union (CDU/CSU), Social Democratic Party (SPD), The Liberals (FDP), The Left, The Greens and others sum up to one in each district. As covariates, we consider district specific quantities i.e. the proportion of electorates (PoE) compared to the total population (turnout), the rate of unemployment (*unemployment*) in 2008, the gross domestic product per capita ($GDPpc$) in 2008 (measured in thousands) and one of 38 administrative regions (*region*) the districts are located in. The predictors $\eta_i^{\alpha_d} = \log(\alpha_{d,i})$ for $d = 1, \ldots, 6$ and $i = 1, \ldots, 413$ are hence of the form

$$\eta_i = \beta_0 + f_1(PoE_i) + f_2(GDPpc_i) + f_3(unemployment_i) + f_{spat}(region_i).$$

As described in Section 3, f_1 to f_3 are smooth functions modelled with Bayesian P-splines and f_{spat} is assigned a Markov random field prior on 38 administrative regions in Germany. For better interpretation, we computed expected proportions $\frac{\exp(\alpha_k)}{\sum_{d=1}^{6}\exp(\alpha_d)}$ of votes for each party and for every effect while all other effects are constant with estimated effects evaluated at mean covariate values. In Figure 6, the expected values can be compared for different regions in Germany and in Figure 7, the expected proportions are shown in dependency of the covariates PoE, $GDPpc$ and *unemployment*. Note that in order to enhance visibility axes ranges are not the same for different parties.

For the CDU/CSU and FDP we estimate the smallest spatial variations of electorates while for the SPD we observe a higher constituency in the Western part of Germany in general. The Greens seem to have most votes in the South West (since 2011 the prime minister of Baden-Württemberg is provided by the Greens) whereas the most notable spatial effect is recorded for the Left. For historic reasons the Left is traditionally very strong in the East of Germany as well as the Saarland because of the prominent position of its former chairman Oskar Lafontaine in this federal state.

The effect of $GDPpc$ is most pronounced for the Greens with a steady increase up to 50,000 which can be explained by their stronger support in urban areas which compiles with political literature, see for example Walter [2010]. While for SPD, FDP and others the effect of PoE does not have clear trends, the expected number of votes increases for the Greens and decreases for CDU/CSU as well as The Left when the voter turnouts rise. Looking at *unemployment*, SPD and the Left show similar behaviour where up to a rate of 10%, increasing unemployment is estimated to have

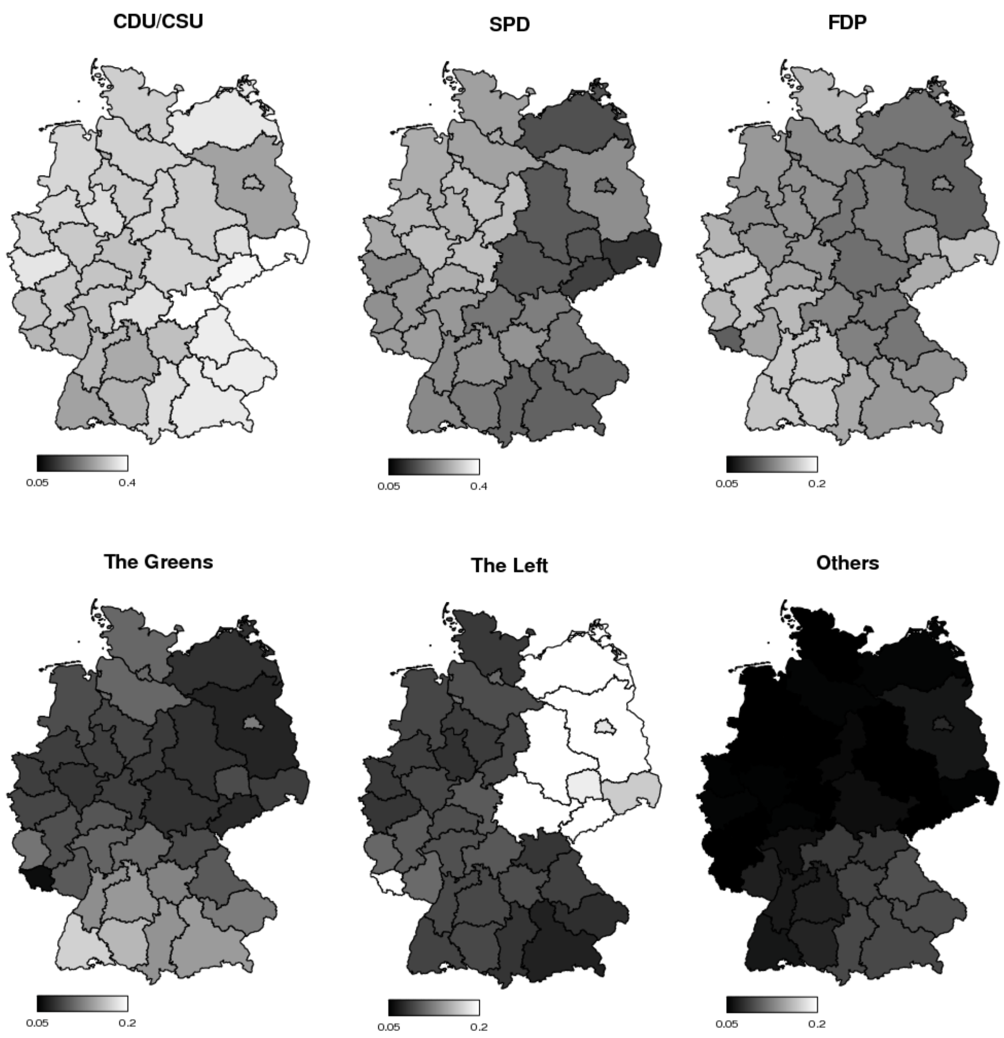

Figure 6: Federal election. Expected proportions of votes for each region, all others effects are evaluated at mean covariate effects. Axes are equal for CDU/CSU, SPD as well as FDP, The Greens, The Left and others, respectively in order to enhance visibility of estimated effects. Geometric information provided by Bundesinstitut fuer Bau-, Stadt- und Raumforschung [BBSR]

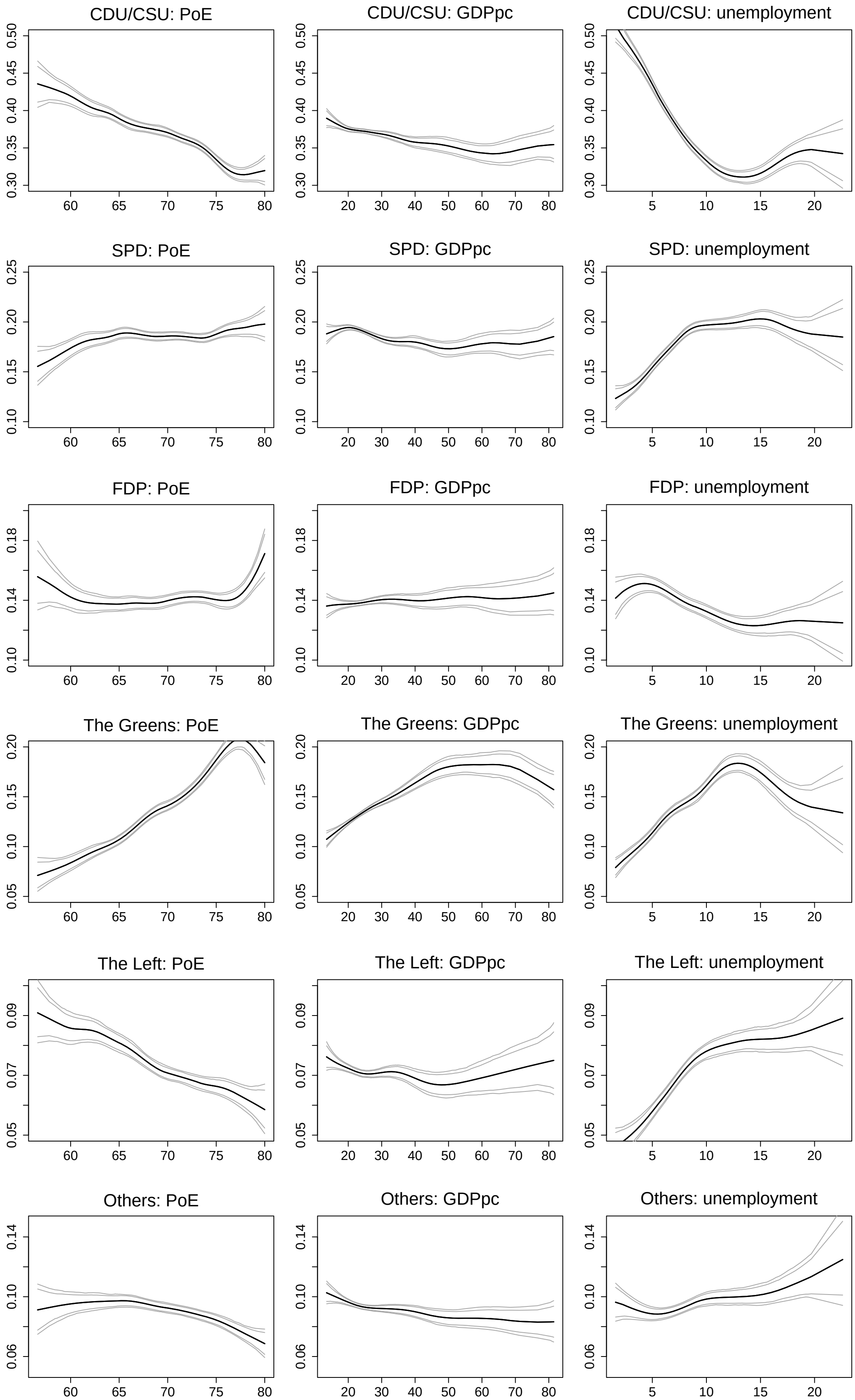

Figure 7: Federal election. Expected proportions of votes together with 80% and 95% pointwise credible intervals, all others effects are evaluated at mean covariate effects. Axes scales are only identical row-wise in order to enhance visibility of estimated effects.

a positive effect on votes for these two parties. For the Greens we observe the same tendency but with a negative effect for high unemployment. In contrast, the CDU reveals most electorates in regions with high rates of employment.

In conclusion it can be said that even with data on a highly aggregated level (we do not have any individual information) clear trends and differences between the considered parties can be identified and explained by the four included covariates geographical region, turnouts, rate of unemployment and the gross domestic product per capita.

5 Summary and Conclusions

In this paper, we have proposed a generic Bayesian framework for structured additive distributional regression with various types of multivariate responses. The flexibility of the approach allows to gain detailed insights into the joint stochastic behaviour of response vectors accounting for a variety of complex regression effects. While interpretation remains feasible for bivariate models for continuous responses, the truly multivariate case with dimensions higher than two remains is challenging since the complex restrictions on the parameter space have to be ensured for positive definite dispersion matrices. A possible remedy would be to work with the Cholesky factor but in this the interpretation of estimated effects gets difficult.

Nevertheless, we believe that multivariate distributional regression is an important contribution to the toolbox of applied statisticians with a variety of applications. In particular, the distributional variant of SUR models with all parameters depending on covariates provides a natural counterpart to recent attempts to define bivariate quantile regression models. Albeit its admittedly much stronger assumption of a particular response distribution, the multivariate normal and the multivariate t model have the considerable advantage to define a coherent, interpretable model for bivariate responses.

In future research, we will consider multivariate hierarchical distributional regression models following the ideas of Lang et al. [2013]. In addition, the multivariate normal model, the multivariate probit model and combinations of binary and continuous covariates in a joint latent normal model will be studied in detail to assess their potential as alternatives to well known model types in economics such as Heckman's selection model. The fact that we provide both joint estimation of effects on several responses and allow for effects not only on the means can be expected to give rise to interesting new insight in corresponding applications.

A Full Conditionals of Latent Variables in the Bivariate Probit Model

In the probit model, the observable binary outcomes are replaced by latent variables as introduced in Section 2.1.1. For simplicity reasons we describe the procedure at the example of a bivariate probit model where $\boldsymbol{y}^*$ is jointly bivariate normal distributed with expectation $\boldsymbol{\mu} = (\eta^{\mu_1}, \eta^{\mu_2})$ and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

It follows that the full conditionals of $\boldsymbol{y}^*$ are truncated bivariate normal distributions

$$\boldsymbol{y}^*|\cdot \sim \mathrm{N}_{[a,b]}\left(\mu, \boldsymbol{\Sigma}\right)$$

with

$$[a,b] = \begin{cases} [0,\infty) \times [0,\infty) & y_1 > 0, y_2 > 0 \\ [0,\infty) \times (-\infty, 0] & y_1 > 0, y_2 < 0 \\ (-\infty, 0] \times [0,\infty) & y_1 < 0, y_2 > 0 \\ (-\infty, 0] \times (-\infty, 0] & y_1 < 0, y_2 < 0 \end{cases}$$

Sampling from a truncated bivariate normal distribution has e.g. been studied by Robert [1995] and a common way is to derive a Gibbs sampler for realising random numbers from the desired truncated bivariate normal distribution. This means that in several steps of the MCMC algorithm another Gibbs sampler would be needed and it is due to ending up with a time demanding procedure. A more efficient way is to draw the components of $\boldsymbol{y}^*$ separately from the conditional distributions. Although the marginal distributions are not truncated Gaussian any more, it can be shown easily that the conditional distributions follow truncated univariate normal distributions i.e.

$$y_i^*|\cdot \sim \begin{cases} \mathrm{N}_{[0,\infty)}\left(\mu_i + \rho(y_j^* - \mu_j), (1-\rho^2)\right) & \text{if} \quad y_i = 1 \\ \mathrm{N}_{(-\infty,0]}\left(\mu_i + \rho(y_j^* - \mu_j), (1-\rho^2)\right) & \text{if} \quad y_i = 0 \end{cases} \tag{3}$$

for $d \neq j$ and $i, j = 1, 2$. If therefore $\vartheta_k = \mu_d$ for $k, d = 1, 2$ holds in step 2 of the generic algorithm summarised in Section 2.2, draw an additional update of y_d^* from a truncated normal distribution (3).

References

J. H. Albert and S. Chib. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association, 88(422):669–679, 1993.

C. Belitz and S. Lang. Simultaneous selection of variables and smoothing parameters in structured additive regression models. Computational Statistics and Data Analysis, 53: 61–81, 2008.

A. Brezger and S. Lang. Generalized structured additive regression based on Bayesian P-splines. Computational Statistics & Data Analysis, 50:967–991, 2006.

Bundesinstitut fuer Bau-, Stadt- und Raumforschung (BBSR): Regierungsbezirke, GeoBasis-BKG, 2002.

M. Chen and D. Dey. Bayesian analysis for correlated ordinal data models. In: Dey, D., Ghosh, S. and Mallick, B. (des.), Generalized Linear Models. A Bayesian Perspective, pages 133–159, 2000.

P. K. Dunn and G. K. Smyth. Randomized quantile residuals. Computational and Graphical Statistics, 5:236–245, 1996.

P. H. Eilers and B. D. Marx. Flexible smoothing using B-splines and penalized likelihood. Statistical Science, 11:89–121, 1996.

L. Fahrmeir and T. Kneib. Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data. Oxford University Press, New York, 2011.

L. Fahrmeir and S. Lang. Bayesian semiparametric regression analysis of multicategorical time-space data. Annals of the Institute of Statistical Mathematics, 53:11–30, 2001.

L. Fahrmeir, T. Kneib, S. Lang, and B. Marx. Regression - Models, Methods and Applications. Springer, 2013.

N. Fenske, T. Kneib, and T. Hothorn. Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. Journal of the American Statistical Association, 106:494–510, 2011.

S. Frühwirth-Schnatter, R. Frühwirth, L. Held, and H. Rue. Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. Statistics and Computing, 19: 479–492, 2009.

D. Gamerman. Sampling from the posterior distribution in generalized linear mixed models. Statistics and Computing, 7:57–68, 1997.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102(477):359–378, 2007.

W. H. Greene. Econometric Analysis. Prentice Hall, 2011.

T. J. Hastie and R. J. Tibshirani. Generalized Additive Models. Chapman & Hall, 1990.

J. Heckman. Dummy endogenous variables in a simultaneous equation system. Econometrica, 46:931–959, 1978.

K. Imai and K. I. van Dyk. A bayesian analysis of the multinomial probit model using marginal data augmentation. Journal of Econometrics, 124:311–334, 2005.

N. Klein, T. Kneib, and S. Lang. Bayesian structured additive distributional regression. Technical report, 2013. URL `http://eeecon.uibk.ac.at/wopec2/repec/inn/wpaper/2013-23.pdf`.

S. Kotz, N. Balakrishnan, and N. L. Johnson. Continuous Multivariate Distributions: Volume 1: Models and Applications. Wiley Series in Probability and Statistics, New York, 2005.

S. Lang and L. Fahrmeir. Bayesian generalized additive mixed models. a simulation study. discussion paper 230, sfb 386. supplement paper to Fahrmeir, L. and Lang, S. (2001): Bayesian semiparametric regression analysis of multicategorical time-space data. annals of the institute of statistical mathematics, 53, 10-30. Technical report, 2001. URL `http://www.uibk.ac.at/statistics/personal/lang/publications/`.

S. Lang, S. B. Adebayo, L. Fahrmeir, and W. J. Steiner. Bayesian geoadditive seemingly unrelated regression. Computational Statistics, 18:263–292, 2003.

S. Lang, N. Umlauf, P. Wechselberger, K. Harttgen, and T. Kneib. Multilevel structured additive regression. Statistics and Computing, 23, 2013.

G. S. Maddala. Limited Dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge, 1983.

P McCullagh and J. A. Nelder. Generalized Linear Models. Chapman & Hall, 1989.

R. A. Rigby and D. M. Stasinopoulos. Generalized additive models for location, scale and shape (with discussion). Applied Statistics, 54:507–554, 2005.

C. P. Robert. Simulation of truncated normal variables. Statistics and Computing, 5: 121–125, 1995.

H. Rue and L. Held. Gaussian Markov Random Fields. Chapman & Hall / CRC, 2005.

D. Ruppert, M. P. Wand, and R. J. Carroll. Semiparametric Regression. Cambridge University Press, 2003.

M. Smith and R. Kohn. Nonparametric seemingly unrelated regression. Journal of Econometrics, 98:257–281, 2000.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, 65(B):583–639, 2002.

G. Tutz. Regression for Categorical Data. Cambridge University Press, 2011.

F. Walter. Gelb oder Grün? Kleine Parteiengeschichte der besserverdienenden Mitte in Deutschland. transcript, 2010.

M. Wiesenfarth, T. Krivobokova, S. Klasen, and S. Sperlich. Direct simultaneous inference in additive models and its application to model undernutrition. Journal of the American Statistical Association, 107(500):1286–1296, 2012.

R. Winkelmann. Econometric Analysis of Count Data. Springer, 2008.

S. N. Wood. Generalized Additive Models : An Introduction with R. Chapmann & Hall, 2006.

J. M. Woolridge. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, 2002.

A. Zellner. An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. Journal of the American Statistical Association, 57:500–509, 1962.

Bayesian Structured Additive Distributional Regression for Multivariate Responses Supplement

Nadja Klein, Thomas Kneib
Chair of Statistics
Georg-August-University Göttingen

Stephan Klasen
Chair of Development Economics
Georg-August-University Göttingen

Stefan Lang
Department of Statistics
University of Innsbruck

A Simulation Study for the Bivariate Normal and Bivariate t-Distribution

In order to validate the performance of Bayesian structured additive distributional regression for the bivariate normal and the bivariate t-distribution we consider the following model setup:

$$
\begin{aligned}
f^{\mu_1}(x_1) &= \sin(x_1) & f^{\mu_2}(x_2) &= \cos(x_2) \\
f^{\sigma_1}(x_1) &= \sqrt{x_1}\cos(x_1) & f^{\sigma_2}(x_2) &= 0.3x_1\cos(x_2) \\
f^{\rho}(x_1) &= \log(x_1) & f^{n_{df}}(x_2) &= \frac{1}{3}x_2^2
\end{aligned}
$$

where $f^{n_{df}}$ is only used in the bivariate t-model. The covariates x_1 and x_2 are uniform random numbers from the intervals $[1, 6]$ and $[-3, 3]$, respectively. For each of the two models we simulated 250 data sets with sample size $n = 1,000$.

Figure A1 shows the comparison of logarithmic mean squared errors (MSE) of all effects in the bivariate normal and the bivariate t-model. Figures A2 and A6 show pointwise 95% coverage rates for each simulated effect in the bivariate normal model and the bivariate t-model respectively. The posterior mean estimates of all effects compared to the true simulated functions for three specific replications (chosen as the $5\%, 50\%$ and 95% quantile of the MSE sample distribution) as well as 95% posterior probabilities can be compared in Figures A3 to A5 for the bivariate normal model and in Figures A7 to A9 for the bivariate t-model. All considered results are predictions resulting from a grid of step size 0.01 within the covariate ranges $[1, 6]$ and $[-3, 3]$, respectively.

Results can shortly be summarised as follows:

- Coverage rates are generally close to the nominal level such that they can be considered as a helpful tool in applied analyses to estimate uncertainties of effects. Except of the effect f^{μ_1} in the t-model, our approach tends to slightly overestimate uncertainty in the simulated example.

- From MSE in form of boxplots depicted in Figure A1 the indication of slightly better results for the bivariate normal model can be confirmed.

- The approach yields points estimates which are quite close to the true functions for all distribution parameters, even for the replication with the 95% worst MSE.

B Supplemenatry Material to Section 3

In this section we provide summary plots of all included nonlinear and spatial effects of the model applied in Section 3 of the main paper for the remaining parameters $\mu_{stunting}$, $\mu_{wasting}$, $\sigma_{stunting}$ and $\sigma_{wasting}$. The predictors for these parameters are of the form (3) given in the main paper. All shown effects are centred around zero.

C Supplemenatry Material to Section 4

In Figures C16 and C17 we show posterior mean estimates of all covariates *region*, *PoE*, *GDPpc* and *unemployment*. In Figure C17 80% and 95% pointwise credible intervals indicate uncertainties of the estimated effects. In each graph, effects are centred around zero.

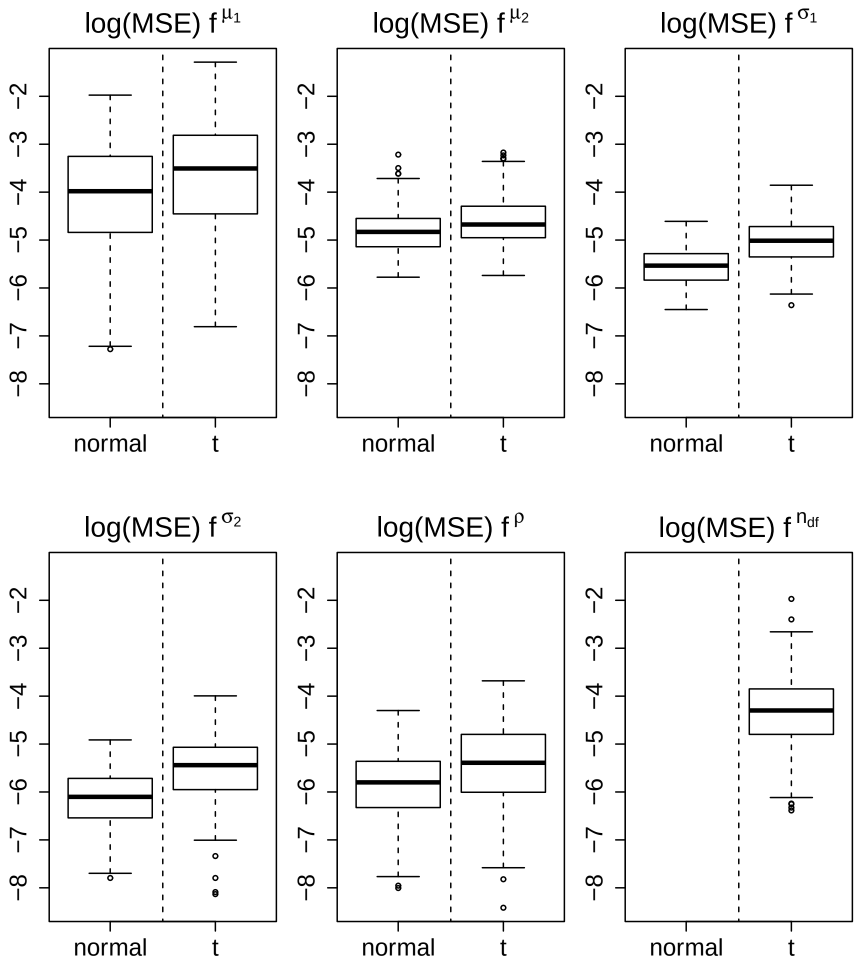

Figure A1: Simulation Study. Boxplots of logarithmic MSE of nonlinear effects in the bivariate normal distribution(left panels each) and the bivariate t-distribution (right panels each).

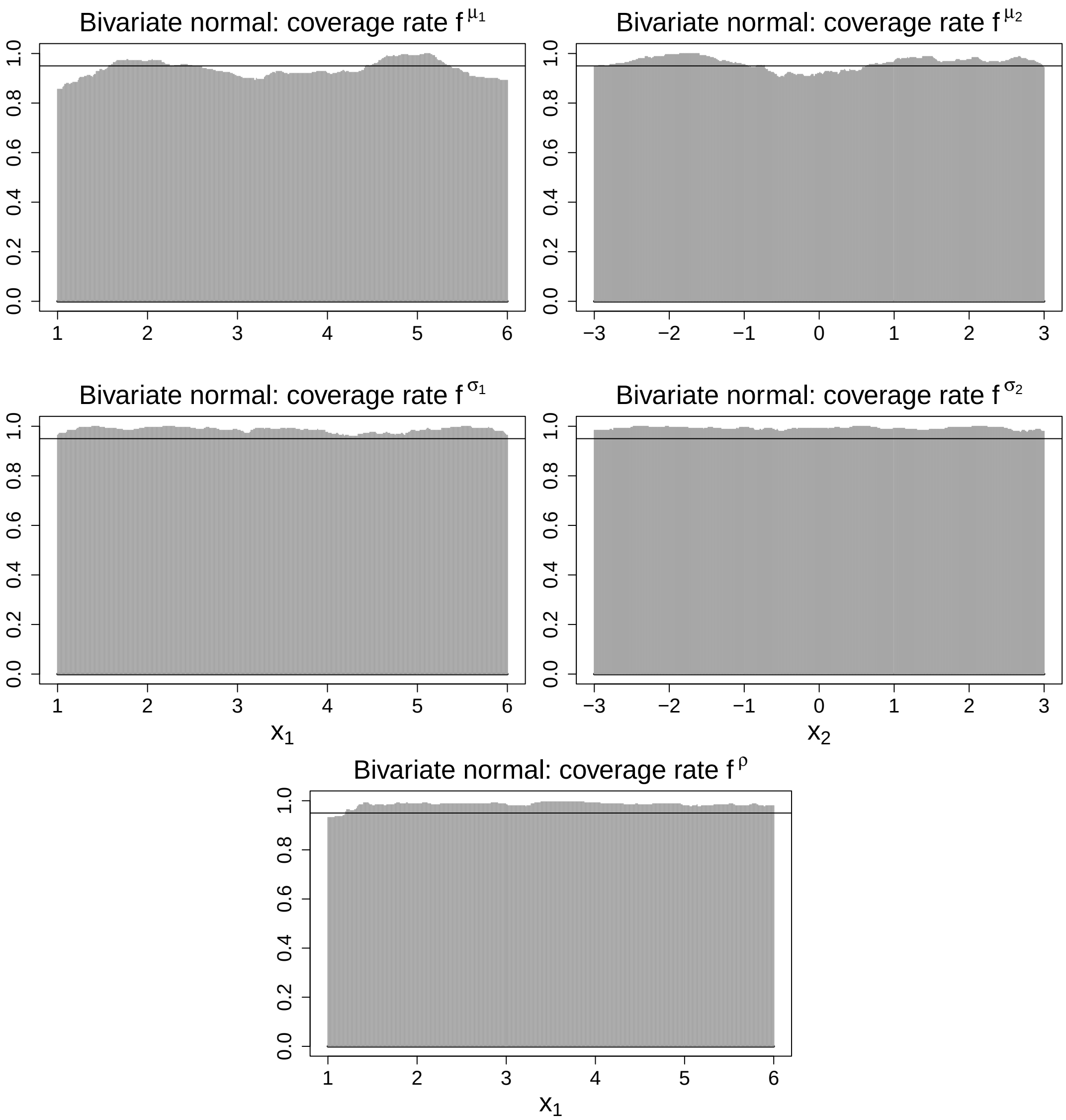

Figure A2: Simulation Study, bivariate normal distribution. Pointwise coverage probabilities of 95% intervals. The horizontal line displays the 95% mark. Vertical lines display the coverage over all 250 simulated models an on a grid of x_1 or x_2 of size 0.01.

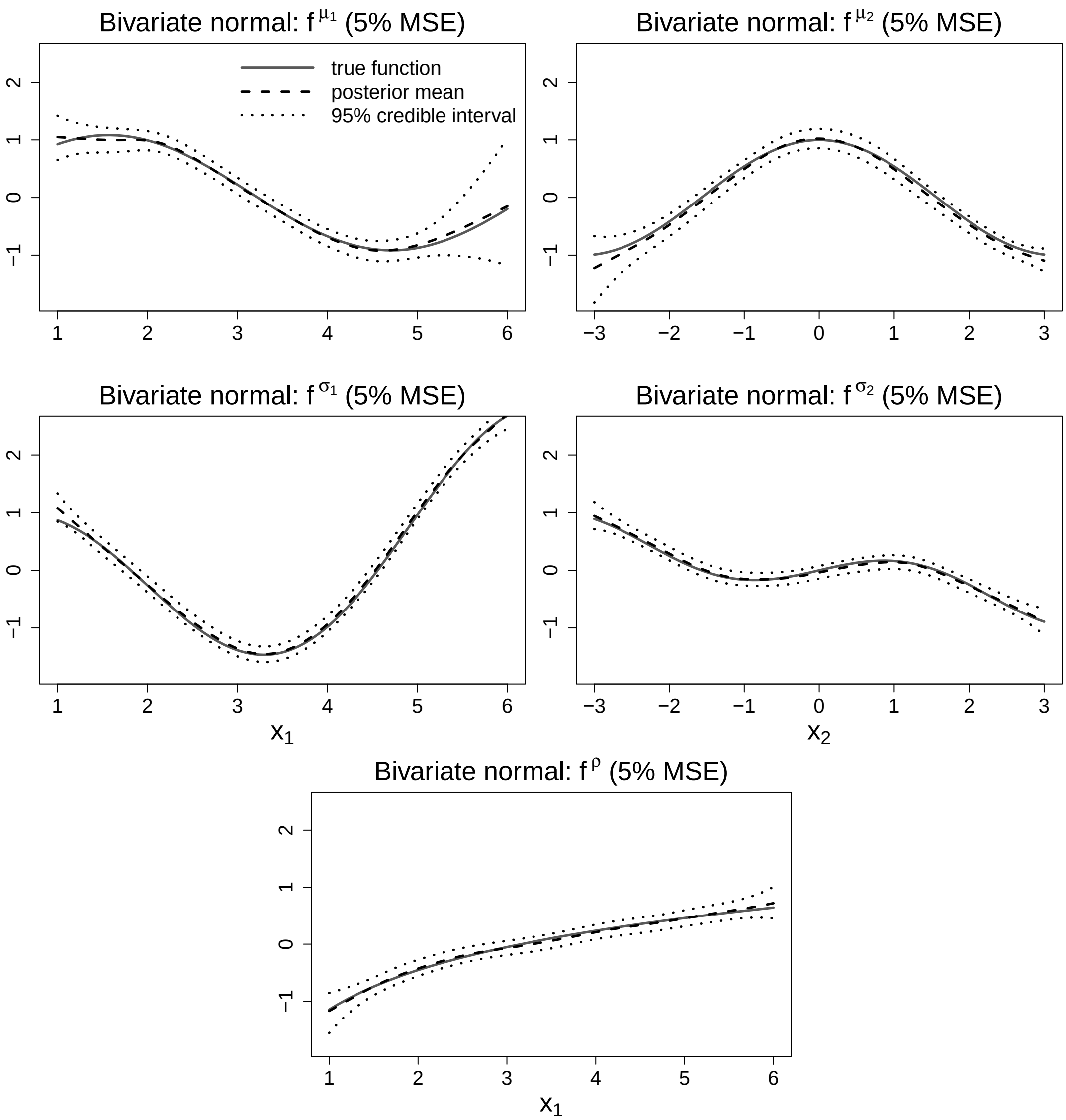

Figure A3: Simulation Study, bivariate normal distribution. Posterior mean estimates (dashed lines) together with 95% credible intervals (dotted lines). The solid lines are the true functions. Estimates were chosen according to their 5% quantile of the MSE distribution obtained from 250 simulation replications.

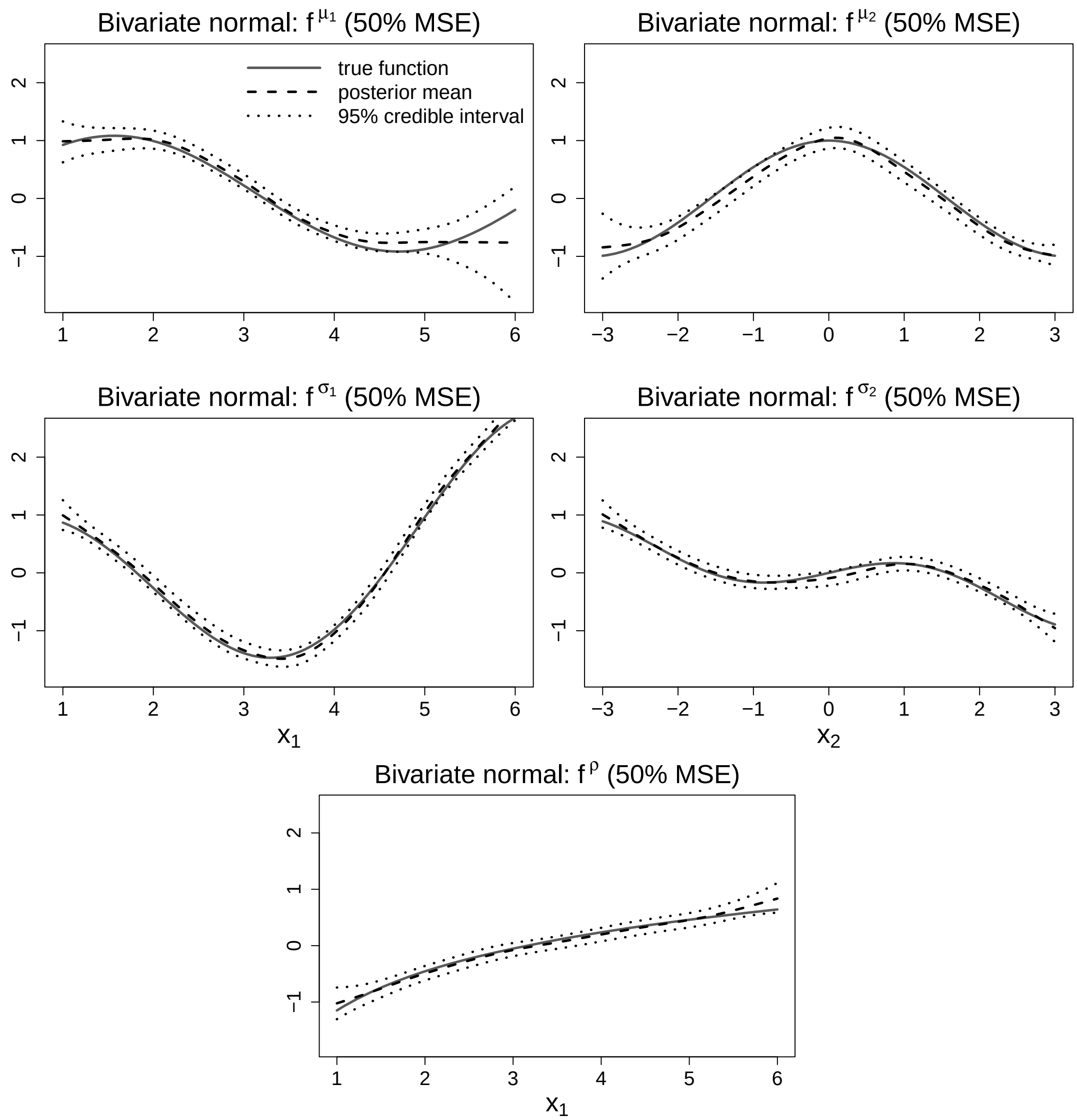

Figure A4: Simulation Study, bivariate normal distribution. Posterior mean estimates (dashed lines) together with 95% credible intervals (dotted lines). The solid lines are the true functions. Estimates were chosen according to their 50% quantile of the MSE distribution obtained from 250 simulation replications.

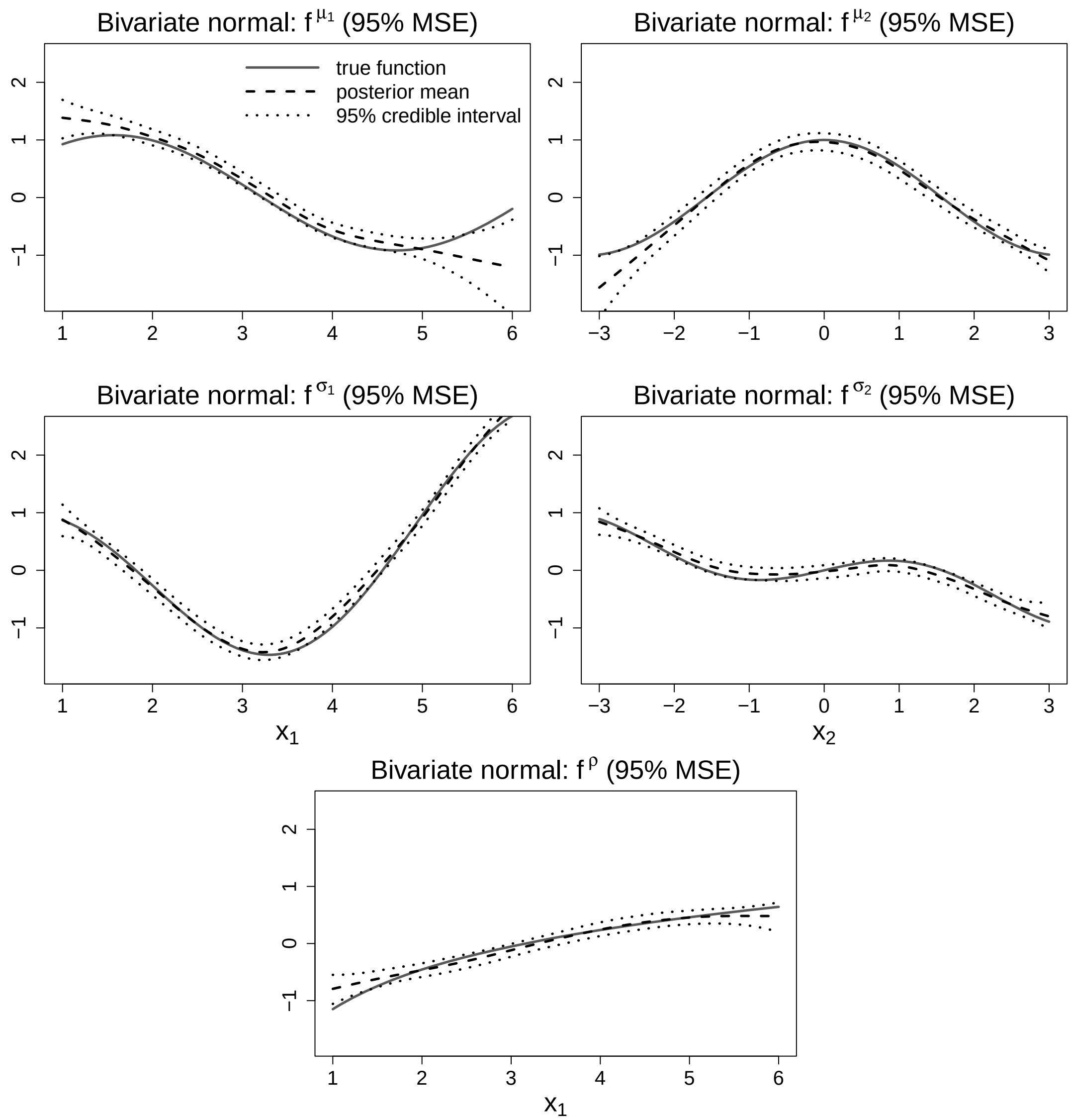

Figure A5: Simulation Study, bivariate normal distribution. Posterior mean estimates (dashed lines) together with 95% credible intervals (dotted lines). The solid lines are the true functions. Estimates were chosen according to their 95% quantile of the MSE distribution obtained from 250 simulation replications.

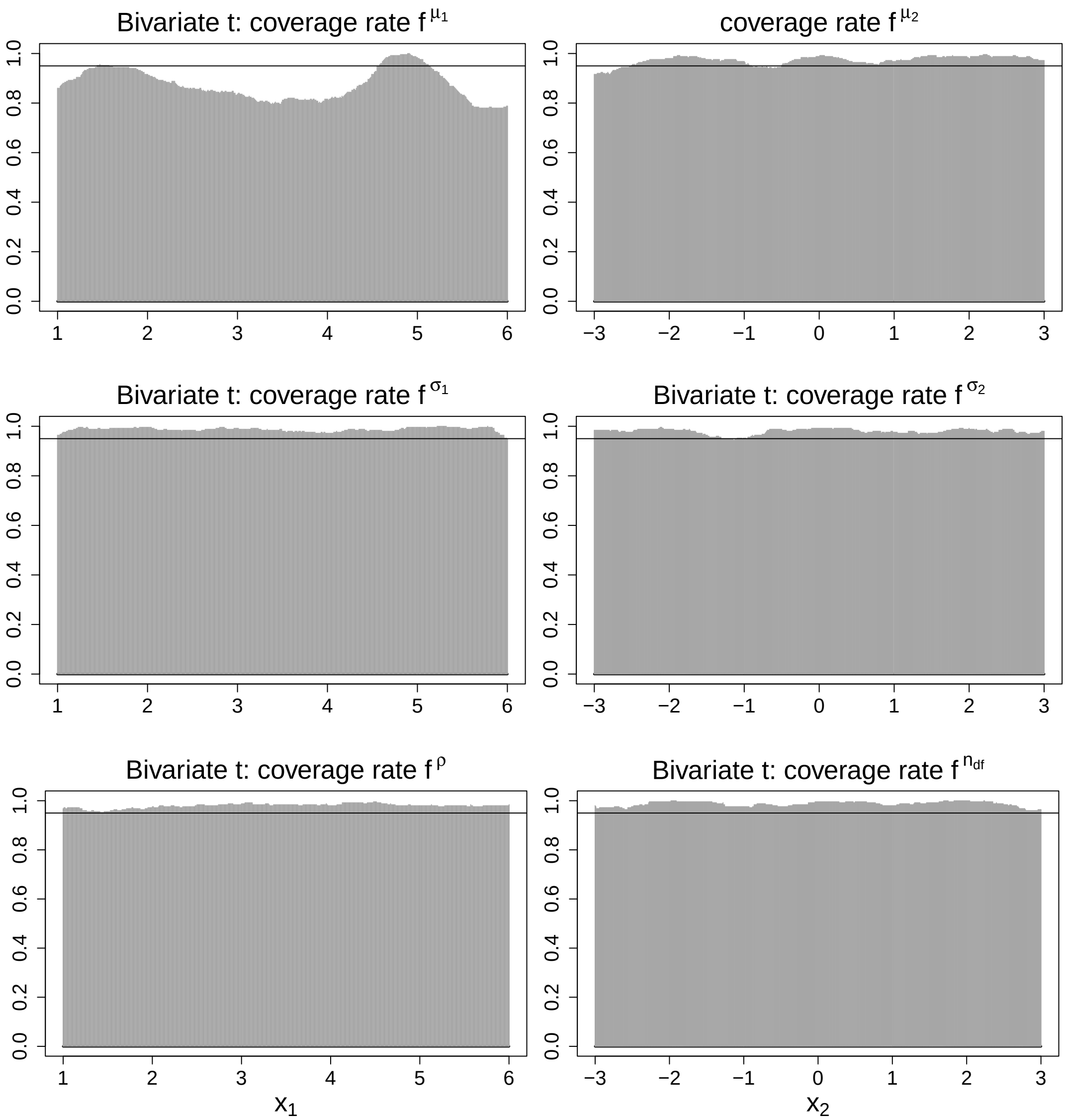

Figure A6: Simulation Study, bivariate t-distribution. Pointwise 95% coverage rates. The horizontal line displays the 95% mark. Vertical lines display the coverage over all 250 simulated models and on a grid of step size 0.01 within the range of x_1 or x_2.

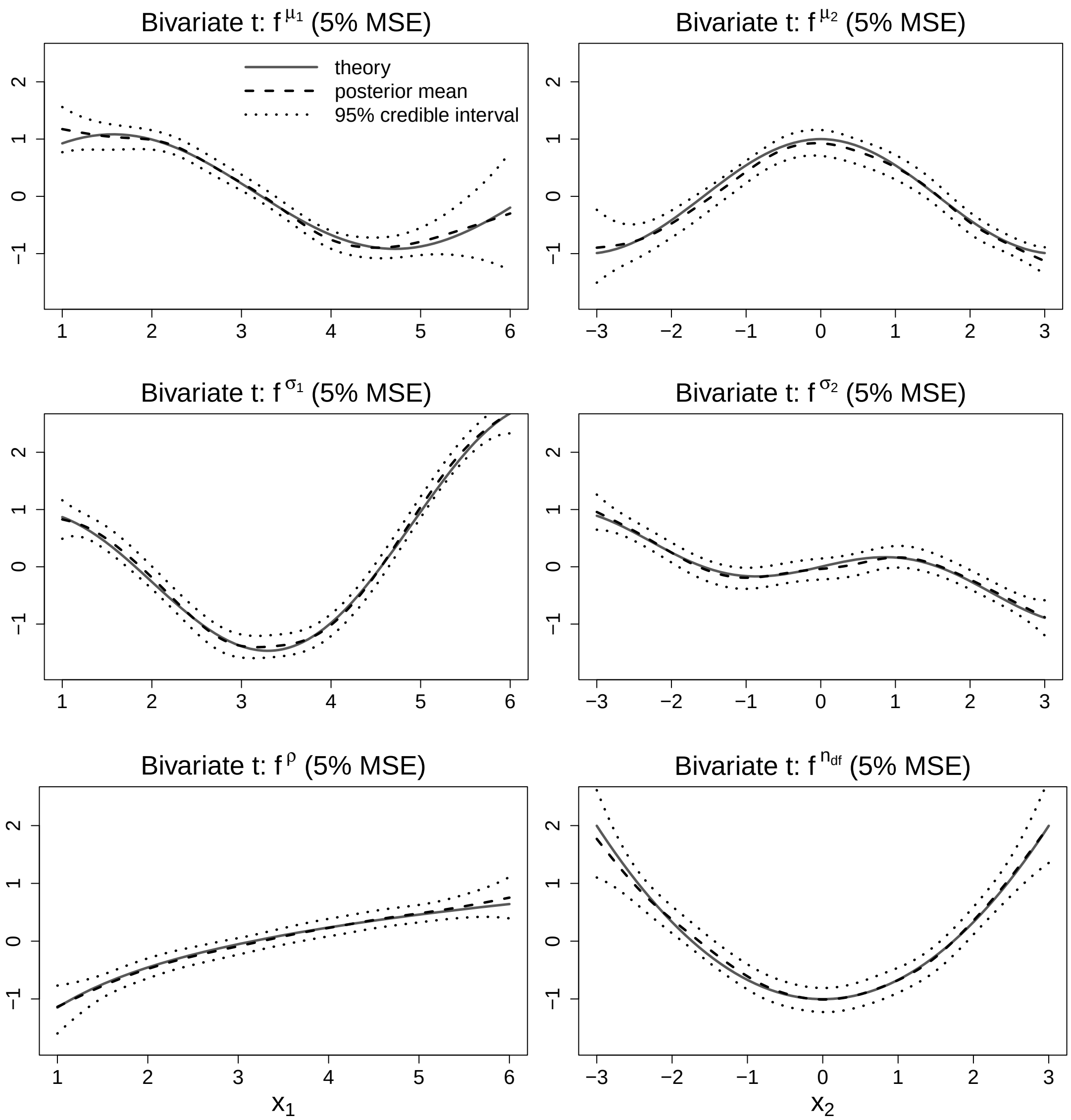

Figure A7: Simulation Study, bivariate t-distribution. Posterior mean estimates (dashed lines) together with 95% credible intervals (dotted lines). The solid lines are the true functions. Estimates were chosen according to their 5% quantile of the MSE distribution obtained from 250 simulation replications.

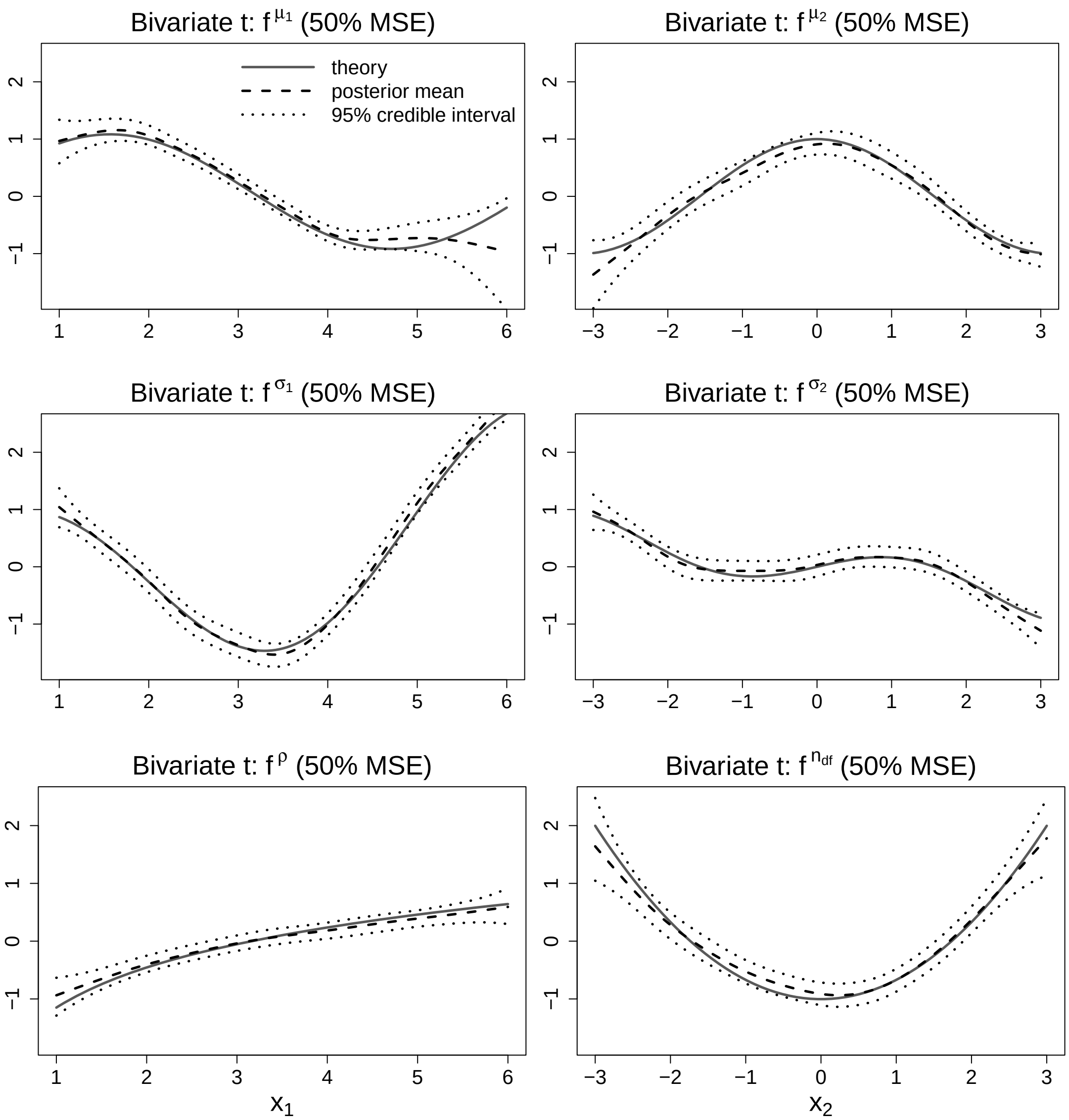

Figure A8: Simulation Study, bivariate t-distribution. Posterior mean estimates (dashed lines) together with 95% credible intervals (dotted lines). The solid lines are the true functions. Estimates were chosen according to their 50% quantile of the MSE distribution obtained from 250 simulation replications.

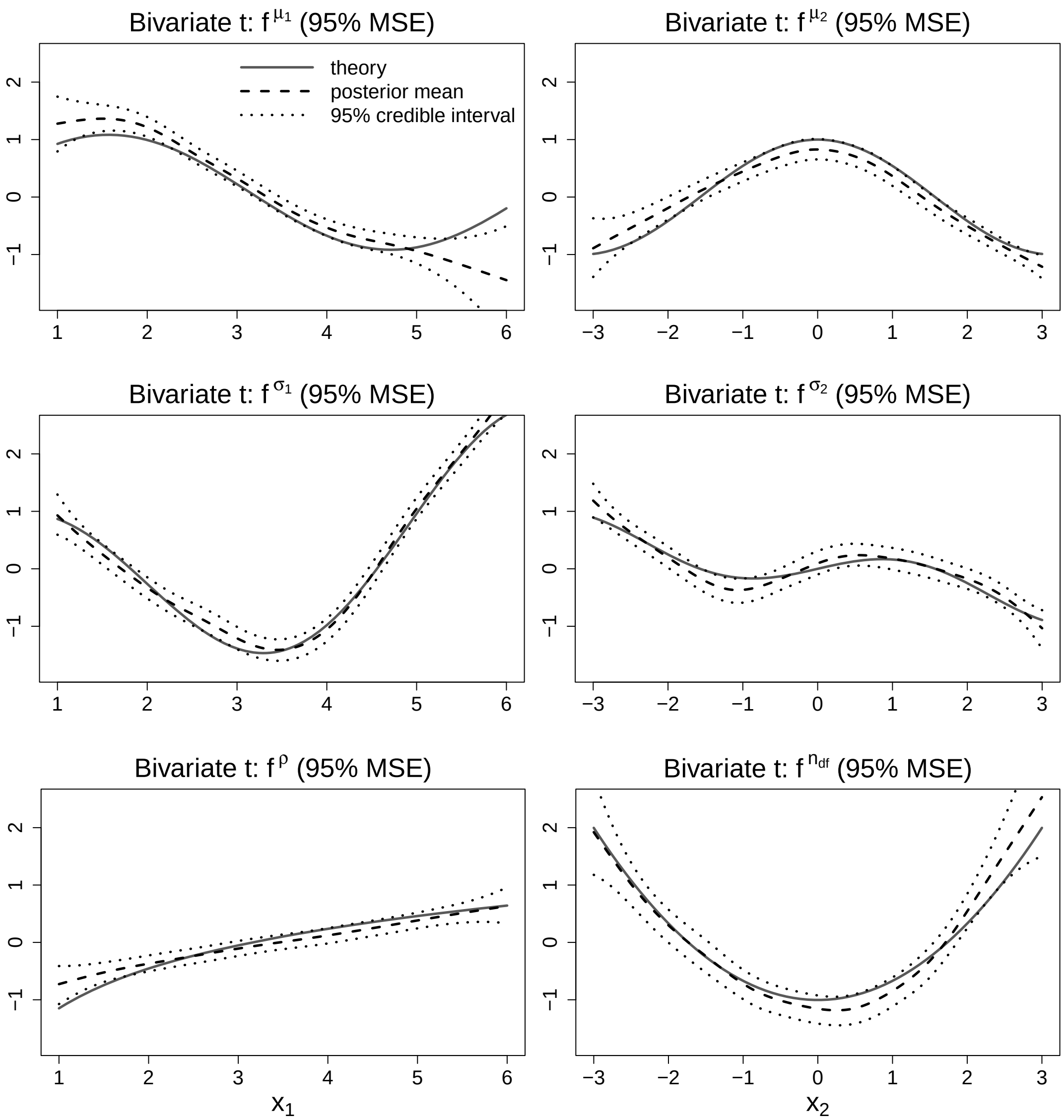

Figure A9: Simulation Study, bivariate t-distribution. Posterior mean estimates (dashed lines) together with 95% credible intervals (dotted lines). The solid lines are the true functions. Estimates were chosen according to their 95% quantile of the MSE distribution obtained from 250 simulation replications.

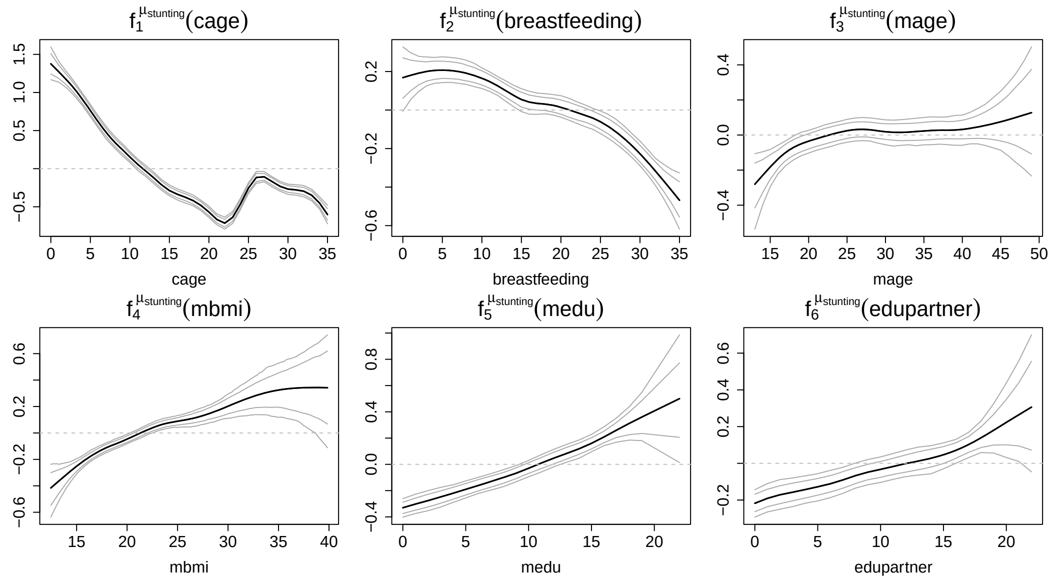

Figure B10: Childhood malnutrition, bivariate t-distribution. Posterior mean estimates of nonparametric effects on $\mu_{stunting}$ together with 80% and 95% pointwise credible intervals, centred around zero each.

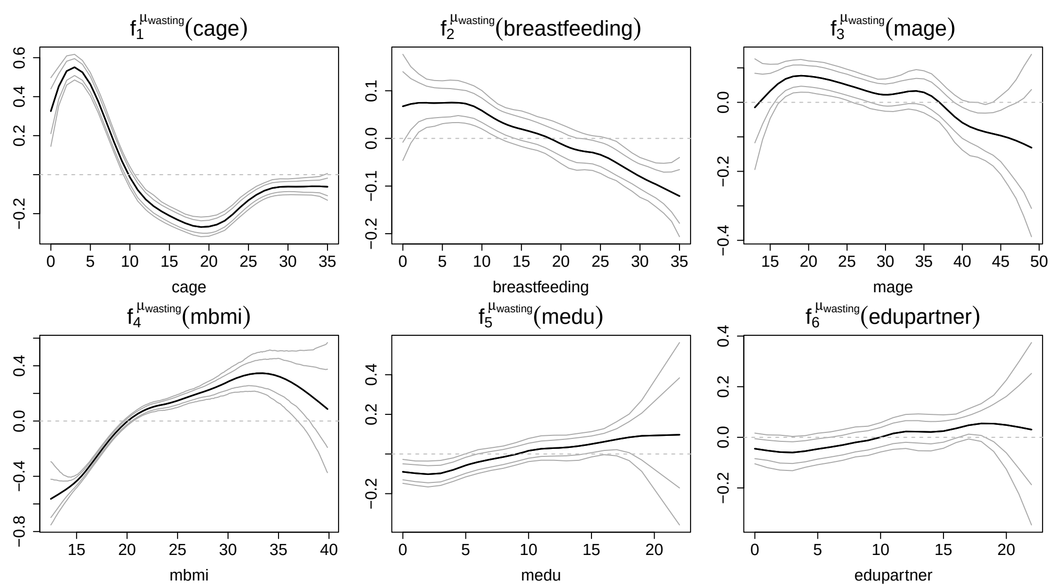

Figure B11: Childhood malnutrition, bivariate t-distribution. Posterior mean estimates of nonparametric effects on $\mu_{wasting}$ together with 80% and 95% pointwise credible intervals, centred around zero each.

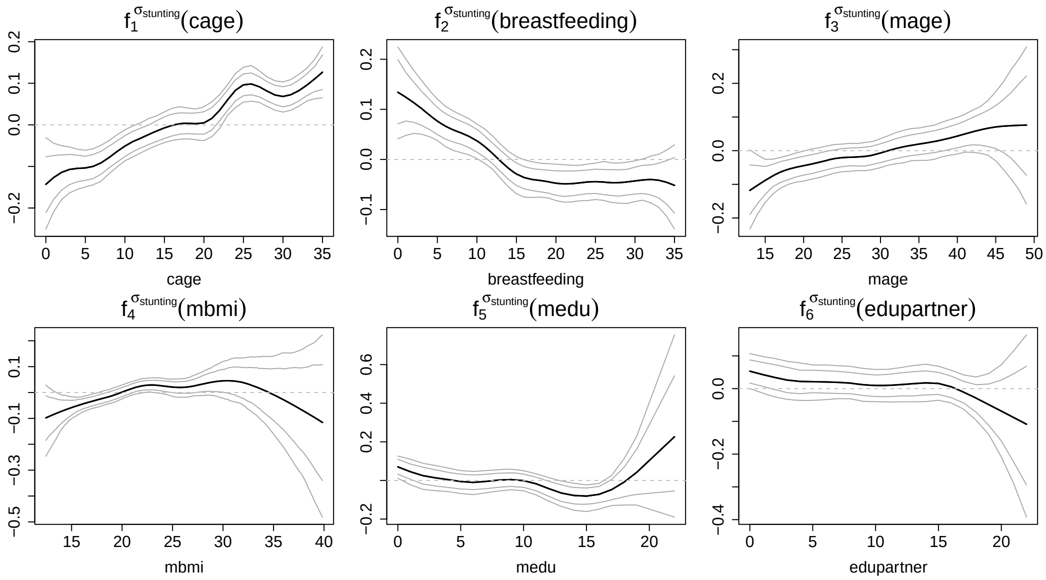

Figure B12: Childhood malnutrition, bivariate t-distribution. Posterior mean estimates of nonparametric effects on $\sigma_{stunting}$ together with 80% and 95% pointwise credible intervals, centred around zero each.

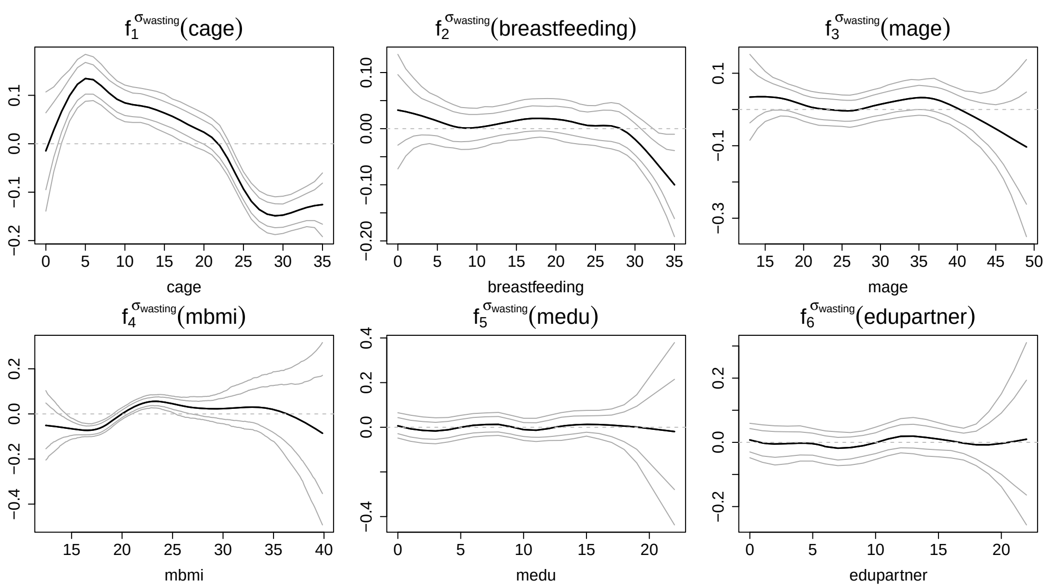

Figure B13: Childhood malnutrition, bivariate t-distribution. Posterior mean estimates of nonparametric effects on $\sigma_{wasting}$ together with 80% and 95% pointwise credible intervals, centred around zero each.

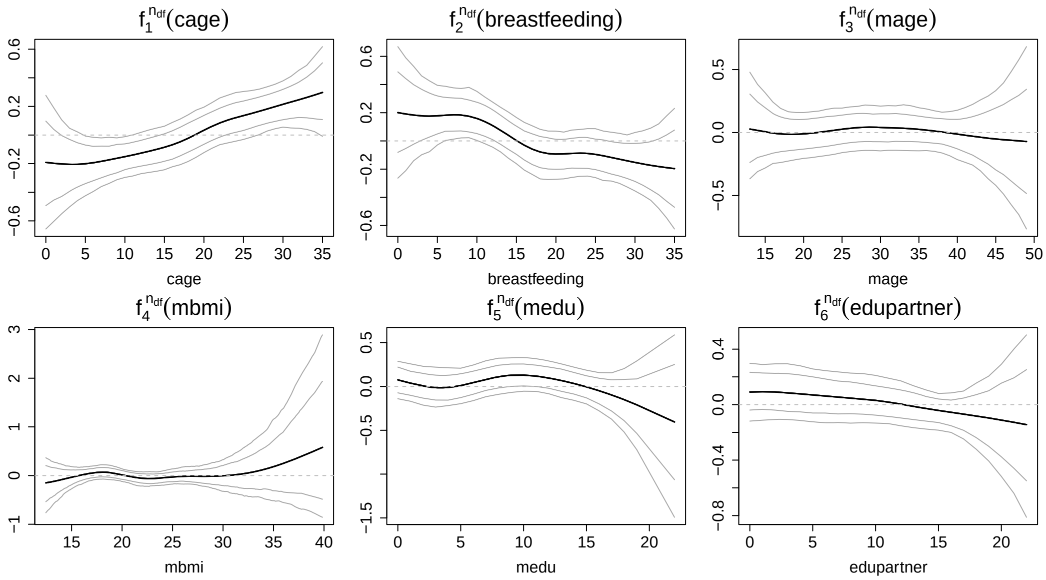

Figure B14: Childhood malnutrition, bivariate t-distribution. Posterior mean estimates of nonparametric effects on n_{df} together with 80% and 95% pointwise credible intervals, centred around zero each.

References

Bundesinstitut fuer Bau-, Stadt- und Raumforschung (BBSR): Regierungsbezirke, GeoBasis-BKG, 2002.

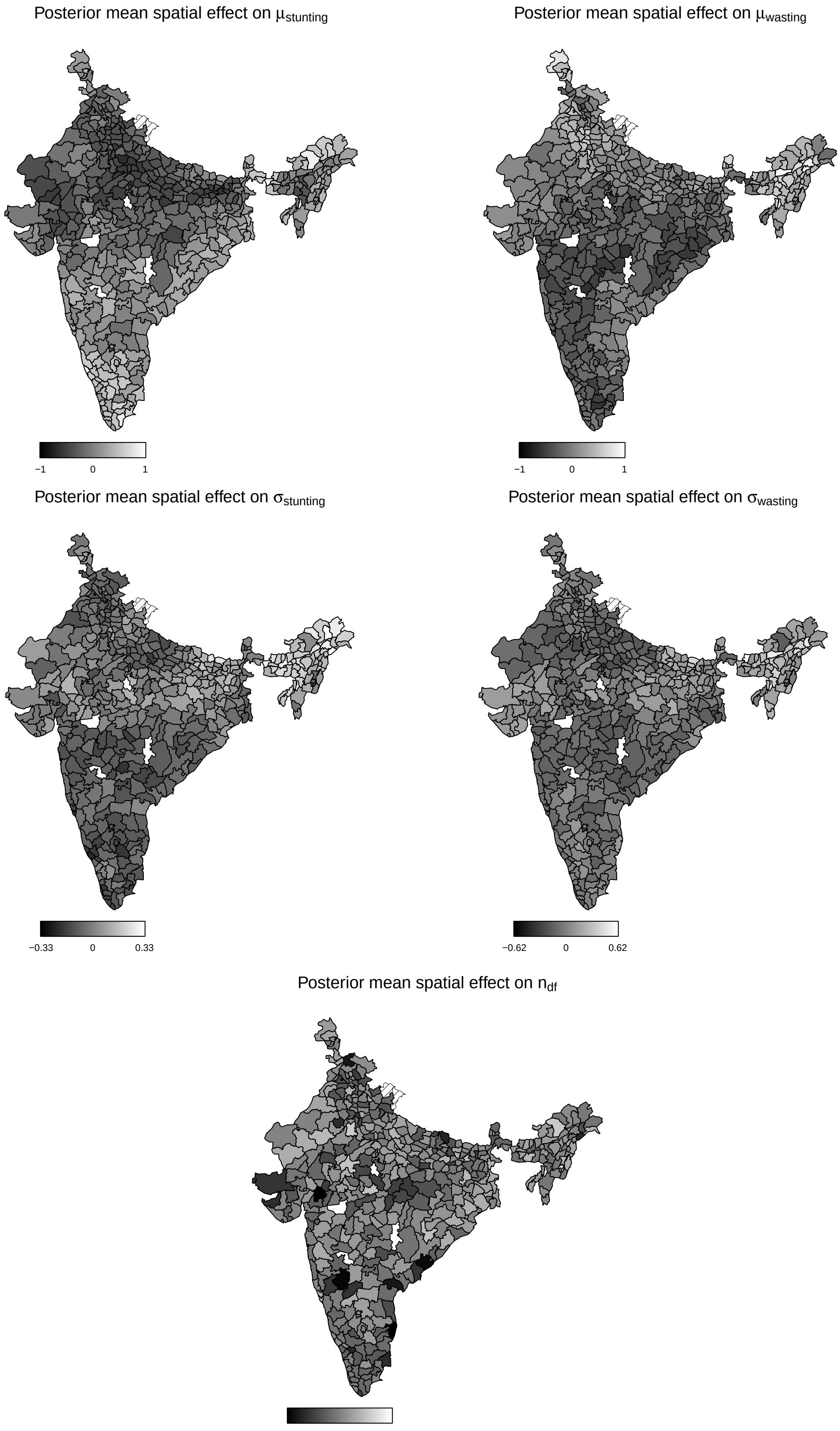

Figure B15: Childhood malnutrition, bivariate t-distribution. Posterior mean estimates of the complete spatial effects on $\mu_{stunting}$, $\mu_{wasting}$, $\sigma_{stunting}$ and $\sigma_{wasting}$, centred around zero each.

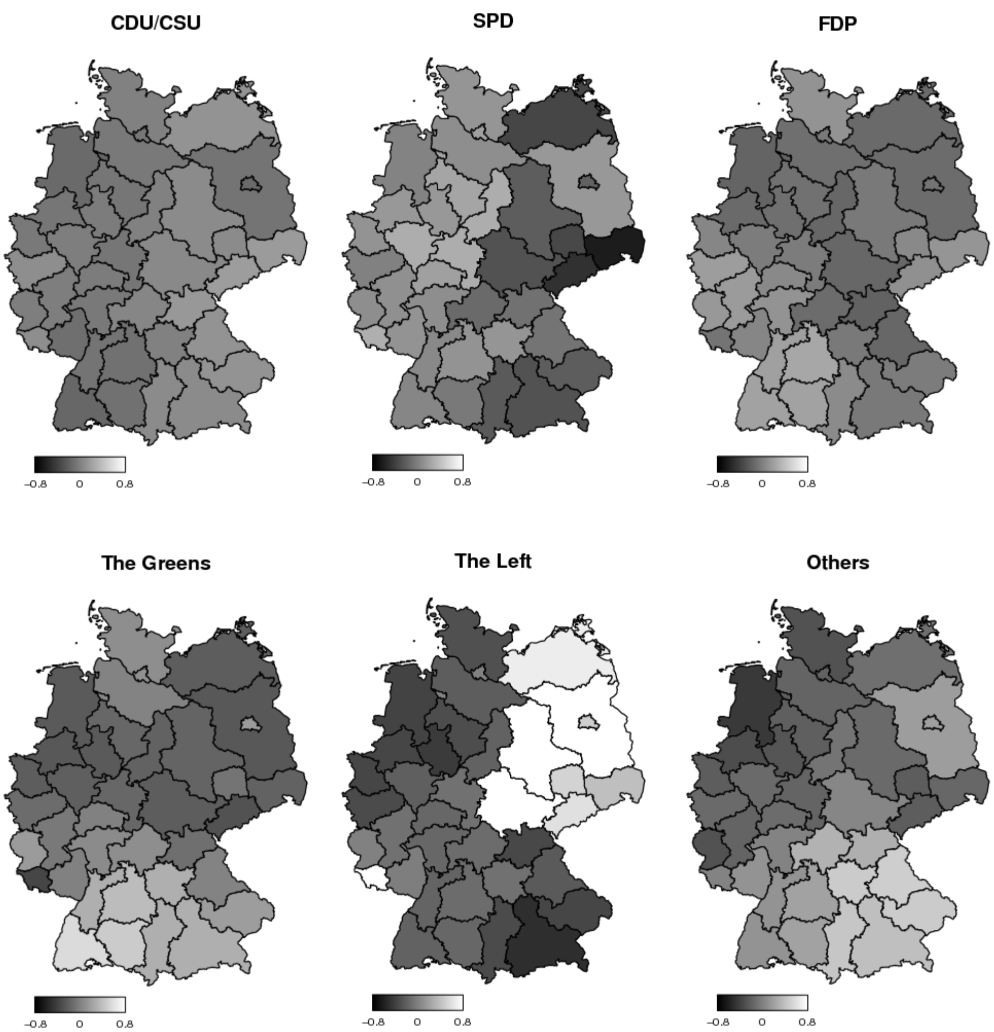

Figure C16: Federal election. Posterior mean estimates of spatial effects, centred around zero each. Geometric information provided by Bundesinstitut fuer Bau-, Stadt- und Raumforschung [BBSR]

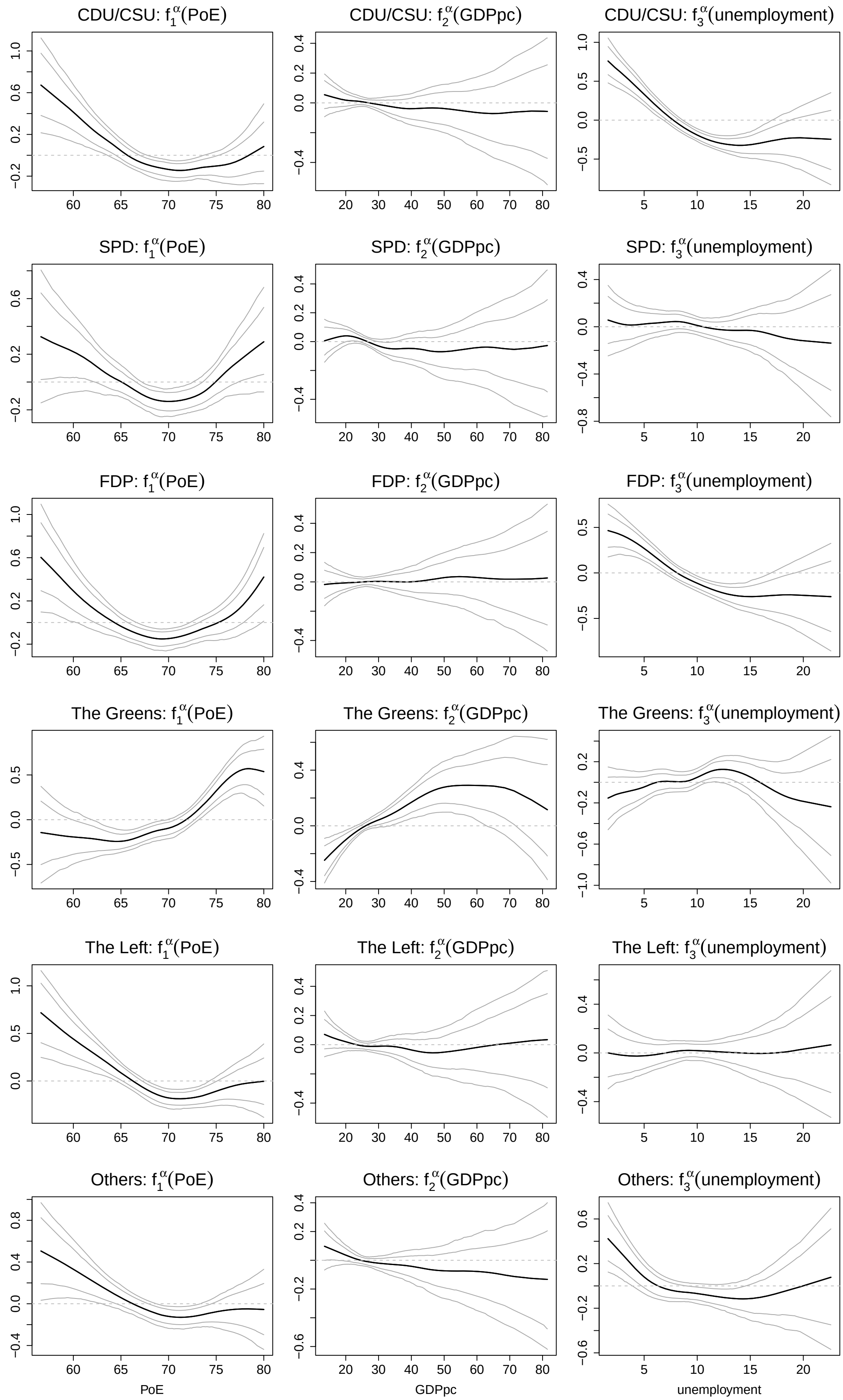

Figure C17: Federal election. Posterior mean estimates of nonparametric effects together with 80% and 95% pointwise credible intervals, centred around zero each.

University of Innsbruck - Working Papers in Economics and Statistics
Recent Papers can be accessed on the following webpage:

http://eeecon.uibk.ac.at/wopec/

2013-35 **Nadja Klein, Thomas Kneib, Stephan Klasen, Stefan Lang:** Bayesian structured additive distributional regression for multivariate responses

2013-34 **Sylvia Kaufmann, Johann Scharler:** Bank-lending standards, loan growth and the business cycle in the Euro area

2013-33 **Ting Wang, Edgar C. Merkle, Achim Zeileis:** Score-based tests of measurement invariance: Use in practice

2013-32 **Jakob W. Messner, Georg J. Mayr, Daniel S. Wilks, Achim Zeileis:** Extending extended logistic regression for ensemble post-processing: Extended vs. separate vs. ordered vs. censored

2013-31 **Anita Gantner, Kristian Horn, Rudolf Kerschbamer:** Fair division in unanimity bargaining with subjective claims

2013-30 **Anita Gantner, Rudolf Kerschbamer:** Fairness and efficiency in a subjective claims problem

2013-29 **Tanja Hörtnagl, Rudolf Kerschbamer, Rudi Stracke, Uwe Sunde:** Heterogeneity in rent-seeking contests with multiple stages: Theory and experimental evidence

2013-28 **Dominik Erharter:** Promoting coordination in summary-statistic games

2013-27 **Dominik Erharter:** Screening experts' distributional preferences

2013-26 **Loukas Balafoutas, Rudolf Kerschbamer, Matthias Sutter:** Second-degree moral hazard in a real-world credence goods market

2013-25 **Rudolf Kerschbamer:** The geometry of distributional preferences and a non-parametric identification approach

2013-24 **Nadja Klein, Michel Denuit, Stefan Lang, Thomas Kneib:** Nonlife ratemaking and risk management with bayesian additive models for location, scale and shape

2013-23 **Nadja Klein, Thomas Kneib, Stefan Lang:** Bayesian structured additive distributional regression

2013-22 **David Plavcan, Georg J. Mayr, Achim Zeileis:** Automatic and probabilistic foehn diagnosis with a statistical mixture model

- 2013-21 **Jakob W. Messner, Georg J. Mayr, Achim Zeileis, Daniel S. Wilks:** Extending extended logistic regression to effectively utilize the ensemble spread
- 2013-20 **Michael Greinecker, Konrad Podczeck:** Liapounoff's vector measure theorem in Banach spaces *forthcoming in Economic Theory Bulletin*
- 2013-19 **Florian Lindner:** Decision time and steps of reasoning in a competitive market entry game
- 2013-18 **Michael Greinecker, Konrad Podczeck:** Purification and independence
- 2013-17 **Loukas Balafoutas, Rudolf Kerschbamer, Martin Kocher, Matthias Sutter:** Revealed distributional preferences: Individuals vs. teams
- 2013-16 **Simone Gobien, Björn Vollan:** Playing with the social network: Social cohesion in resettled and non-resettled communities in Cambodia
- 2013-15 **Björn Vollan, Sebastian Prediger, Markus Frölich:** Co-managing common pool resources: Do formal rules have to be adapted to traditional ecological norms?
- 2013-14 **Björn Vollan, Yexin Zhou, Andreas Landmann, Biliang Hu, Carsten Herrmann-Pillath:** Cooperation under democracy and authoritarian norms
- 2013-13 **Florian Lindner, Matthias Sutter:** Level-k reasoning and time pressure in the 11-20 money request game *forthcoming in Economics Letters*
- 2013-12 **Nadja Klein, Thomas Kneib, Stefan Lang:** Bayesian generalized additive models for location, scale and shape for zero-inflated and overdispersed count data
- 2013-11 **Thomas Stöckl:** Price efficiency and trading behavior in limit order markets with competing insiders *forthcoming in Experimental Economics*
- 2013-10 **Sebastian Prediger, Björn Vollan, Benedikt Herrmann:** Resource scarcity, spite and cooperation
- 2013-09 **Andreas Exenberger, Simon Hartmann:** How does institutional change coincide with changes in the quality of life? An exemplary case study
- 2013-08 **E. Glenn Dutcher, Loukas Balafoutas, Florian Lindner, Dmitry Ryvkin, Matthias Sutter:** Strive to be first or avoid being last: An experiment on relative performance incentives.
- 2013-07 **Daniela Glätzle-Rützler, Matthias Sutter, Achim Zeileis:** No myopic loss aversion in adolescents? An experimental note
- 2013-06 **Conrad Kobel, Engelbert Theurl:** Hospital specialisation within a DRG-Framework: The Austrian case

2013-05 **Martin Halla, Mario Lackner, Johann Scharler:** Does the welfare state destroy the family? Evidence from OECD member countries

2013-04 **Thomas Stöckl, Jürgen Huber, Michael Kirchler, Florian Lindner:** Hot hand belief and gambler's fallacy in teams: Evidence from investment experiments

2013-03 **Wolfgang Luhan, Johann Scharler:** Monetary policy, inflation illusion and the Taylor principle: An experimental study

2013-02 **Esther Blanco, Maria Claudia Lopez, James M. Walker:** Tensions between the resource damage and the private benefits of appropriation in the commons

2013-01 **Jakob W. Messner, Achim Zeileis, Jochen Broecker, Georg J. Mayr:** Improved probabilistic wind power forecasts with an inverse power curve transformation and censored regression

2012-27 **Achim Zeileis, Nikolaus Umlauf, Friedrich Leisch:** Flexible generation of e-learning exams in R: Moodle quizzes, OLAT assessments, and beyond

2012-26 **Francisco Campos-Ortiz, Louis Putterman, T.K. Ahn, Loukas Balafoutas, Mongoljin Batsaikhan, Matthias Sutter:** Security of property as a public good: Institutions, socio-political environment and experimental behavior in five countries

2012-25 **Esther Blanco, Maria Claudia Lopez, James M. Walker:** Appropriation in the commons: variations in the opportunity costs of conservation

2012-24 **Edgar C. Merkle, Jinyan Fan, Achim Zeileis:** Testing for measurement invariance with respect to an ordinal variable *forthcoming in Psychometrika*

2012-23 **Lukas Schrott, Martin Gächter, Engelbert Theurl:** Regional development in advanced countries: A within-country application of the Human Development Index for Austria

2012-22 **Glenn Dutcher, Krista Jabs Saral:** Does team telecommuting affect productivity? An experiment

2012-21 **Thomas Windberger, Jesus Crespo Cuaresma, Janette Walde:** Dirty floating and monetary independence in Central and Eastern Europe - The role of structural breaks

2012-20 **Martin Wagner, Achim Zeileis:** Heterogeneity of regional growth in the European Union

2012-19 **Natalia Montinari, Antonio Nicolo, Regine Oexl:** Mediocrity and induced reciprocity

2012-18 **Esther Blanco, Javier Lozano:** Evolutionary success and failure of wildlife conservancy programs

2012-17 **Ronald Peeters, Marc Vorsatz, Markus Walzl:** Beliefs and truth-telling: A laboratory experiment

2012-16 **Alexander Sebald, Markus Walzl:** Optimal contracts based on subjective evaluations and reciprocity

2012-15 **Alexander Sebald, Markus Walzl:** Subjective performance evaluations and reciprocity in principal-agent relations

2012-14 **Elisabeth Christen:** Time zones matter: The impact of distance and time zones on services trade

2012-13 **Elisabeth Christen, Joseph Francois, Bernard Hoekman:** CGE modeling of market access in services

2012-12 **Loukas Balafoutas, Nikos Nikiforakis:** Norm enforcement in the city: A natural field experiment *forthcoming in European Economic Review*

2012-11 **Dominik Erharter:** Credence goods markets, distributional preferences and the role of institutions

2012-10 **Nikolaus Umlauf, Daniel Adler, Thomas Kneib, Stefan Lang, Achim Zeileis:** Structured additive regression models: An R interface to BayesX

2012-09 **Achim Zeileis, Christoph Leitner, Kurt Hornik:** History repeating: Spain beats Germany in the EURO 2012 Final

2012-08 **Loukas Balafoutas, Glenn Dutcher, Florian Lindner, Dmitry Ryvkin:** The optimal allocation of prizes in tournaments of heterogeneous agents

2012-07 **Stefan Lang, Nikolaus Umlauf, Peter Wechselberger, Kenneth Harttgen, Thomas Kneib:** Multilevel structured additive regression

2012-06 **Elisabeth Waldmann, Thomas Kneib, Yu Ryan Yu, Stefan Lang:** Bayesian semiparametric additive quantile regression

2012-05 **Eric Mayer, Sebastian Rueth, Johann Scharler:** Government debt, inflation dynamics and the transmission of fiscal policy shocks *forthcoming in Economic Modelling*

2012-04 **Markus Leibrecht, Johann Scharler:** Government size and business cycle volatility; How important are credit constraints? *forthcoming in Economica*

2012-03 **Uwe Dulleck, David Johnston, Rudolf Kerschbamer, Matthias Sutter:** The good, the bad and the naive: Do fair prices signal good types or do they induce good behaviour?

2012-02 **Martin G. Kocher, Wolfgang J. Luhan, Matthias Sutter:** Testing a forgotten aspect of Akerlof's gift exchange hypothesis: Relational contracts with individual and uniform wages

2012-01 **Loukas Balafoutas, Florian Lindner, Matthias Sutter:** Sabotage in tournaments: Evidence from a natural experiment *published in* *Kyklos*

University of Innsbruck

Working Papers in Economics and Statistics

2013-35

Nadja Klein, Thomas Kneib, Stephan Klasen, Stefan Lang

Bayesian structured additive distributional regression for multivariate responses

Abstract
In this paper, we propose a unified Bayesian approach for multivariate structured additive distributional regression analysis where inference is applicable to a huge class of multivariate response distributions, comprising continuous, discrete and latent models, and where each parameter of these potentially complex distributions is modelled by a structured additive predictor. The latter is an additive composition of different types of covariate effects e.g. nonlinear effects of continuous variables, random effects, spatial variations, or interaction effects. Inference is realised by a generic, efficient Markov chain Monte Carlo algorithm based on iteratively weighted least squares approximations and with multivariate Gaussian priors to enforce specific properties of functional effects. Examples will be given by illustrations on analysing the joint model of risk factors for chronic and acute childhood malnutrition in India and on ecological regression for German election results.

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)