General Multivariate Effect Priors in Recursive Bivariate Gaussian Models

Hauke Thaden

Georg-August University of Goettingen

hauke.thaden@uni-goettingen.de

February 28, 2017

Abstract

Modeling complex relationships and interactions between variables is an ongoing statistical challenge. In particular, the joint modeling of multiple response variables has recently gained interest among methodological and applied researchers. In this article, we contribute to this development by incorporating semiparametric predictors into recursive simultaneous equation models. In particular, we extend the existing framework by imposing effect priors that account for correlation of the effects across equations. This idea can be seen as a generalization of multivariate conditional autoregressive priors used for the analysis of multivariate spatial data.
We implement a Gibbs sampler for the estimation and evaluate the model in an elaborate simulation study. Finally, we illustrate the applicability of our approach with real data examples on malnutrition in Asia and Africa as well as the analysis of plant and species richness with respect to environmental diversity.

Keywords: simultaneous equation models, correlated effects, semiparametric predictors

1 Introduction

Joint modeling of multiple response variables has recently gained rising popularity in statistical research. Examples include the development of models for multivariate responses in the context of distributional regression (Klein and Kneib, 2016), joint modeling of e.g. survival and longitudinal data (Waldmann et al., 2017) or extensions of simultaneous equation models (for example Thaden and Kneib, 2017, and Thaden et al., 2017). Furthermore, not only different compositions of response

variables, but also flexibility concerning the type of covariate effects constitutes an important field of current research. In particular, the interest in regression models with structured additive predictors including linear, nonlinear, random or spatial effects increased, both from a methodological and applied perspective (e.g. Klein et al., 2015 or Pata et al., 2012).

In this article, we contribute to this development by joining the framework of simultaneous equation models (SEM) with the flexibility of semiparametric effect types. More precisely, we summarize how different effects can be conceived using a unifying basis function approach (as shown in e.g. Wood, 2006) and integrate those into recursive SEM with bivariate response. In contrast to Song et al. (2013), who focus on modeling nonlinear effects using Bayesian P-splines in SEM with latent variables, we further generalize their approach by introducing additional effect types (i.e. spatial and random effects). Furthermore, in order to capture potential complex correlation structures among the occurring variables, we allow the semiparametric effects to be correlated across equations. This is achieved by choosing appropriate priors for these effects. These priors serve as a fundament for our Markov Chain Monte Carlo (MCMC) estimation procedure. Overall, our approach extends the idea of multivariate conditionally autoregressive (MCAR, see Gelfand and Vounatsou, 2003, for example) spatial effect priors, which has - to the best of our knowledge - neither been generalized to alternative effect types nor integrated into bivariate SEM before.

We will not only evaluate our proposed approach in several simulation scenarios but also apply the method to real data examples representing different semiparametric effects in a typical structured additive predictor. They illustrate the complex underlying relationships arising in various applied areas in which typically different sources of correlation occur. Specifically, we start with the problem of correlated spatial effects in childhood undernutrition in developing countries. Childhood undernutrition is one of the major public health problems in these countries. It is expected that acute undernutrition, wasting, has an effect on chronic undernutrition, stunting. We analyze this question in more detail across countries in Africa and Asia. We will use data from the global health observatory data repository published by the World Health Organization (WHO, 2016).

The second example illustrates the applicability of our approach in an ecological context. We analyze how species richness of plants and animals are interrelated with respect to their environment. The research question is based on the original paper of Jetz et al. (2009) along the sub-dataset applied in Klein and Kneib (2016).

The rest of the paper is structured as follows: In Section 2, we explain our proposed approach in detail. We build up the model starting from simple linear SEM and include more complex predictor structures step by step in several subsections. Our implementation of the Bayesian estimation strategy via Gibbs sampling can be found in Section 3. The simulations in Section 4 evaluate the performance of the proposed method while Section 5 illustrates the applicability in two complex real datasets coming from different applied areas (health and ecology) and dealing with relevant research questions. Finally, Section 6 summarizes our findings and gives possible extensions for future research.

2 Methodology

2.1 Recursive Bivariate Simultaneous Equation Models

Simultaneous equation models (SEM) allow for representing potentially complex relationships between variables in a multivariate setting. More specifically, they consist of multiple regression equations with the additional feature that response variables of one equation (endogenous variables) are allowed to appear as covariates in another equation (hence the term *recursive* SEM). Exogenous variables only appear as covariates in one or more of the equations. We refer to Bollen (1989) for a detailed general overview on SEM.

Figure 1 shows a typical representation of a SEM with one exogenous variable x and two related endogenous variables y_1 and y_2 as a path diagram. Effects between variables are illustrated as arrows.

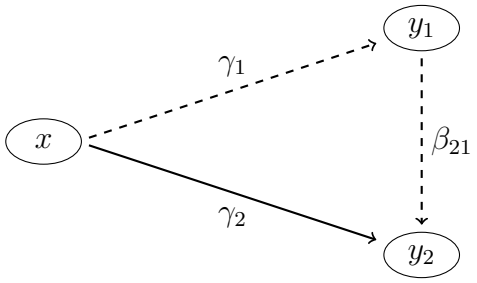

Figure 1: Path diagram of a simple linear recursive SEM. The exogenous variable x simultaneously influences the endogenous variables y_1 and y_2 via γ_1 and γ_2, respectively. Additionally y_1 has a direct effect β_{21} on y_2.

In this simple linear setting, Figure 1 translates to the simultaneous equations

$$y_1 = \gamma_1 x + \varepsilon_1 \tag{1}$$

$$y_2 = \gamma_2 x + \beta_{21} y_1 + \varepsilon_2, \tag{2}$$

where γ_1, γ_2 and β_{21} are the linear regression coefficients and ε_1 and ε_2 are the error terms within the equations, respectively. In this article, we investigate the performance of more flexible predictor structures in order to overcome the - in many practical applications unrealistic - assumption of linearity. More precisely, we allow the exogenous variable to have some functional influence on the endogenous variables. We generalize (1) and (2) and obtain

$$y_1 = f^{(1)}(x) + \varepsilon_1 \tag{1'}$$

$$y_2 = f^{(2)}(x) + \beta_{21} y_1 + \varepsilon_2. \tag{2'}$$

The unknown functions $f^{(1)}$ and $f^{(2)}$ are not necessarily continuous. They can represent different types of effects such as nonlinear, spatial or random effects. Figure 2 illustrates this extension. In the subsequent section, we explain how the unknown functions $f^{(1)}$ and $f^{(2)}$ can be approximated via basis functions. Additionally, we further generalize the approach by allowing these functions to be correlated. By this, we extend the framework of classical linear recursive bivariate SEM by allowing for semiparametric predictors with general correlation structure.

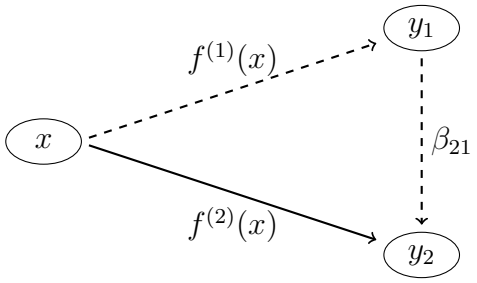

Figure 2: Illustration of a nonlinear simultaneous equation model with two endogenous variables y_1 and y_2. The exogenous variable x is allowed to simultaneously affect both endogenous variables via the unknown functions $f^{(1)}(x)$ and $f^{(2)}(x)$, respectively.

2.2 Effect Specific Basis Function Representations

In this section, we explain how multiple effect types can be expressed in the unifying semiparametric framework of a linear basis function approach (see Wood, 2006 for a detailed overview). For a simple start, assume that as in (2') y_2 is affected by x via an unknown function f, i.e. for each observation $i = 1, \ldots, n$,

$$y_{2i} = f^{(2)}(x_i) + \varepsilon_{2i}$$

holds. The unknown function f is approximated as a linear combination of effect-specific basis functions, namely

$$f^{(2)}(x_i) = \sum_{l=1}^{L} \alpha_{2l} B_l(x_i), \tag{3}$$

such that (3) can be written as

$$f^{(2)}(\mathbf{x}) = \mathbf{B}\boldsymbol{\alpha}_2, \tag{4}$$

with $\mathbf{x} = (x_1, \ldots, x_n)'$ and coefficient vector $\boldsymbol{\alpha}_2 = (\alpha_{21}, \ldots, \alpha_{2L})'$. Above, $\mathbf{B}$ is the design matrix with entries $\mathbf{B}[i, l] = B_l(x_i)$ for $i = 1, \ldots, n$ and $l = 1, \ldots, L$. The choice of basis functions depends on the effect type. Widely used examples include:

- *Linear effects:* The basis functions simply correspond to the observations $x_1, \ldots, x_n$. Consequently, $\mathbf{B} = \mathbf{x}$ reduces to an $n \times 1$ design vector.

- *Continuous effects:* A common way to include smooth functions for univariate covariates are B-spline basis functions (see Eilers and Marx, 1996), such that $\mathbf{B}$ contains the basis functions evaluated at the observed data points.

- *Discrete spatial effects:* For discrete spatial information (e.g. administrative regional units) the design matrix corresponds to an indicator matrix capturing which region each observation is located in.

- *Random effects:* Random effects are used to model group or individual specific effects. Similar to discrete spatial effects, the design matrix $\mathbf{B}$ indicates which group or individual the observations belong to.

Along with the choice of suitable basis functions, the resulting regression coefficients are usually regularized in order to ensure a certain smoothness (e.g. across space or of nonlinear functions) or to avoid overfitting. From a frequentist perspective, this regularization is obtained by penalizing the coefficients during the estimation (as explained in Wood, 2006).

Basic Prior Structures

In the Bayesian formulation - as in our case - the desired smoothness is obtained by choosing appropriate Gaussian priors of the form

$$p(\boldsymbol{\alpha}_2 \mid \tau_2^2) \propto \exp\left(-\frac{1}{2\tau_2^2}\boldsymbol{\alpha}_2'\mathbf{K}\boldsymbol{\alpha}_2\right), \tag{5}$$

where τ_2^2 is the smoothing variance replacing the function of a penalty parameter in a penalized likelihood approach. Similar to the basis functions, the precision matrix $\mathbf{K}$ is effect specific and in some cases rank deficient. For the above mentioned examples of effect types, the corresponding precision matrices are:

- *Linear effects:* Either $\mathbf{K} = 0$ for no smoothing or $\mathbf{K} = 1$ which corresponds to a ridge penalization from a frequentist perspective.

- *Nonlinear effects:* The Bayesian analogue of penalized B-splines or P-splines is obtained by choosing $\mathbf{K} = \mathbf{D}'\mathbf{D}$, where $\mathbf{D}$ is the difference matrix of

desired order (frequent choices are first or second order differences). Consequently, the prior is partially improper as the rank of $\mathbf{K}$ is reduced by the order of the difference approach (again, see Eilers and Marx, 1996 or Lang and Brezger, 2004 for details).

- *Discrete spatial effects:* When analyzing spatially structured data, a commonly made assumption is that nearby observations are more similar than observations far apart. This idea is reflected by choosing $\mathbf{K}$ to be the adjacency matrix incorporating the neighborhood structure of the regions under consideration which results in partially improper Markov random field priors for the regional effects. Rue and Held (2005) give a detailed overview on construction and properties of $\mathbf{K}$.

- *Random effects:* The idea that observations within a group may behave different than across groups is represented by setting $\mathbf{K} = \mathbf{I}_L$, where L is the number of groups (i.i.d. random effects).

A detailed summary of available basis functions and smoothing matrices is given in (Fahrmeir et al., 2013, Chapter 9).

Bivariate Semiparametric SEM Formulation

A natural extension of the basis function approach in a bivariate setting is to define the predictor structure of each equation in the SEM individually. Based on the structure of a simple SEM in (1') and (2'), we get

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{B}\boldsymbol{\alpha}_1 + \boldsymbol{\varepsilon}_1 \\ \mathbf{y}_2 &= \mathbf{B}\boldsymbol{\alpha}_2 + \beta_{21}\mathbf{y}_1 + \boldsymbol{\varepsilon}_2, \end{aligned} \tag{6}$$

where the individual contributions of the observed responses and error terms are stored in the vectors $\mathbf{y}_j = (y_{j1}, \ldots, y_{jn})'$ and $\boldsymbol{\varepsilon}_j = (\varepsilon_{j1}, \ldots, \varepsilon_{jn})'$, $j = 1, 2$, respectively. As explained above, $\mathbf{B}$ captures the evaluations of the effect specific basis functions at the observations of the exogenous variables. The SEM in (6)

can be rewritten as a large multivariate model via

$$\begin{pmatrix} \mathbf{y}_1' \\ \mathbf{y}_2' \end{pmatrix} = \mathbf{M}^{-1} \begin{pmatrix} \boldsymbol{\alpha}_1' \\ \boldsymbol{\alpha}_2' \end{pmatrix} \mathbf{B}' + \mathbf{M}^{-1} \begin{pmatrix} \boldsymbol{\varepsilon}_1' \\ \boldsymbol{\varepsilon}_2' \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} 1 & 0 \\ -\beta_{21} & 1. \end{pmatrix} \tag{7}$$

For the rest of this article, we will assume that the error terms $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ are normally distributed and independent across equations, i.e.

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \overset{iid}{\sim} \mathcal{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \underbrace{\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}}_{=:\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}} \right), \quad i = 1, \ldots, n. \tag{8}$$

This assumption is based on the following reasoning:

- At least linear recursive SEM as investigated in this article (i.e. the same set of exogenous variables builds up the predictor of both endogenous variables) are technically identified, if and only if the error terms are independent (see Bollen, 1989). Although the regularization priors as introduced above in general reduce the effective degrees of freedom compared to unpenalized estimation, it is not within the scope of this article to thoroughly investigate the identifiability of regularized SEM with correlated error terms. Consequently, we stick to this - potentially too conservative - assumption.

- Due to the recursive structure of the model, the independence between ε_1 and ε_2 does not imply independence between y_1 and y_2. Instead, the relationship between the response variables is captured (a) by the direct effect β_{21}, (b) by the simultaneous influence of x on both responses and (c) by allowing the latter to be correlated itself. The underlying assumption thus corresponds to the idea that the correlation between y_1 and y_2 is completely captured by (a)-(c).

2.3 A Priori Correlated Semiparametric Predictors

Instead of using independent priors for $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ as in (5), i.e.

$$p(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \mid \tau_1^2, \tau_2^2) \propto \exp\left(-\frac{1}{2\tau_1^2}\boldsymbol{\alpha}_1'\mathbf{K}\boldsymbol{\alpha}_1\right) \cdot \exp\left(-\frac{1}{2\tau_2^2}\boldsymbol{\alpha}_2'\mathbf{K}\boldsymbol{\alpha}_2\right), \qquad (9)$$

we aim at allowing these coefficients to be correlated. As a typical example, in which the implicit assumption of uncorrelated effects might be too restrictive, consider the analysis of ecological data. Often, regional effects (modeled with Markov random field prior, see above) can be included to represent environmental factors. On the other hand, different species might react similarly or in a conflictive manner to their environment. This feature is potentially not accounted for, if the effects are forced to be independent by the choice of according priors. An example is given in Thaden et al. (2017) who simultaneously study the environmental sensitivity of young and adult mussels at the Galician coast.

Hence, inspired by the idea of multivariate conditionally autoregressive (MCAR) regional effects (e.g. Gelfand and Vounatsou, 2003) which are used for multivariate spatial data, we combine the individual priors in Equation (9) to a joint prior for $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$, namely

$$p\left(\underbrace{(\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2')'}_{=\boldsymbol{\alpha}} \mid \mathbf{A}\right) \propto \exp\left(-\frac{1}{2}\boldsymbol{\alpha}'(\mathbf{A}^{-1} \otimes \mathbf{K})\boldsymbol{\alpha}\right), \quad \mathbf{A} = \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix}. \qquad (10)$$

Again, τ_1^2 and τ_2^2 are the smoothing variances of the effects of x on y_1 and y_2, respectively while with the structure of $\mathbf{A}$ the correlation between the individual effects can be captured by ρ. Note that (9) is a special case of (10) via setting $\rho = 0$.

2.4 Interpretation

Joining SEM techniques with correlated semiparametric predictors yields some noteworthy features concerning the interpretation of the effects. First, as is typical for SEM, the overall influence of the exogenous variable x can be decomposed into a direct and an indirect effect on y_2. Following the arrows in Figure 1, the direct

effect corresponds to γ_2. The indirect effect results from multiplying along the dashed arrows, such that it is given by $\gamma_1 \cdot \beta_{21}$. Consequently, the total effect of x on y can be written as

$$\text{effect}_{x,\text{total}} = x \cdot (\gamma_2 + \gamma_1 \cdot \beta_{21}).$$

The same decomposition is obtained when semiparametric effects are involved. Replacing $x\gamma_1$ and $x\gamma_2$ by $f^{(1)}(x)$ and $f^{(2)}(x)$, respectively, yields

$$\begin{aligned} \text{effect}_{x,\text{total}} &= f^{(2)}(x) + \beta_{21} f^{(1)}(x) \\ &= \mathbf{B}\left(\boldsymbol{\alpha}_2 + \boldsymbol{\alpha}_1 \beta_{21}\right). \end{aligned} \tag{11}$$

Figure 3 illustrates the extension of Figure 1 and shows the influence of the model parameters within the recursive bivariate SEM. The overall variability of the basis functions (e.g. variability across space or between groups) is captured by τ_1^2 and τ_2^2 (dotted arrows). The correlation between the corresponding effects is captured by ρ (dashed arrow). The interpretation of this parameter is straightforward: it captures whether the influences of the exogenous variable on the endogenous variables are similar ($\rho > 0$), conflictive ($\rho < 0$) or not related at all ($\rho = 0$). The interpretation of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ depends on the type of effect (nonlinear, spatial, random, ...), whereas β_{21} is the classical linear effect.

2.5 Mixed Model Representation of the Predictors

Unfortunately, implementing the basis function approach along with its correlated effects as explained in Section 2.2, yields computational problems. We explain these for the special case of nonlinear effects via Bayesian P-splines with coefficient vectors $\boldsymbol{\alpha}_j = (\alpha_{j1}, \ldots, \alpha_{jL})', j = 1, 2$. As stated in Fahrmeir et al. (2013, Chapter 8.1), the first order difference penalty of an individual P-spline corresponds to a stochastic formulation, a first order random walk defined by

$$\alpha_{jl} = \alpha_{j,l-1} + u_{jl}, \quad u_{jl} \sim \mathcal{N}(0, \tau_j^2), j = 1, 2. \tag{12}$$

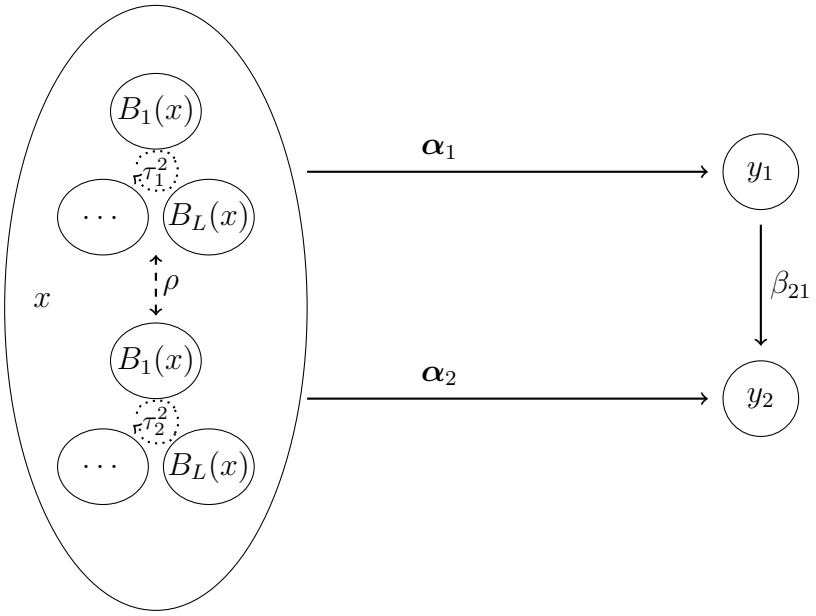

Figure 3: Decomposition of the overall interrelationships within the bivariate recursive model. Basis functions $B_1(x), \ldots, B_L(x)$ are evaluated at the exogenous variable x (left large ellipse). The variability across basis functions is controlled for via τ_1^2 and τ_2^2 (variances of the semiparametric effect, dotted arrows) and the penalty matrix $\mathbf{K}$ (see Section 2.2 for details and examples). Additionally, the correlation between the effects on the endogenous variables y_1 and y_2 is captured by ρ (dashed arrow). The effects of the basis functions on y_1 and y_2 are collected in $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$, respectively. Finally, y_1 has a linear effect β_{21} on y_2.

On the one hand, this representation is fundamental for deriving the prior distribution for $\boldsymbol{\alpha}_j$ in (5). On the other hand, two random walks as in (12) are usually highly correlated even if the stochastic parts $\mathbf{u}_1$ and $\mathbf{u}_2$ of the two random walks are independent - a phenomenon often called spurious correlation by econometricians (Simon, 1954). Consequently, the two vectors of coefficients $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ of the splines in our bivariate recursive SEM will be correlated as well, no matter what the true value of ρ is. As a result, the Markov chains (see Section 3) for ρ generally converge to ± 1 during the estimation via sampling. In particular, this parameter is not or only weakly identified in our model.

To overcome this problem, we use the mixed model representation of the basis function approach as outlined byFahrmeir et al. (2004):

$$\mathbf{B}\boldsymbol{\alpha}_j = \mathbf{X}\boldsymbol{\beta}_j + \mathbf{Z}\boldsymbol{\gamma}_j,$$

where $\mathbf{X}$ and $\mathbf{Z}$ result from the eigendecomposition of the precision matrix $\mathbf{K}$. Depending on the type of effect, $\mathbf{K}$ is rank deficient (e.g. has rank $L-1$ for MRF and rank $L-k$ for P-splines based on k-th order differences). The dimension of $\boldsymbol{\beta}_j$ corresponds to the rank deficit k and $\mathbf{X}\boldsymbol{\beta}_j$ can be interpreted as the unregularized baseline effect. The component $\mathbf{Z}\boldsymbol{\gamma}_j$ then captures the deviations from this baseline effect and is the regularized (or smoothed) part. For example,

- $\mathbf{X}\boldsymbol{\beta}_j$ is a horizontal line for P-splines based on first order differences and a linear trend for second order differences.
- $\mathbf{X}\boldsymbol{\beta}_j$ is the average regional effect in a MRF. The differences between regions are included in $\mathbf{Z}\boldsymbol{\gamma}_j$.
- $\mathbf{X} = \mathbf{0}$ for i.i.d. random effect, since in this case $\mathbf{K} = \mathbf{I}_L$ has full rank.

It can be shown that, with $\mathbf{X}$ and $\mathbf{Z}$ chosen as above, $p(\boldsymbol{\alpha}_j) \propto \exp\left(-\frac{1}{2\tau_j^2}\boldsymbol{\alpha}_j'\mathbf{K}\boldsymbol{\alpha}_j\right)$ implies

$$\boldsymbol{\gamma}_j \sim \mathcal{N}\left(\mathbf{0}, \tau_j^2\mathbf{I}_{L-k}\right).$$

Consequently, we convert the joint prior structure in Equation (10) to

$$p\left(\underbrace{(\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2')'}_{=\boldsymbol{\gamma}}\right) \propto \exp\left(-\frac{1}{2}\boldsymbol{\gamma}'(\mathbf{A}^{-1} \otimes \mathbf{I}_{L-k})\boldsymbol{\gamma}\right), \quad \mathbf{A} = \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix}.$$

3 Bayesian Inference and Estimation

The full conditional distributions based on the joint posterior $p(\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, \tau_1^2, \tau_2^2, \rho|\mathbf{y})$ are estimated via the Gibbs sampler described in this section.

3.1 Likelihood

The likelihood of $\mathbf{y}$ is based on the normality assumption in (8) and the model formulation in Equation (7). Consequently, we find

$$p\left(\begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} \mid \boldsymbol{\theta}\right) \propto \exp\left(-\frac{1}{2}\left(\begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} - \boldsymbol{\mu}_{y_i}\right)' \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \left(\begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} - \boldsymbol{\mu}_{y_i}\right)\right)$$

with

$$\boldsymbol{\mu}_{\mathbf{y}_i} = \mathbf{M}^{-1} \begin{pmatrix} \boldsymbol{\beta}_1' & \boldsymbol{\gamma}_1' \\ \boldsymbol{\beta}_2' & \boldsymbol{\gamma}_2' \end{pmatrix} \begin{pmatrix} \mathbf{X}[i,\ ]' \\ \mathbf{Z}[i,\ ]' \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{M}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} \left(\mathbf{M}^{-1}\right)'$$

and $\boldsymbol{\theta}$ collects all unknown model parameters.

3.2 Prior Choices

Coefficient for the Direct Covariate Effect

The coefficient β_{21} is assigned a flat prior, i.e. $p(\beta_{21}) \propto const.$

Unpenalized Part of the Coefficient Vector

We employ weakly informative conjugate priors for $(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$

$$(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')' \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \nu^2 & 0 \\ 0 & \nu^2 \end{pmatrix} \otimes \mathbf{I}_k\right), \quad \text{with large } \nu.$$

Correlated Part of the Coefficient Vector

For the regularized part of the effect specific coefficients, we use the prior distribution introduced in Section 2.2 and 2.5, namely

$$\boldsymbol{\gamma} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{A} \otimes \mathbf{I}_{L-k}\right)$$

Error Variances

Similar to $(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$, we use the conjugate inverse gamma distribution for the error variances σ_1^2 and σ_2^2 with small scale and shape parameters:

$$\sigma_j^2 \sim \text{IG}(0.001, 0.001), \quad j = 1, 2.$$

Prior Covariance Structure of $\boldsymbol{\gamma}$

As hyperprior for $\mathbf{A}$ we chose an inverse Wishart distribution with $\kappa = 4$ degrees of freedom and scale matrix $\boldsymbol{\Psi} = \mathbf{I}_2$. This choice corresponds to the least informative

hyperprior for which the mean still exists.
The resulting full conditional distributions for the model parameters then establish a Gibbs sampler for the estimation of all model parameters (see Appendix A).

3.3 Model Selection

The MCMC draws resulting from the Gibbs sampler allow for the application of well known model selection criteria in order to compare the overall predictive performance of the model. We focus on the Watanabe-Akaike information criterion (WAIC, see Watanabe, 2010). Similar to the deviance information criterion (DIC, see Spiegelhalter et al., 2002) it is easily computable from the MCMC samples, but has the advantage that it does not rely on posterior point estimates. Instead, it incorporates the complete simulated posterior distribution of the parameters and can hence be seen as fully Bayesian. As DIC, it is usually interpreted as a compromise between predictive capability of the model and its complexity in terms of the effective number of parameters. It is defined as

$$\text{WAIC} = -2 \cdot \sum_{i=1}^{n} \log \left(\frac{1}{S} \sum_{s=1}^{S} p(\mathbf{y}_i \mid \boldsymbol{\theta}^{(s)}) \right) + 2 \cdot \hat{p}_{\text{WAIC}},$$

where S is the number of MCMC samples. For details on the implementation and the effective number of parameters $\hat{p}_{\text{WAIC}}$, see Gelman et al. (2014).

4 Simulation study

4.1 Setup

We investigate the capability of our approach to identify the model parameters in a broad set of simulated scenarios (36 in total). The performance of the model is examined for different sample sizes and various correlation structures within the data. Also, the influence of presence or absence of a recursive structure, incorporated via $\beta_{21} \neq 0$ or $\beta_{21} = 0$, respectively, is analyzed. Data are generated according to Equation (7). We refer to Table 1 for details on the simulation process.

Effect type	β_{21}	ρ	n
P-Splines	$\{0, 1\}$	$\{0, 0.5, -0.8\}$	$\{150, 500\}$
Discrete spatial	$\{0, 1\}$	$\{0, 0.5, -0.8\}$	$\{150, 500\}$
Random intercept	$\{0, 1\}$	$\{0, 0.5, -0.8\}$	$\{150, 500\}$

Table 1: Simulated model parameters for the different effect types. The remaining parameters where fixed to $\sigma_1^2 = \sigma_2^2 = \tau_1^2 = \tau_2^2 = 1$. The total number of scenarios is 36, each scenario has been replicated 100 times.

4.2 Results

In the following, our results are evaluated via estimation bias and mean squared error (MSE) for the 100 repetitions of each scenario. Bayesian inference is based on 1000 post burnin and post thinning MCMC samples (we discard 500 burnin samples and use a thinning of 5).

P-Splines

In this section, we illustrate the estimation results for the scenarios mentioned in Table 1 applied to P-splines with 10 inner knots and second order differences. The exogenous variable x ranges from 0 to 5. Figure 4 exemplarily shows the simulated nonlinear functions $f^{(1)}(x)$ on y_1 (left) and $f^{(2)}(x)$ on y_2 (right). The functions look similar across repetitions within a simulation scenario. The solid, dashed and dotted lines correspond to true correlations between the spline coefficients of $\rho = 0, \rho = 0.5$ and $\rho = -0.8$, respectively. Furthermore, Figure 4 nicely illustrates how a nontrivial correlation $\rho \neq 0$ induces additional smoothing of the nonlinear functions.

Estimation results for the direct effect $\hat{\beta}_{21}$ of y_1 on y_2 as well as the correlation $\hat{\rho}$ between the splines coefficients are summarized in Figure 5. The structure of the four individual plots is as follows:

- The solid lines together with '$\circ$' correspond to the case $\beta_{21} = 1, n = 150$.
- The dashed lines together with '$\triangle$' correspond to the case $\beta_{21} = 0, n = 150$.
- The dotted lines together with '$+$' correspond to the case $\beta_{21} = 1, n = 500$.
- The dashed and dotted lines together with '$\times$' correspond to the case $\beta_{21} = 0, n = 500$.

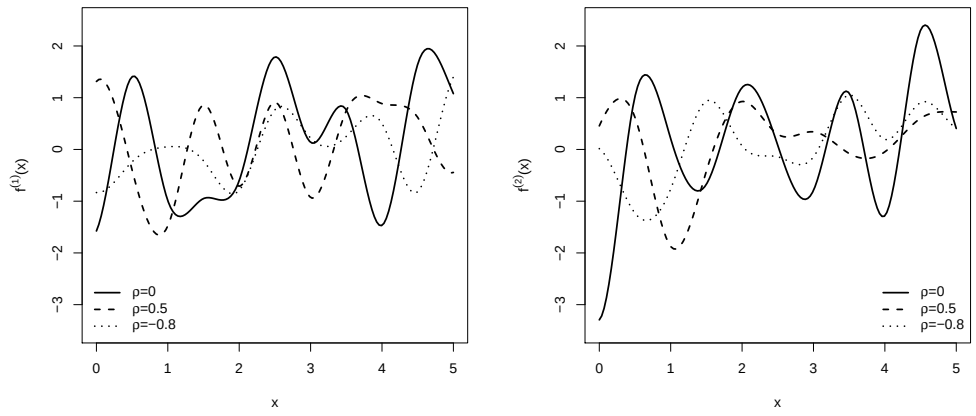

Figure 4: Simulated nonlinear functions of x on y_1 (left) and y_2 (right). Displayed are examples for the cases $\rho = 0$ (solid lines), $\rho = 0.5$ (dashed lines) and $\rho = -0.8$ (dotted lines).

In all four individual plots, MSE and bias are plotted against the absolute value of the correlation $|\rho|$. The upper left plot shows the MSE of $\hat{\beta}_{21}$ compared across simulation scenarios. Estimation of this coefficient seems to be independent of the effect size (i.e. the cases $\beta_{21} = 0$ and $\beta_{21} \neq 0$ are estimated similarly well). On the other hand, an increasing sample size heavily reduces both, the bias and MSE. For a given sample size, the bias is the same for $\beta_{21} = 0$ and $\beta_{21} = 1$, whereas it can be reduced with larger sample sizes.

The estimates for ρ shows a similar behavior though not as pronounced as for the linear effect. Especially for highly correlated splines, the MSE is substantially smaller in larger samples (upper right plot). The bias is reduced in all scenarios when n is large (lower right plot). Again, the size of β_{21} has no effect on the performance of $\hat{\rho}$ with respect to MSE or bias. It should be noted that MSE and bias are in general larger for $\hat{\rho}$ than for $\hat{\beta}_{21}$, i.e. it seems (as expected) to be more difficult to identify the correlation structure of the semiparametric effects than the recursive linear effect.

Comparison to Models with Uncorrelated Effects

When comparing the performance of our model which incorporates effects with a general correlation structure to the alternative approach using independent priors as in Equation (9), we find that – though on a small scale – the estimation of the

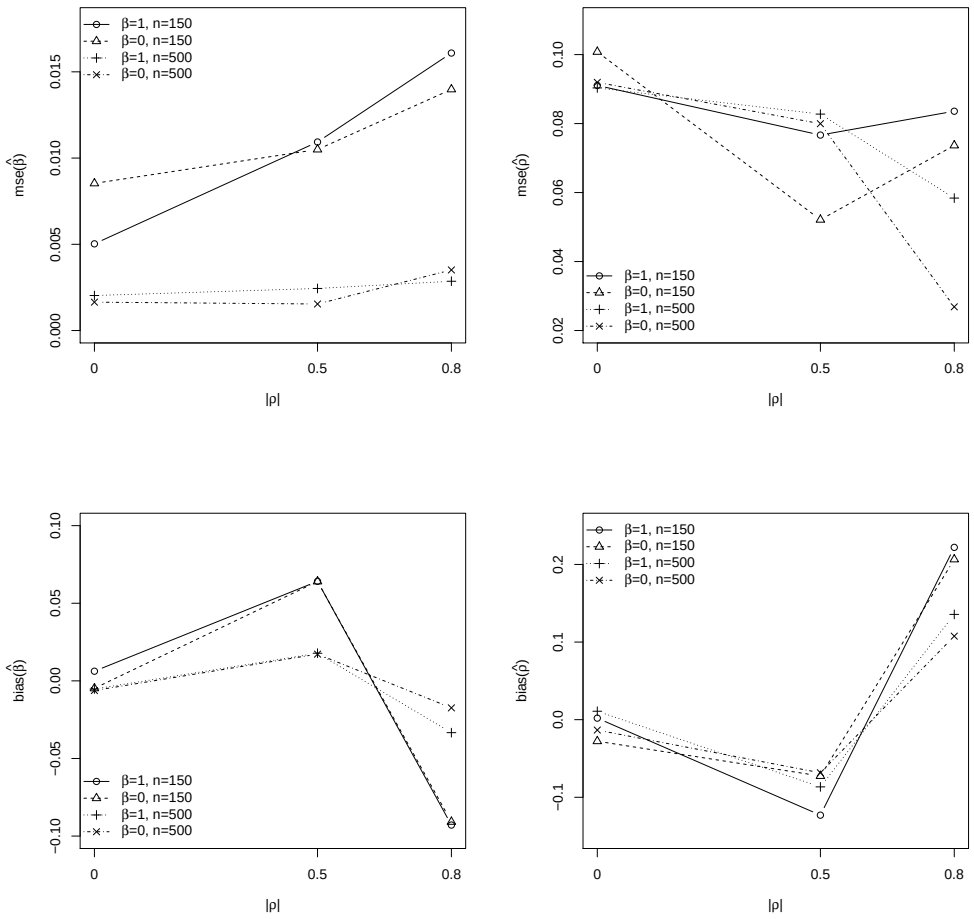

Figure 5: Top row: Mean squared error of $\hat{\beta}$ (left) and $\hat{\rho}$ (right) across simulation scenarios. Bottom row: Estimation bias of $\hat{\beta}$ (left) and $\hat{\rho}$ (right) across simulation scenarios.

linear covariate effect β_{21} tends to be more stable in the general case. Exemplarily, we show the values of MSE and bias of the corresponding estimator for three simulation scenarios in Table 2. The results for the other scenarios are comparable. In addition, we found that the additional smoothing which results from allowing the effects to be correlated results in a generally smaller WAIC compared to a model with uncorrelated effects (not shown for the simulations, see Section 5.3 for a comparison of WAIC in the applications).

Discrete Spatial Effects and i.i.d. Random Effects

Additionally, we simulated spatially structured data on an artificial map with 49 regions (as in Thaden and Kneib, 2017) as well as clustered data (10 clusters)

Scenario / Model	$\text{mse}(\hat{\beta}_{21})$	$\text{bias}(\hat{\beta}_{21})$
$\beta_{21} = 1, \rho = 0$ / $\rho = 0$	0.007	0.011
$\beta_{21} = 1, \rho = 0$ / $\rho \neq 0$	0.005	0.006
$\beta_{21} = 1, \rho = 0.5$ / $\rho = 0$	0.023	0.105
$\beta_{21} = 1, \rho = 0.5$ / $\rho \neq 0$	0.011	0.064
$\beta_{21} = 1, \rho = -0.8$ / $\rho = 0$	0.030	-0.140
$\beta_{21} = 1, \rho = -0.8$ / $\rho \neq 0$	0.016	-0.093

Table 2: Comparison of mean squared error and bias of $\hat{\beta}_{21}$ between the general approach and a model with uncorrelated semiparametric effects.

for the evaluation of our model with the effect types explained in Section 2.2. In summary and as expected (due to the similarities in the basis function approach), the results for simulations with discrete spatial data and i.i.d. random effects correspond to those obtained for P-splines. Consequently, we abstain from showing these results in this section.

5 Applications

5.1 Correlated Spatial Effects: Malnutrition in Africa and Asia

Childhood undernutrition is one of the major health problems in developing countries. Specific forms of malnutrition have been linked to individual characteristics of children and their parents in various studies (e.g. Klein and Kneib, 2016). The term *wasting* characterizes low weight for height and is generally associated with acute starvation. On the other hand, *stunting* is a sign for long-term suboptimal nutritional conditions and is defined as low height for age. Both measures are usually reported as Z-scores that compare the individual nutrition status with a predefined reference population:

$$z = \frac{\text{observed value} - \text{median value in reference population}}{\text{standard deviation in reference population}}. \tag{13}$$

Based on the WHO definition, an individual is considered to suffer from wasting or stunting, if its weight or height is two standard deviations below the median weight or height of the reference population, respectively.

Our analysis is based on country specific proportions of children under five years of age who are affected by wasting and stunting while controlling for environmental and political circumstances. The latter factors are included as correlated discrete spatial effects, see Figure 6 for an illustration.

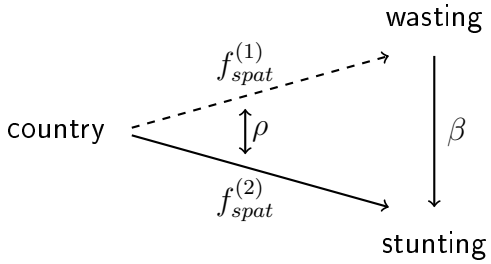

Figure 6: Flexible SEM approach with correlated regional effects to explain the relationships between the location and wasting and stunting proportions. The endogenous variables are on the log scale.

Concretely, we estimate the interrelations using the model formulation

$$\log(wasting_i) = f^{(1)}_{spat}(country_i) + \varepsilon_{1i}$$
$$\log(stunting_i) = f^{(2)}_{spat}(country_i) + \beta \log(wasting_i) + \varepsilon_{2i},$$

where the spatial functions $f^{(1)}_{spat}$ and $f^{(2)}_{spat}$ are included as correlated Markov random fields. Our findings are based on WHO data for African and western Asian developing countries (WHO, 2016). For each country, multiple observations (between 1990 and 2014) are available. The number of data points differs from country to country and ranges between 1 (e.g. in Turkmenistan) and 21 (in Bangladesh). We estimate two separate models for 47 African ($n = 237$ observations) and 23 Asian ($n = 127$) countries.

We find a significant effect of the log proportion of wasted children on that of stunted children (and hence of acute on chronic undernutrition) of $\hat{\beta}_{Asia} = 0.57$ (in a 90% credibility interval $[0.28, 0.83]$) along with a correlation of $\hat{\rho}_{Asia} = 0.53$ ($[0.02, 0.85]$) in Asia. At the same time, the estimates for African developing countries point in the same direction ($\hat{\beta}_{Africa} = 0.10$ and $\hat{\rho}_{Africa} = 0.11$). However, they are not significant as the 90% credibility intervals are $[-0.05, 0.23]$ and $[-0.30, 0.50]$, respectively.

The country specific spatial effects are shown in Figure 7 for Africa and in Figure 8

for Asia. Both, wasting (left plot in Figure 7) and stunting (central plot) proportions are more strongly affected in countries in or close to the Sahel, which can at least partly be explained by the challenging climatic conditions in that area. In Asia, more children suffer from wasting (left plot in Figure 8) and stunting (central plot) in the south east, whereas northern countries are not that strongly affected. In this case, the positively estimated spatial correlation ($\hat{\rho}_{Asia} = 0.53$) is visible as similar spatial patterns in wasting and stunting. The total country-specific effects based on Equation 11 are shown in the right plots of Figures 7 for Africa and 8 for Asia, respectively. Note that the spatial effects are significant in general. However, for both continents a large proportion of the marginal credibility intervals of individual country-specific effects on stunting overlaps 0 (for details, see Figure 11 and 12 in Appendix B).

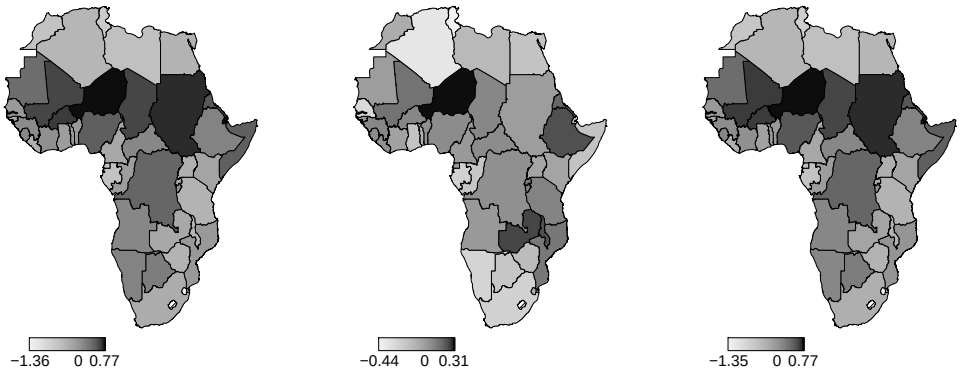

Figure 7: Estimated country specific effect for acute (wasting, left) and chronic (stunting, center) in Africa. The right plot shows the total regional effect for stunting in Africa based on Equation (11).

5.2 Correlated P-Splines: Species Richness of Plants and Animals

Understanding the environmental drivers of species richness is of crucial ecological interest. Illustrating the complex interactions between different species and their environments often is an incentive for extending standard statistical models in order to capture these interrelations. Jetz et al. (2009), for example, relate the relationship of species richness of plants and animals to environmental factors

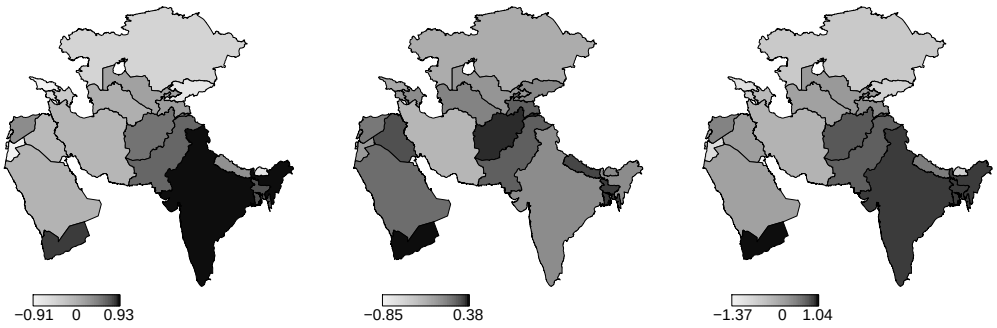

Figure 8: Estimated country specific effect for acute (wasting, left) and chronic (stunting, center) in Asia. The right plot shows the total regional effect for stunting in Asia based on Equation (11).

(e.g. temperature and number of different ecosystems) at 639 sampling locations worldwide using a linear structural equation model approach (see Figure 9 (a)).

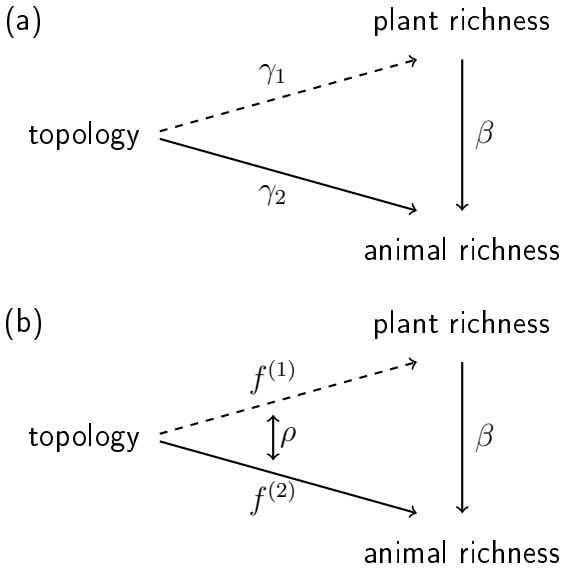

Figure 9: (a) SEM approach similar to Jetz et al. (2009): The topological diversity simultaneously affects plant and animal richness linearly. (b) Proposed approach using correlated splines to explain the relationships between topology and plant and animal species richness. All variables are on the log scale.

They analyze if the resulting correlation is due to a direct effect of plant richness on animal richness (i.e. from producer to consumer) or if it "emerges more strongly from similar responses to environmental gradients". Klein and Kneib (2016) apply a structured additive copula model to a sub-dataset from Jetz et al. (2009) in order to explain the aforementioned dependency structure by environmental covariates

but do not allow for a direct association between plant and animal species richness. We use the dataset (consisting of $n = 480$ observations) from Klein and Kneib (2016) and extend the approach of Jetz et al. (2009) as explained in Section 2. As illustrated in Figure 9 (b), we allow the effects of the topological diversity ($topo$, measured as maximal range of elevation within the sampling region) on plant and animal species richness to be nonlinear and correlated. We estimate the effects based on the model

$$\begin{aligned}\log(plants_i) &= f^{(1)}(\log(topo_i)) + \varepsilon_{1i} \\ \log(animals_i) &= f^{(2)}(\log(topo_i)) + \beta \log(plants_i) + \varepsilon_{2i}.\end{aligned}$$

From the results we find a significant direct effect $\hat{\beta} = 0.45$ (in a 90% credibility interval $[0.37, 0.54]$) of plant species richness on animal species richness. The estimated nonlinear effects on plant and animal species richness along with 80% and 90% pointwise credibility intervals are illustrated in the left and central panel of Figure 10, respectively. The right plot shows the total effect of topological diversity on animal species richness as a linear combination of the individual nonlinear effects based on Equation (11).

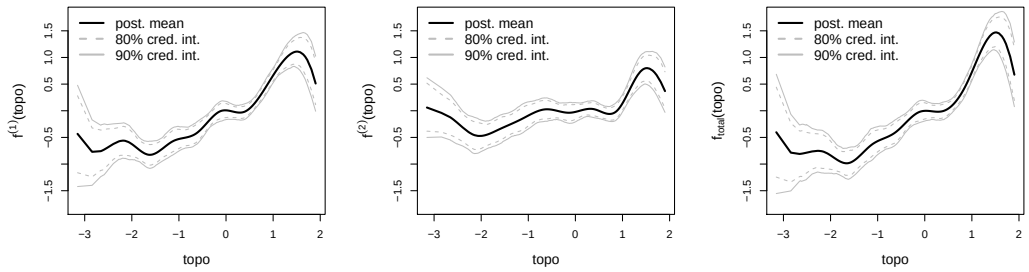

Figure 10: Nonlinear effects of topological diversity on species richness of plants (left) and animals (center). The total topological effect on animal species richness based on Equation (11) is shown in the right plot. The 80% and 90% pointwise credibility intervals are shown as gray dashed and solid lines, respectively.

Overall, we find a positive influence of topological diversity on both species richness in plants and animals. The effect is more pronounced for plants but still significant based on the 80% and 90% credibility intervals. The slope appears to be larger for higher topological diversity and is negative for extremely high elevation ranges. The negative slopes for highly heterogeneous regions could possibly be explained

by the area-heterogeneity trade off described by Allouche et al. (2012).
The estimated splines show a similar pattern which is also reflected by the estimated correlation $\hat{\rho} = 0.72$ ($[0.39, 0.91]$). This estimate thus quantifies the "similar responses to environmental gradients" as mentioned by Jetz et al. (2009) and hence supports the interpretation of the authors.

5.3 Model Selection for the Applications

Evaluating WAIC for the examples explained above illustrates another feature of our approach: besides its ability to capture potential correlations between the included effects, this additional smoothing reduces the overall model complexity. As a result, WAIC is lower (compared to a model with uncorrelated effects) in all cases as shown in Table 3.

	Malnutrition in Africa	Malnutrition in Asia	Species richness
$\rho = 0$	530.41	424.57	2385.43
$\rho \neq 0$	465.49	404.61	2383.40

Table 3: WAIC values for the models applied in Sections 5.1 and 5.2. Model performance is evaluated for uncorrelated (first row) and correlated (second row) semiparametric effects.

6 Discussion

In this contribution, we extend the framework of simultaneous equation models with the flexibility of semiparametric effects. In order to capture potentially complex correlation structures within the data, we illustrate how different types of effects (e.g. nonlinear or spatial) can be incorporated into SEM via a basis function approach using correlated prior structures for the corresponding effects. We show how the resulting effects can be interpreted and how the overall variability and correlation in the model can be decomposed using path diagrams. For the estimation, we implement a Gibbs sampler based on conjugate priors for all parameters.
In an extensive simulation study, we evaluate the model's performance in a large variety of scenarios. Independent from the effect type, our approach is capable of

identifying the occurring effects and of recovering the correlation of these effects. Compared to a model with independent effects across equations, introducing a general covariance structure yields additional smoothing of the effects and hence a reduction of the WAIC (which we also find in all applications).

We illustrate the interdisciplinary applicability of our approach in two examples. Firstly, we analyze the effect of the proportion of acutely malnourished children on that of chronically undernourished children in African and Asian developing countries. Correlated country-specific effects are included in order to capture environmental (e.g. climatic) and political factors within these countries. Based on WHO data between 1990 and 2014, we find a positive effect of acute on chronic undernutrition and positively correlated spatial effects.

In a second example, we use correlated Bayesian P-splines in order to simultaneously quantify the effect of topological diversity on plant and animal species richness. Along with a positive direct effect of plant on animal species richness (representing the idea that producers attract consumers), the estimated nonlinear effects of topological diversity are highly correlated, indicating that plants and animals react in a similar way to their environment in the sampling regions.

Identification is a crucial aspect in SEM. For that reason, we impose the conservative (i.e. emerging from linear SEM) assumption of uncorrelated error terms within the two recursive model equations. As mentioned above, smoothing in general and in particular smoothing across equations will lead to a reduction of the overall effective number of parameters. It appears worthwhile to investigate under which circumstances (i.e. the degree of smoothing) the assumption of independence can be relaxed. Additionally, it is conceptually straightforward to further extend the predictor structure of the endogenous variables by using multiple semiparametric effects at once. With the appropriate prior choices, our Gibbs sampler only needs slight adjustments. Finally, the field of applications of our approach can be widened by allowing for distributional flexibility in the responses (i.e. relaxing the assumption of normally distributed error terms). Depending on the choices of other parametric distributions, full conditionals of the parameters will no longer be obtained from simple Gibbs steps, but from a Metropolis-Hastings type step with appropriate proposal distributions.

Acknowledgments

The work of Hauke Thaden was partly supported by the German Research Foundation (DFG) via the research training group 1644 on *Scaling problems in statistics*. The author would like to thank Holger Kreft from the Department of Biodiversity, Macroecology & Biogeography at the University of Göttingen for providing the data for the application on species richness patterns and his helpful advice concerning the interpretation of the results. Additionally, the exchange of ideas and concepts with Thomas Kneib and Nadja Klein (both from the University of Goettingen) concerning this manuscript is highly appreciated.

References

Allouche, O., Kalyuzhny, M., Moreno-Rueda, G., Pizarro, M., and Kadmon, R. (2012). Area-heterogeneity tradeoff and the diversity of ecological communities. *Proceedings of the National Academy of Sciences*, 109(43):17495–17500.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Wiley series in probability and mathematical statistics. Wiley, New York.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, pages 89–102.

Fahrmeir, L., Kneib, T., and Lang, S. (2004). Penalized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica*, pages 731–761.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression – Models, Methods and Applications*. Springer, Berlin, Heidelberg.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4(1):11–15.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016.

Jetz, W., Kreft, H., Ceballos, G., and Mutke, J. (2009). Global associations between terrestrial producer and vertebrate consumer diversity. *Proceedings of the Royal Society B: Biological Sciences*, 276(1655):269–278.

Klein, N. and Kneib, T. (2016). Simultaneous inference in structured additive conditional copula regression models: a unifying Bayesian approach. *Statistics and Computing*, 26(4):841–860.

Klein, N., Kneib, T., Lang, S., and Sohn, A. (2015). Bayesian structured additive distributional regression with an application to regional income inequality in Germany. *The Annals of Applied Statistics*, 9(2):1024–1052.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212.

Pata, M. P., Kneib, T., Cadarso-Suarez, C., Lustres-Perez, V., and Fernandez-Pulpeiro, E. (2012). Categorical structured additive regression for assessing habitat suitability in the spatial distribution of mussel seed abundance. *Environmetrics*, 23(1):75–84.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Simon, H. A. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association*, 49(267):467–479.

Song, X.-Y., Lu, Z.-H., Cai, J.-H., and Ip, E. H.-S. (2013). A Bayesian Modeling Approach for Generalized Semiparametric Structural Equation Models. *Psychometrika*, 78(4):624–647.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Thaden, H. and Kneib, T. (2017). Structural equation models for dealing with spatial confounding. *The American Statistician, to appear*.

Thaden, H., Pata, M. P., Klein, N., Cadarso-Suarez, C., and Kneib, T. (2017). Integrating multivariate conditionally autoregressive spatial priors into recursive bivariate models for analyzing environmental sensitivity of mussels. *Spatial Statistics, submitted*.

Waldmann, E., Taylor-Robinson, D., Klein, N., Kneib, T., Pressler, T., Schmid, M., and Mayr, A. (2017). Boosting joint models for longitudinal and time-to-event data. *Biometrical Journal, accepted*.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.

WHO (2016). *Food Security Indicator Data Base*. World Health Organization.

Wood, S. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, New York.

A Full Conditionals and Gibbs Sampler

Based on the posterior distribution

$$p(\beta_{21}, \tilde{\boldsymbol{\delta}}, \sigma_1^2, \sigma_2^2, \mathbf{A} \mid \mathbf{y}) \propto \mathcal{L} \times p(\beta_{21}) \times p(\tilde{\boldsymbol{\delta}}) \times p(\sigma_1^2) \times p(\sigma_2^2) \times p(\mathbf{A}),$$

we derive the full conditionals for the model parameters as follows.

Direct Covariate Effect

The full conditional distribution of the direct covariate effect β_{21} given all other model parameters is

$$\beta_{21} | \cdots \sim \mathcal{N}\left(\frac{\mathbf{y}_1'\mathbf{y}_2 - \sum_{l=1}^{L} \left(\sum_{i=1}^{n} y_{1i} v_{il} \right) \delta_{2l}}{\mathbf{y}_1'\mathbf{y}_1}, \frac{\sigma_2^2}{\mathbf{y}_1'\mathbf{y}_1} \right),$$

where v_{il} is the entry at position $[i, l]$ of the matrix $\mathbf{V} = (\mathbf{X} \mid \mathbf{Z})$ and δ_{2l} captures the components of the coefficients vector $\boldsymbol{\delta}_2 = (\boldsymbol{\beta}_2', \boldsymbol{\gamma}_2')'$.

Unpenalized and Penalized Part of the Coefficient Vector

After some re-ordering of the coefficients, we derive a joint full conditional for $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ and $\boldsymbol{\gamma}$. More precisely, with

$$\begin{aligned}
\tilde{\boldsymbol{\delta}} &= (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2')', \\
\tilde{\mathbf{V}} &= \left(\tilde{\mathbf{X}} \mid \tilde{\mathbf{Z}}\right), \quad \text{where } \tilde{\mathbf{X}} = \mathbf{M}^{-1} \otimes \mathbf{X} \text{ and } \tilde{\mathbf{Z}} = \mathbf{M}^{-1} \otimes \mathbf{Z},
\end{aligned}$$

as well as

$$\begin{aligned}
\tilde{\boldsymbol{\Sigma}} &= (\mathbf{M}^{-1} \otimes \mathbf{I}_n)(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} \otimes \mathbf{I}_n)(\mathbf{M}^{-1} \otimes \mathbf{I}_n)' \text{ and} \\
\tilde{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{\delta}}} &= \begin{pmatrix} \begin{pmatrix} \nu^2 & 0 \\ 0 & \nu^2 \end{pmatrix} \otimes \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \otimes \mathbf{I}_{L-k} \end{pmatrix},
\end{aligned}$$

it can be shown that the full conditional of $\tilde{\boldsymbol{\delta}}$ given the remaining model components is again a Gaussian distribution, namely

$$\tilde{\boldsymbol{\delta}} | \cdots \sim \mathcal{N}\left(\left(\tilde{\mathbf{V}}'\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{V}} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\delta}}}^{-1}\right)^{-1} \tilde{\mathbf{V}}'\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{y}}, \left(\tilde{\mathbf{V}}'\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{V}} + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\delta}}}^{-1}\right)^{-1}\right),$$

where $\tilde{\mathbf{y}} = (\mathbf{y}_1', \mathbf{y}_2')'$.

Error Variances and Covariance Structure of $\boldsymbol{\gamma}$

The full conditional distributions of the error variances calculate as

$$\sigma_j | \cdots \sim \text{IG}\left(a_{\sigma_j^2}, b_{\sigma_j^2}\right), \; j = 1, 2,$$

with

$$
\begin{aligned}
a_{\sigma_1^2} &= a_{\sigma_2^2} = 0.001 + \frac{n}{2}, \\
b_{\sigma_1^2} &= 0.001 + 0.5 \sum_{i=1}^{n} y_{1i}^2 - \sum_{l=1}^{L} \left(\sum_{i=1}^{n} y_{1i} v_{il} \right) \delta_{1l} - 0.5 \sum_{i=1}^{n} \left(\sum_{l=1}^{L} v_{il} \delta_{1l} \right)^2 \text{ and} \\
b_{\sigma_2^2} &= 0.001 + \frac{\beta_{21}^2}{2} \sum_{i=1}^{n} y_{1i}^2 - \beta_{21} \sum_{i=1}^{n} y_{1i} y_{2i} + 0.5 \sum_{i=1}^{n} y_{2i}^2 \\
&\quad - \beta_{21} \sum_{l=1}^{L} \left(\sum_{i=1}^{n} y_{1i} v_{il} \right) \delta_{2l} + \sum_{l=1}^{L} \left(\sum_{i=1}^{n} y_{2i} v_{il} \right) \delta_{2l} + 0.5 \sum_{i=1}^{n} \left(\sum_{l=1}^{L} v_{il} \delta_{2l} \right)^2,
\end{aligned}
$$

with v_{il} and δ_{jl} as above.

Finally, the full conditional distribution of the covariance structure of the coefficient vector $\boldsymbol{\gamma}$ is given by

$$
\mathbf{A} | \cdots \sim \text{IW} \left(\tilde{\boldsymbol{\gamma}}' \tilde{\boldsymbol{\gamma}} + \boldsymbol{\Psi}, L - k + \kappa \right).
$$

Iteratively drawing from these full conditional distributions consequently constitutes the Gibbs sampler on which our inference is based.[1]

B Uncertainty of Country-specific Effects for the Malnutrition Data

The figures in this section of the appendix show the marginal credibility intervals of the country-specific effects from the application on malnutrition in Africa and Asia (see Section 5). The intervals are calculated based on the MCMC samples for the regression coefficients.

[1] When comparing our results to those from a model with uncorrelated semiparametric effects during the simulation study in Section 4, we employ inverse-gamma priors for the individual variance τ_1^2 and τ_2^2. The Gibbs sampler is accordingly adjusted in this case.

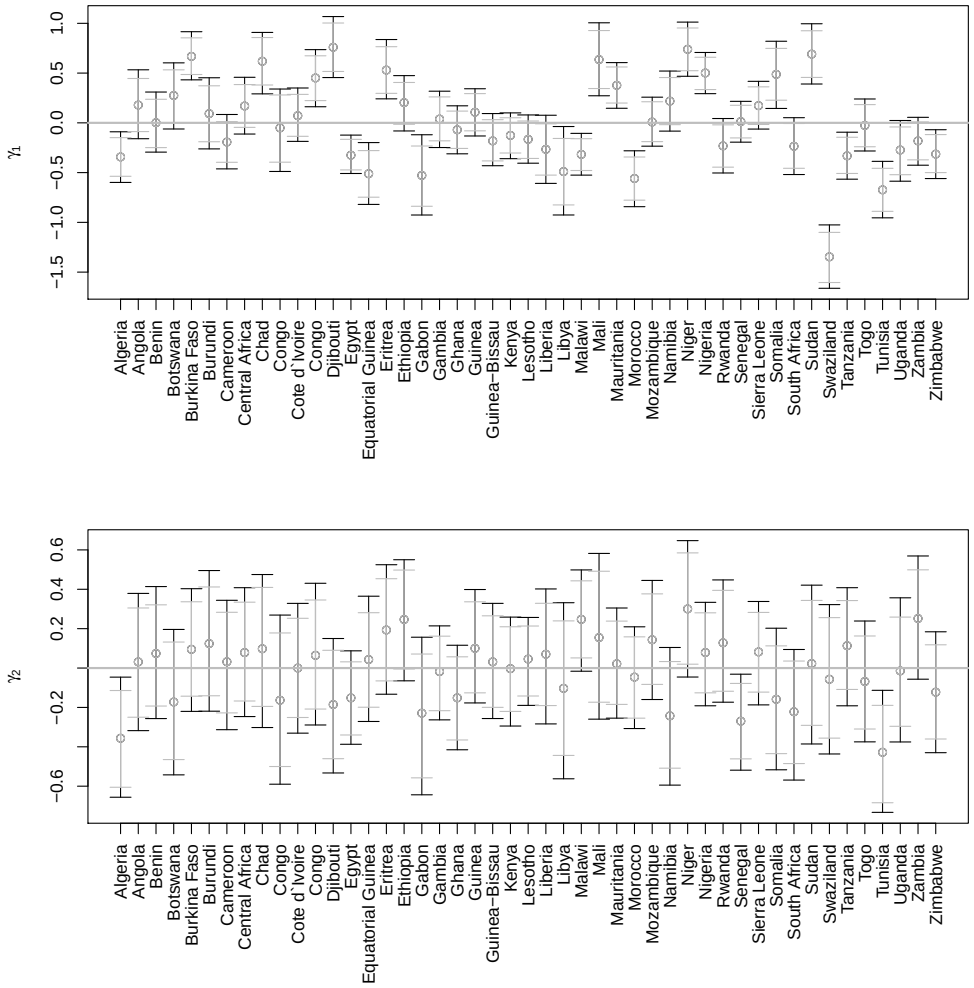

Figure 11: Marginal credibility intervals for the country-specific effects on wasting (upper panel) and stunting (lower panel) in Africa. The 80% and 90% credibility intervals are illustrated in gray and black, respectively.

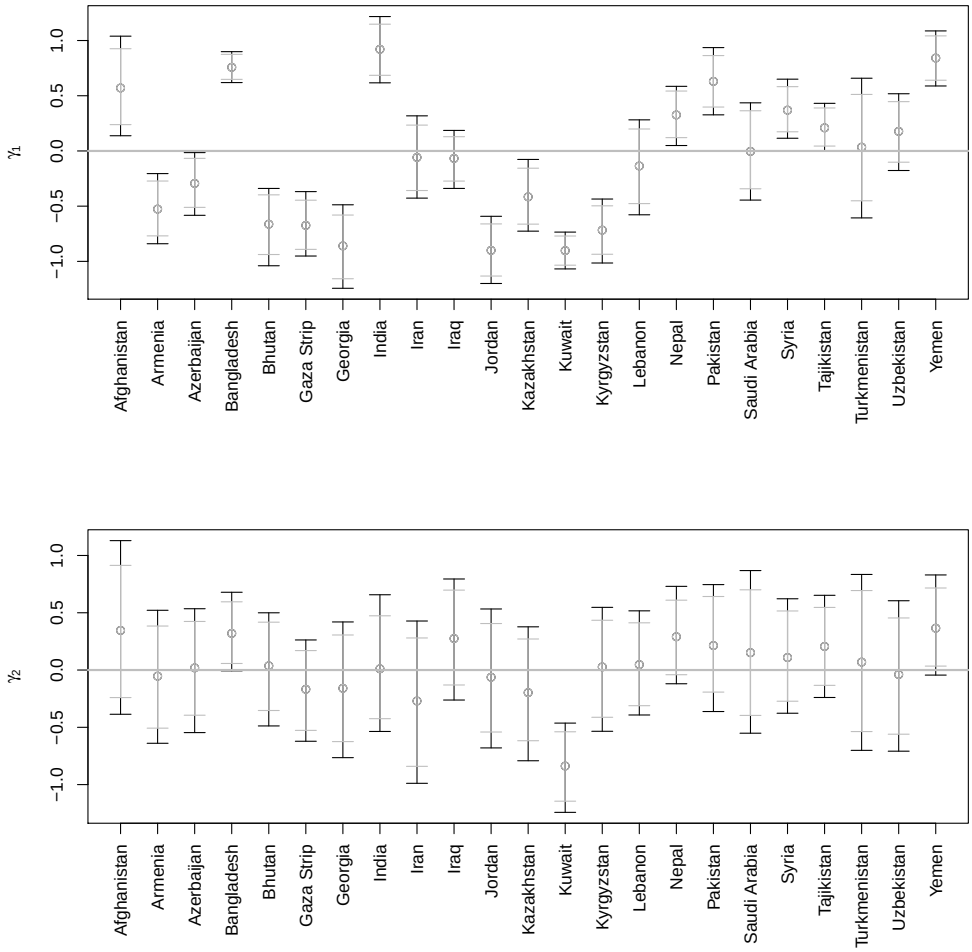

Figure 12: Marginal credibility intervals for the country-specific effects on wasting (upper panel) and stunting (lower panel) in Asia. The 80% and 90% credibility intervals are illustrated in gray and black, respectively.