Roth, Helene; Lang, Stefan; Wagner, Helga

Working Paper

Random intercept selection in structured additive regression models

Working Papers in Economics and Statistics, No. 2015-02

Provided in Cooperation with:
Institute of Public Finance, University of Innsbruck

Suggested Citation: Roth, Helene; Lang, Stefan; Wagner, Helga (2015) : Random intercept selection in structured additive regression models, Working Papers in Economics and Statistics, No. 2015-02, University of Innsbruck, Research Platform Empirical and Experimental Economics (eeecon), Innsbruck

This Version is available at:
https://hdl.handle.net/10419/122202

University of Innsbruck

Random intercept selection in structured additive regression models

Helene Roth, Stefan Lang, Helga Wagner

Working Papers in Economics and Statistics

2015-02

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 Banking and Finance
- Department of Economics
- Department of Public Finance
- Department of Statistics

Contact address of the editor:
Research platform “Empirical and Experimental Economics”
University of Innsbruck
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.

Random intercept selection in structured additive regression models

Helene Roth, Stefan Lang
Department of Statistics,
University of Innsbruck

Helga Wagner
Institute of Applied Statistics (IFAS)
Johannes Kepler University Linz

Abstract

This paper discusses random intercept selection within the context of semiparametric regression models with structured additive predictor (STAR). STAR models can deal simultaneously with nonlinear covariate effects and time trends, unit- or cluster-specific heterogeneity, spatial heterogeneity and complex interactions between covariates of different type. The random intercept selection is based on spike and slab priors for the variances of the random intercept coefficients. The aim is to achieve shrinkage of small random intercept coefficients to zero similar as for the LASSO in frequentist linear models. The mixture structure of the spike and slab prior allows for selective shrinkage, as coefficients are either heavily shrunk under the spike component or left almost unshrunk under the slab component. The hyperparameters of the spike and slab prior are chosen by theoretical considerations based on the prior inclusion probability of a particular random coefficient given the true effect size. Using extensive simulation experiments we compare random intercept models based on spike and slab priors for variances with the usual Inverse Gamma priors. A case study on malnutrition of children in Zambia illustrates the methodology in a real data example.

Keywords: Bayesian hierarchical models, Bayesian model choice, MCMC, P-splines, spike and slab priors

1 Introduction

A particularly broad and rich framework for semiparametric regression is provided by generalized structured additive regression (STAR) models proposed in Fahrmeir, Kneib, and Lang (2004) and Brezger and Lang (2006), see Fahrmeir, Kneib, Lang, and Marx (2013) for an introduction from first principles. Models of similar complexity have been developed in a mostly frequentist setting by Simon Wood (see e.g. Wood 2006) and in Ruppert, Wand, and Carroll (2003), Rigby and Stasinopoulos (2005) or Rue, Martino, and Nicolas (2009). STAR models can deal simultaneously with nonlinear covariate effects and time trends, unit- or cluster-specific heterogeneity, spatial heterogeneity and complex interactions between covariates of different type.

The purpose of this paper is to discuss random intercept selection within the context of STAR models. We aim at shrinking small random intercept coefficients to zero similar to the LASSO (Tibshirani 1996) in frequentist linear models. The selection is based on spike and slab priors for the variances of the random intercept coefficients as proposed for ordinary linear random effects models in Frühwirth-Schnatter and Wagner (2011). The hyperparameters are chosen by theoretical considerations based on the prior inclusion probability of a particular random coefficient given the true effect size. An implementation of the approach is provided in the software package BayesX (Belitz, Brezger, Kneib, Lang, and Umlauf 2013) and the corresponding R interface BayesR (Umlauf, Adler, Kneib, Lang, and Zeileis 2014).

Spike and slab priors are mixtures of two components, the spike with small variance and a comparably flat slab. They were introduced in Mitchell and Beauchamp (1988) for Bayesian variable selection in linear regression models. Whereas initially spike and slab prior distributions were specified for the regression coefficients, e.g. George and McCulloch (1993), recently Ishwaran and Rao (2003) and Ishwaran and Rao (2005) introduced spike and slab priors for their variances. Since then spike and slab priors have been frequently used for Bayesian model choice and variable selection see e.g. Fahrmeir, Kneib, and Konrath (2010)

for variable selection in linear models and Scheipl, Fahrmeir, and Kneib (2012) for variable selection in STAR models. Introductions into the usage of spike and slab priors can be found in Malsiner-Walli and Wagner (2011) and Fahrmeir, Kneib, Lang, and Marx (2013).

Frühwirth-Schnatter and Wagner (2011) demonstrated that spike and slab priors for the random intercept variances are useful for random intercept selection. These priors induce selective shrinkage as random effects assigned to the spike component are heavily shrunk to zero whereas those assigned to the slab experience only little shrinkage. The mixture structure of the prior easily allows for classification of effects as not different from the prior mean– for effects assigned to the spike – and deviating considerably, when assigned to the slab component.

Frühwirth-Schnatter and Wagner (2011) compare different specifications of the spike and the slab priors and unimodal priors for simulated data. They conclude that spike and slab priors outperform unimodal priors with respect to correct classification of non-zero random intercepts and there is little difference between different versions of spike and slab priors. In their analysis the hyperparameters were chosen to yield resulting random effects variance of 1, but for no other values. We fill this gap by an intensive simulation study where we compare spike and slab priors for a wide range of hyperparameters.

The rest of the paper is organized as follows: Section 2 reviews the modeling framework of this paper. Section 3 introduces the spike and slab prior for variances of random effects and discusses the choice of hyperparameters. Technical details regarding the properties of the stochastic search variable selection (SSVS) prior are deferred to the Appendix A. Simulation experiments are presented in Section 4. An illustration for real data is given in Section 5 with an application of the methodology on malnutrition of children in Zambia. The final Section 6 summarizes the paper and provides an outlook for future research.

2 Structured additive models

2.1 Observation model

Suppose that observations $(y_i, \boldsymbol{z}_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$, are given, where y_i is a continuous response variable, and $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iq})'$ and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$ are vectors of covariates. For the variables in $\boldsymbol{z}$ possibly nonlinear effects are assumed whereas the variables in $\boldsymbol{x}$ are modeled in the usual linear way. The components of $\boldsymbol{z}$ are not necessarily continuous covariates. A component may also indicate a time scale, a cluster- or a spatial index (e.g. municipality, district or county) a particular observation pertains to. We assume an additive decomposition of the effects of z_{ij} (and x_{ij}) and obtain the model

$$y_i = f_1(z_{i1}) + \ldots + f_q(z_{iq}) + \boldsymbol{x}_i'\boldsymbol{\gamma} + \varepsilon_i. \tag{1}$$

Here, $f_1, \ldots, f_q$ are nonlinear functions of the covariates $\boldsymbol{z}_i$ and $\boldsymbol{x}_i'\boldsymbol{\gamma}$ is the usual linear part of the model. The errors ε_i are assumed to be mutually independent Gaussian with mean 0 and variance σ^2, i.e. $\varepsilon_i \sim N(0, \sigma^2)$.

The nonlinear effects in (1) are modeled by a basis functions approach, i.e. a particular function f of covariate z is approximated by a linear combination of basis or indicator functions

$$f(z) = \sum_{k=1}^{K} \beta_k B_k(z). \tag{2}$$

The B_k's are known basis functions and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)'$ is a vector of unknown regression coefficients to be estimated. Defining the $n \times K$ design matrix $\mathbf{Z}$ with elements $\mathbf{Z}[i, k] = B_k(z_i)$, the vector $\mathbf{f} = (f(z_1), \ldots, f(z_n))'$ of function evaluations can be written in matrix notation as $\mathbf{f} = \mathbf{Z}\boldsymbol{\beta}$. Accordingly, we obtain

$$\boldsymbol{y} = \boldsymbol{\eta} + \boldsymbol{\varepsilon} = \boldsymbol{Z}_1\boldsymbol{\beta}_1 + \ldots + \boldsymbol{Z}_q\boldsymbol{\beta}_q + \boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{3}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)'$ and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I})$.

2.2 Priors

Effect modeling and priors depend on the covariate or term type. We first describe the general form of priors. In a frequentist setting, overfitting of a particular function $\boldsymbol{f} = \boldsymbol{Z}\boldsymbol{\beta}$ is avoided by defining a roughness penalty on the regression coefficients, see for instance Fahrmeir, Kneib, Lang, and Marx (2013). In a Bayesian framework a standard smoothness prior is a (possibly improper) Gaussian prior of the form

$$p(\boldsymbol{\beta}|\tau^2) \propto \left(\frac{1}{\tau^2}\right)^{\text{rk}(\boldsymbol{K})/2} \exp\left(-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{K}\boldsymbol{\beta}\right) \cdot I(\boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{0}), \tag{4}$$

where $I(\cdot)$ is the indicator function. The key components of the prior are the penalty matrix $\boldsymbol{K}$, the variance parameter τ^2 and the constraint $\boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{0}$. Usually the penalty matrix is rank deficient, i.e. $\text{rk}(\boldsymbol{K}) < K$, resulting in a partially improper prior.

The amount of smoothness is governed by the variance parameter τ^2. A conjugate Inverse Gamma prior is employed for τ^2 (as well as for the error variance parameter σ^2 in models with Gaussian responses), i.e. $\tau^2 \sim \mathcal{G}^{-1}(a, b)$ with small values such as $a = b = 0.001$ for the hyperparameters a and b resulting in an uninformative prior on the log scale. The term $I(\boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{0})$ imposes required identifiability constraints on the parameter vector. A straightforward choice is $\boldsymbol{A} = (1, \ldots, 1)$, i.e. the regression coefficients are centered around zero.

For a continuous covariate z, our basic approach for modeling a smooth function f are Bayesian P-splines introduced in a frequentist setting by Eilers and Marx (1996) and in a Bayesian version by Lang and Brezger (2004). P-splines assume that the unknown functions can be approximated by a polynomial spline which can be written in terms of a linear combination of B-spline basis functions. Hence, the columns of the design matrix $\boldsymbol{Z}$ are given by the B-spline basis functions evaluated at the observations z_i. Lang and Brezger (2004) propose to use first or second order random walks as smoothness priors for the

regression coefficients, i.e.

$$\beta_k = \beta_{k-1} + u_k, \quad \text{or} \quad \beta_k = 2\beta_{k-1} - \beta_{k-2} + u_k, \tag{5}$$

with Gaussian errors $u_k \sim N(0, \tau^2)$ and diffuse priors $p(\beta_1) \propto const$, or $p(\beta_1)$ and $p(\beta_2) \propto const$, for initial values. This prior is of the form (4) with penalty matrix given by $\boldsymbol{K} = \boldsymbol{D'D}$, where $\boldsymbol{D}$ is a first or second order difference matrix.

Suppose now that covariate z is an index variable that indicates the unit or cluster a particular observation belongs to. In this case, it is common practice to introduce unit- or cluster specific i.i.d. Gaussian random intercepts or slopes, see e.g. Diggle, Heagerty, Liang, and Zeger (2002). Suppose z can take the values $1, \ldots, K$. Then, an i.i.d. random intercept can be incorporated into our framework of structured additive regression by assuming $f(k) = \beta_k \sim N(0, \tau^2)$, $k = 1, \ldots, K$. The design matrix $\boldsymbol{Z}$ is now a 0/1 incidence matrix with dimension $n \times K$. The penalty matrix is the $K \times K$ identity matrix, i.e. $\boldsymbol{K} = \boldsymbol{I}$.

3 Random intercept selection

In this section we now replace the usual Inverse Gamma prior for the variance τ^2 of an i.i.d. random intercept by spike and slab priors that allow for random effect selection.

Although random intercepts can be formally subsumed within the STAR predictor (see the previous section) we include the random intercept term explicitly in our model, i.e.

$$y_i = f_1(z_{i1}) + \ldots + f_q(z_{iq}) + \boldsymbol{x}_i'\boldsymbol{\gamma} + \beta_{z_i} + \varepsilon_i, \tag{6}$$

where $z_i \in \{1, \ldots, K\}$ is the value of the cluster variable z for the i-th observation.

3.1 Prior

For the random intercepts β_k, $k = 1, \ldots, K$ we specify a spike and slab prior distribution by the following hierarchy:

$$\beta_k|\delta_k, \tau^2 \sim \mathcal{N}\left(0, r(\delta_k)\tau^2\right)$$

$$\tau^2 \sim \mathcal{G}^{-1}\left(\nu, Q\right)$$

$$p(\delta_k|\omega) = \begin{cases} \omega & \delta_k = 1 \\ 1-\omega & \delta_k = 0 \end{cases}$$

$$\omega \sim \text{Beta}\left(a_0, b_0\right)$$

where

$$r(\delta_k) = \begin{cases} r & \delta_k = 0 \\ 1 & \delta_k = 1, \end{cases}$$

and $r << 1$ is a constant. Our default for r in our implementation is $r = 0.005$. The random effects β_k basically follow a centered Gaussian distribution, with variance depending on the indicator variable δ_k. $\delta_k = 1$ indicates the slab component, with variance τ^2 resulting from the Inverse Gamma distribution and $\delta_k = 0$ the spike with considerably smaller variance $r\tau^2$.

Hence random intercepts are assumed to i.i.d follow a scale mixture of Normals

$$p\left(\beta_k|\omega, \tau^2\right) \propto \frac{1-\omega}{r\tau} \exp\left(-\frac{\beta_j^2}{2r\tau^2}\right) + \frac{\omega}{\tau} \exp\left(-\frac{\beta_j^2}{2\tau^2}\right),$$

conditional on the mixture weight ω and the variance parameter τ^2 .

Smoothing of random intercepts is induced by the hyper-priors on these parameter. The hyper-prior on ω induces smoothing of the indicator variables, which are independent conditioning on ω, but marginally dependent for proper prior distributions, i.e. $a_0, b_0 > 0$. As,

conditional on the indicator variables, the marginal distribution of the random intercepts is a scaled Student distribution with 2ν degrees of freedom (see appendix A) the hyper-prior on the variance parameter induces smoothing mainly within the mixture components.

The main purpose of the mixture specification of the prior is to allow for selective shrinkage: due to the smaller variance, random intercepts assigned to the spike component will be shrunk much more severly than those assigned to the slab component. Additionally based on the posterior distribution of the indicator variables, random intercepts can be classified as either negligable or considerably different from zero, if the posterior probability of $\delta_k = 1$ exceeds a certain treshold.

However, indicators are marginally dependent for a proper prior distribution, i.e. $a_0, b_0 > 0$. Considering two random intercepts β_k, β_l, the marginal prior probability that both are assigned to the slab component is

$$P(\delta_k = 1, \delta_l = 1) = \int P(\delta_k = 1, \delta_l = 1|\omega)p(\omega)d\omega = \int \omega^2 p(\omega)d\omega =$$
$$= \mathrm{E}(\omega^2) = \frac{a_0}{a_0 + b_0}\frac{a_0 + 1}{a_0 + b_0 + 1}$$

which is larger than $P(\delta_k = 1)P(\delta_l = 1) = \mathrm{E}(\omega)^2 = \left(\frac{a_0}{a_0+b_0}\right)^2$ for $b_0 > 1$

Correspondingly

$$P(\delta_k = 0, \delta_l = 0) = \mathrm{E}((1-\omega)^2) > P(\delta_k = 0)P(\delta_l = 0)$$

for $a_0 > 1$.

3.2 Full conditionals

Random intercept selection using spike and slab priors for the variances can be easily incoprorated into existing sampling schemes for STAR models. The MCMC updates of the regression coefficients of the nonlinear functions and linear effects remain unchainged. The parameters for the variance τ^2 can be updated by simple Gibbs steps because their full conditionals are known distributions. We have the following full conditionals:

1. Full conditional of δ_k:

$$p(\delta_k = 1 \,|\, \beta_k, \omega, \tau^2) = \frac{\omega p(\beta_k \,|\, \tau^2, \delta_k = 1)}{(1-\omega)p(\beta_k \,|\, \tau^2, \delta_k = 0) + \omega p(\beta_k \,|\, \tau^2, \delta_k = 1)} = \\ = \frac{1}{1 + \frac{1-\omega}{\omega} L_k},$$

where

$$L_k = \frac{\varphi(\beta_k; 0; r\,\tau^2)}{\varphi(\beta_k; 0; \tau^2)} = \frac{1}{\sqrt{r}} e^{-\frac{\beta_k^2}{2\tau^2}(1/r - 1)}.$$

Here $\varphi(\beta_k; 0; \tau^2)$ denotes the density of the $N(0, \tau^2)$ distribution evaluated at β_k.

2. Full conditional of τ^2:

$$\tau^2 \,|\, \nu \sim \mathcal{G}^{-1}\left(\nu + K/2, Q + \frac{1}{2}\sum_{k=1}^{K} \frac{\beta_k^2}{r(\delta_k)}\right)$$

3. Full conditional of ω:

$$\omega \,|\, \boldsymbol{\delta} \sim \text{Beta}\left(a_0 + \sum_{k=1}^{K} \delta_k, b_0 + K - \sum_{k=1}^{K} \delta_k\right)$$

3.3 Choosing the hyperparameters

The parameters β_k are set up with a mixture of prior distributions to separate effects close to zero from larger ones. The range of a priori likely parameter values is thereby driven by the variance τ^2. A criterion to specify the prior distribution of the parameter variance, i.e. the hyperparameters ν and Q of the Inverse Gamma distribution, is therefore favorable. For that purpose we establish a principle to determine ν and Q in order to

- ensure inclusion of a priori relevant effects with a high probability (*sensitivity*)

$$\mathbb{P}\left(\delta_k = 1 \,|\, |\beta_k| \geq q_1\right)$$

and

- at the same time let "small" effects be excluded from the model (*specificity*)

$$\mathbb{P}\left(\delta_k = 0 \,|\, |\beta_k| \leq q_2\right).$$

In Appendix A we show that

$$p\left(\delta_k = 1|\beta_k\right) \approx \frac{1}{1 + \frac{b_0}{a_0} \frac{p_{2\nu}\left(\beta_k; r\frac{Q}{\nu}\right)}{p_{2\nu}\left(\beta_k; \frac{Q}{\nu}\right)}}$$

with $p_{2\nu}\left(\beta_k|\delta_k; r\left(\delta_k\right)\frac{Q}{\nu}\right)$ being the density of a scaled Student-t distribution with 2ν degrees of freedom, that is,

$$\sqrt{\frac{\nu}{r\left(\delta_i\right)Q}}\beta_k|\delta_k \sim t_{2\nu}.$$

Concrete probabilities are then computed after specifying the interesting minimum (or maximum) effect size by

$$\mathbb{P}\left(\delta_k = 1||\beta_k| \leq q\right) = \frac{1}{1 + \frac{b_0}{a_0} \frac{\mathbb{P}_{t_{2\nu}}\left(\sqrt{\frac{\nu}{rQ}}|\beta_k| \leq q\right)}{\mathbb{P}_{t_{2\nu}}\left(\sqrt{\frac{\nu}{Q}}|\beta_k| \leq q\right)}}.$$

For a selection of hyperparameters the corresponding probabilities are listed in the Tables 1 - 5. The following general observations can be made:

- The higher the probability of values around zero in the slab component, the higher the chance that small effects are detected, but also the worse the separation of irrelevant ones.
  - ν drives the degrees of freedom of the Student t-distribution. A higher ν leads to a smaller variance and a smaller kurtosis and both increase the probability mass around zero.
  - Q is part of the scale parameter where a higher Q describes a wider spread of the effects and so reduces the probability of values around zero.
- $\nu = 0.001$, $Q = 0.001$ seems an undesirable combination as sensitivity remains at its low but specificity can be improved. Note, that for a single Inverse Gamma distribution this is a standard choice to incorporate an uninformative prior on the log scale.

As a brief summary we can state that

- an increase in ν (Q fixed) increases the sensitivity but decreases the specificity,
- an increase in Q (ν fixed) decreases the sensitivity but increases the specificity.

To investigate the effect on the posterior inclusion probability we select hyperparameters to model

- high sensitivity and high specificity ($\nu = 10$, $Q = 10$)
- high sensitivity and low specificity ($\nu = 1$, $Q = 0.001$)
- low sensitivity and high specificity ($\nu = 1$, $Q = 25$),

see also Figure 1. These settings were integrated into the simulation study and compared to the well established $\nu = 5$ and $Q = 25$ hyperparameter pair.

q	**Sensitivity**					**Specificity**				
	Q=0.001	Q=1	Q=3	Q=10	Q=25	Q=0.001	Q=1	Q=3	Q=10	Q=25
0.01	0.502	0.500	0.500	0.500	0.500	0.938	0.984	0.984	0.984	0.984
0.05	0.502	0.501	0.501	0.500	0.500	0.837	0.978	0.981	0.983	0.984
0.09	0.502	0.501	0.501	0.501	0.500	0.792	0.971	0.977	0.981	0.983
0.13	0.502	0.501	0.501	0.501	0.500	0.767	0.964	0.973	0.979	0.982
0.17	0.502	0.501	0.501	0.501	0.501	0.750	0.958	0.969	0.977	0.980
0.21	0.502	0.501	0.501	0.501	0.501	0.739	0.952	0.965	0.975	0.979
0.25	0.502	0.501	0.501	0.501	0.501	0.730	0.946	0.962	0.972	0.978

Table 1: Marginal prior inclusion probabilities, $\nu = 0.001$

q	**Sensitivity**					**Specificity**				
	Q=0.001	Q=1	Q=3	Q=10	Q=25	Q=0.001	Q=1	Q=3	Q=10	Q=25
0.01	0.997	0.627	0.570	0.537	0.523	0.821	0.983	0.984	0.984	0.984
0.05	1.000	0.917	0.824	0.701	0.627	0.573	0.963	0.975	0.981	0.983
0.09	1.000	0.969	0.922	0.821	0.726	0.528	0.939	0.962	0.975	0.980
0.13	1.000	0.984	0.958	0.889	0.802	0.514	0.915	0.948	0.968	0.976
0.17	1.000	0.990	0.974	0.926	0.856	0.508	0.893	0.934	0.960	0.972
0.21	1.000	0.993	0.982	0.948	0.892	0.506	0.871	0.920	0.953	0.967
0.25	1.000	0.995	0.987	0.962	0.917	0.504	0.851	0.907	0.945	0.963

Table 2: Marginal prior inclusion probabilities, $\nu = 1$

q	**Sensitivity**					**Specificity**				
	Q=0.001	Q=1	Q=3	Q=10	Q=25	Q=0.001	Q=1	Q=3	Q=10	Q=25
0.01	1	0.758	0.643	0.573	0.545	0.716	0.981	0.983	0.984	0.984
0.05	1	0.998	0.980	0.880	0.758	0.509	0.938	0.962	0.976	0.981
0.09	1	1.000	0.999	0.979	0.910	0.501	0.894	0.936	0.963	0.974
0.13	1	1.000	1.000	0.996	0.971	0.500	0.854	0.910	0.948	0.966
0.17	1	1.000	1.000	0.999	0.990	0.500	0.819	0.885	0.934	0.956
0.21	1	1.000	1.000	1.000	0.996	0.500	0.786	0.863	0.919	0.947
0.25	1	1.000	1.000	1.000	0.998	0.500	0.758	0.841	0.905	0.938

Table 3: Marginal prior inclusion probabilities, $\nu = 3$

q	**Sensitivity**					**Specificity**				
	Q=0.001	Q=1	Q=3	Q=10	Q=25	Q=0.001	Q=1	Q=3	Q=10	Q=25
0.01	1	0.84	0.696	0.600	0.560	0.665	0.979	0.983	0.984	0.984
0.05	1	1.00	0.998	0.952	0.840	0.501	0.920	0.952	0.972	0.979
0.09	1	1.00	1.000	0.997	0.971	0.500	0.865	0.917	0.953	0.969
0.13	1	1.00	1.000	1.000	0.996	0.500	0.818	0.885	0.933	0.957
0.17	1	1.00	1.000	1.000	0.999	0.500	0.776	0.855	0.915	0.944
0.21	1	1.00	1.000	1.000	1.000	0.500	0.740	0.828	0.897	0.932
0.25	1	1.00	1.000	1.000	1.000	0.500	0.708	0.802	0.880	0.920

Table 4: Marginal prior inclusion probabilities, $\nu = 5$

q	Sensitivity					Specificity				
	Q=0.001	Q=1	Q=3	Q=10	Q=25	Q=0.001	Q=1	Q=3	Q=10	Q=25
0.01	1	0.943	0.79	0.650	0.589	0.599	0.974	0.981	0.983	0.984
0.05	1	1.000	1.00	0.995	0.943	0.500	0.890	0.933	0.962	0.974
0.09	1	1.000	1.00	1.000	0.998	0.500	0.819	0.886	0.934	0.957
0.13	1	1.000	1.00	1.000	1.000	0.500	0.761	0.844	0.907	0.939
0.17	1	1.000	1.00	1.000	1.000	0.500	0.713	0.806	0.882	0.922
0.21	1	1.000	1.00	1.000	1.000	0.500	0.673	0.772	0.859	0.906
0.25	1	1.000	1.00	1.000	1.000	0.500	0.640	0.742	0.837	0.890

Table 5: Marginal prior inclusion probabilities, $\nu = 10$

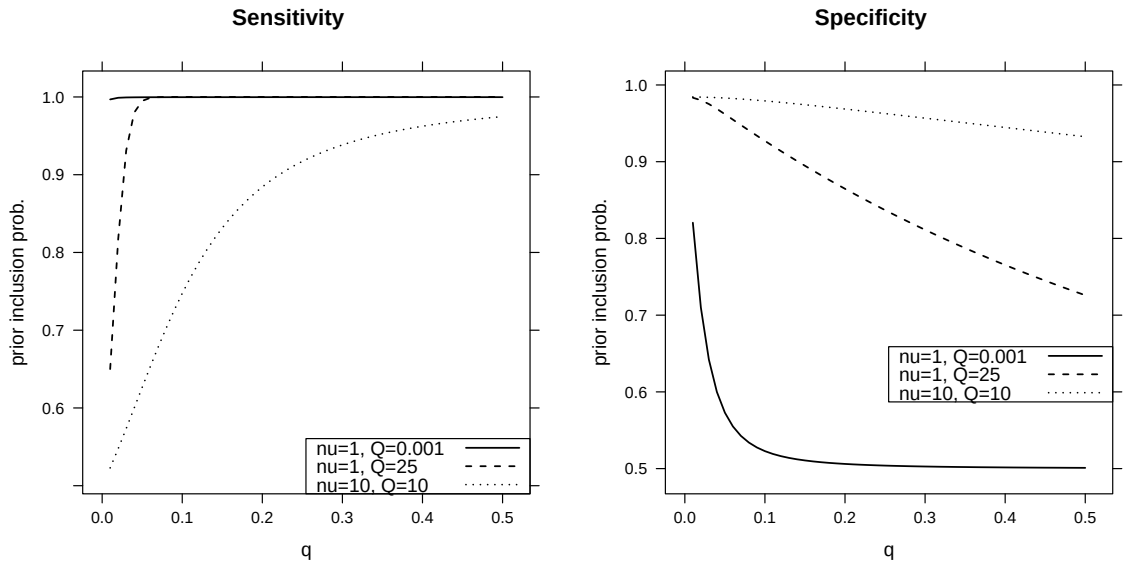

Figure 1: The sensitivity and specificity of the considered scenarios depending on the random effect size q

4 Simulation experiments

A simulation study was conducted to compare the spike and slab prior, denoted by *SSVS*, to the well established Inverse Gamma prior distribution for random intercepts, denoted by *IG*. Therefore we set up different random intercept structures and compare the overall performance, estimated coefficients as well as the random intercept selection.

4.1 Simulation designs

The response is driven by two smooth effects, a random intercept and an error term:

$$y_i \sim \log(z_{i1}) + \sin(z_{i2}) + \beta_{z_i} + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}\left(0, 0.5^2\right).$$

The values of z_1 and z_2 are distributed on equidistant grids in $[-3, 3]$ and $[0.3, 3]$, respectively. The domains of the smooth components were selected in order to ensure effects of similar magnitude. Cubic B-spline representations with 22 basis functions and a random walk two penalty were chosen to model the smooth functions. The $n = 1000$ observations were separated into $N = 100$ clusters, each of size $T = 10$. The different specifications for the random intercept are given as follows:

- **full** $\quad \beta_k \sim \mathcal{N}(0, 1)$

- **none** $\quad \beta_k = 0, \qquad k = 1, \ldots, 100$

- **partial** $\quad \beta_k \begin{cases} = 0, & k = 1, \ldots, 75 \\ \sim \mathcal{N}(0, 1), & k = 76, \ldots, 100. \end{cases}$

The classification of observations to clusters was kept fix over the $M = 500$ repetitions as well as the random effects. The hyperparameters for the prior distribution of the random effects variance were set to $\nu = 0.001$, $Q = 0.001$ for *IG*. For *SSVS* we used the hyperparameter settings described and justified in Section 3.3. Since the results turned out be practically

unaffected by the choice of hyperparameters, the following presentation of results is restricted to the choice $\nu = 5$ and $Q = 25$.

All models were estimated using the open source software `BayesX`.

4.2 Evaluation of model fit

The overall fit (*IG* versus *SSVS*) is compared by the deviance information criterion (DIC), see Spiegelhalter, Best, Carlin, and van der Linde (2002). The DIC also allows for model choice, i.e. the discrimination between the *IG* or the *SSVS* prior. A rough rule of thumb says that DIC differences of 10 and more between two competing models indicate that the model with the lower DIC is superior. Differences between 5 and 10 are "substantial" but not definitely conclusive, while differences below 5 are not supportive for either model.

The quality of the estimated smooth functions are measured by the average estimated function as a measure of the bias and the mean squared error

$$MSE(f_j) = \frac{1}{K}\sum_{i=1}^{K}\left(\hat{f}\left(z_{ij}\right) - f\left(z_{ij}\right)\right)^2, \qquad j = 1, 2.$$

The quality of the estimated random effects is measured by the average estimated random effect as a measure of the bias and the mean squared error

$$MSE_{re} = \frac{1}{K}\sum_{k=1}^{K}\left(\hat{\beta}_k - \beta_k\right)^2.$$

Furthermore, in case of *SSVS* priors we constructed tables of inclusion as follows: A cluster effect is assumed to be included in the model if the posterior inclusion probability exceeds 0.5. The inclusion tables then indicate the percentage of replications the respective cluster effect is included in the model.

4.3 Results

The results of our simulation experiments are summarized through Figures 2 - 6. Figure 2 displays boxplots of DIC differences between the fits based on Inverse Gamma and spike and

slab priors for random effects variances. Figures 3 and 4 show average estimated functions as well as MSE's for f_1 and f_2 in case of a partial random effects. For the other cases (no and full random effects) the results are similar and therefore omitted. Figure 5 displays boxplots of MSE's for the estimated random effects while Figure 6 shows relative inclusion frequencies depending on the size of the random effect.

We can draw the following conclusions:

- *Overall fit:* Figure 2 shows that the SSVS prior primarily pays off in the case were some of the random effects coefficients are zero (the partial scenario). Here the DIC differences are huge with average difference of about 45 units. If in fact there are no nonzero random effects (none scenario) both priors IG and SSVS perform similar with a slight tendency in favor of the SSVS prior. In cases where all random effects coefficients are not equal to zero (full scenario) the IG prior seems to perform slightly better than the SSVS prior although the DIC differences are usually within a range between 5 and 10 units and therefore not conclusive according to our rule of thumb.

- *Fit of smooth functions:* Figures 3 and 4 suggest that the choice of the random effects prior (IG versus SSVS) does not affect the fit of the smooth functions f_1 and f_2.

- *Fit of random effects:* Figure 5 largely confirms our findings based on the DIC. The proposed SSVS prior performs superior if all or some of the random effects coefficients are in fact zero (upper and middle panel). If all random effects are nonzero the usual IG prior performs slightly better than the SSVS prior.

- *Inclusion frequencies:* Figure 6 shows that the random coefficient selection works very well. Random coefficients that are actually zero (partial scenario) are almost never included into the model while for nonzero coefficients the inclusion frequencies are usually close to one. Exceptions are some effects which are nonzero but close to zero. Here the inclusion frequencies can be comparably low in the partial and full setting.

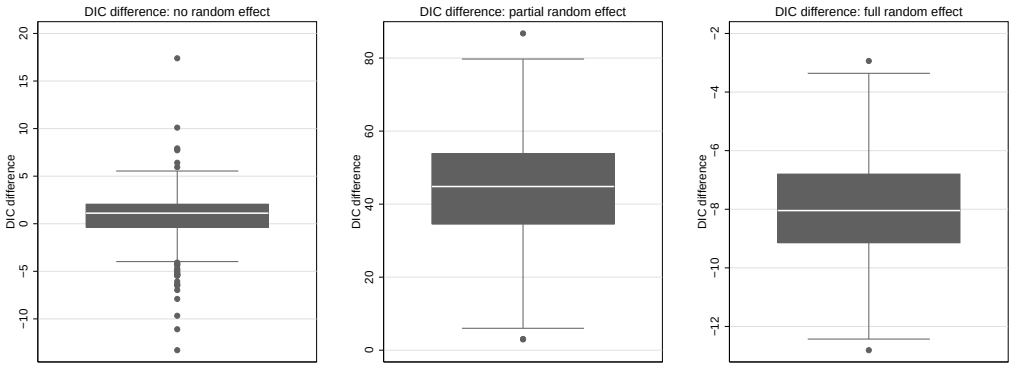

Figure 2: DIC differences between fits based on Inverse Gamma and spike and slab priors for random effects variances.

Note that we have omitted the results for the scenario where all random coefficients are zero. The reason is that the relative inclusion frequencies are practically zero indicating that the selection process works extremely well in this scenario.

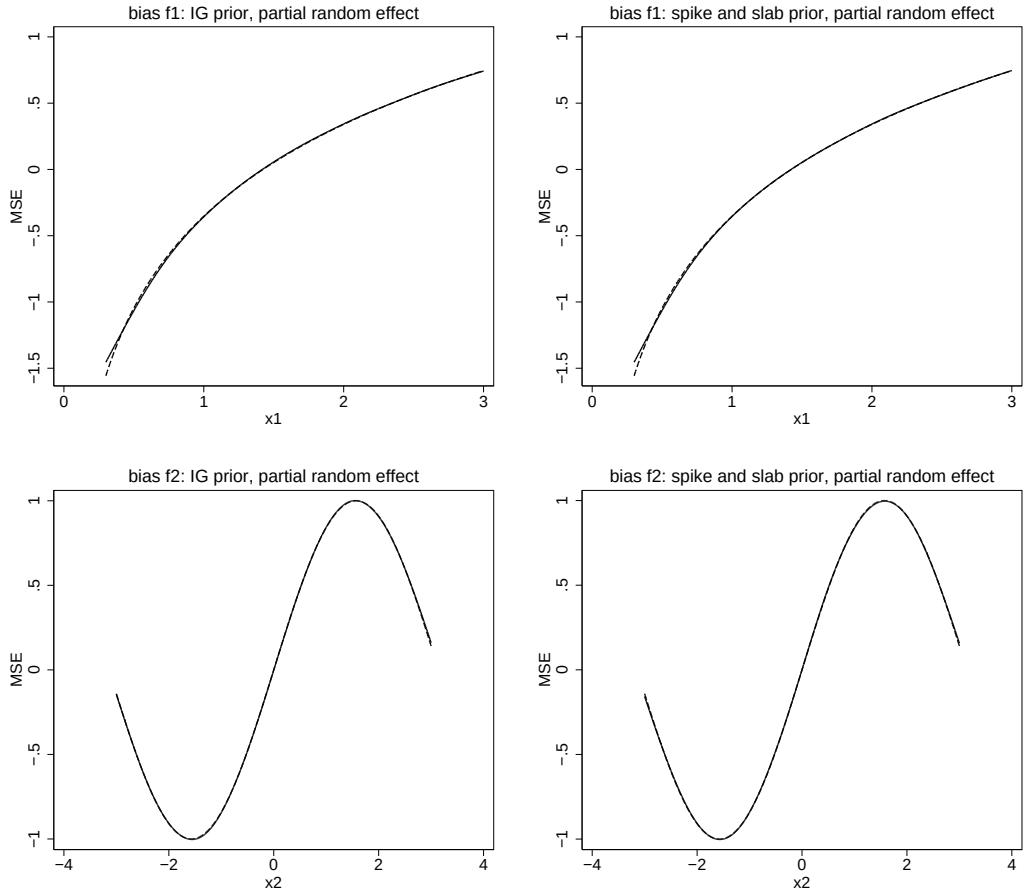

Figure 3: Average estimated functions for f_1 and f_2 in case of a partial random effect. The left panel corresponds to Inverse Gamma priors for the random effects variances and the right panel to spike and slab priors.

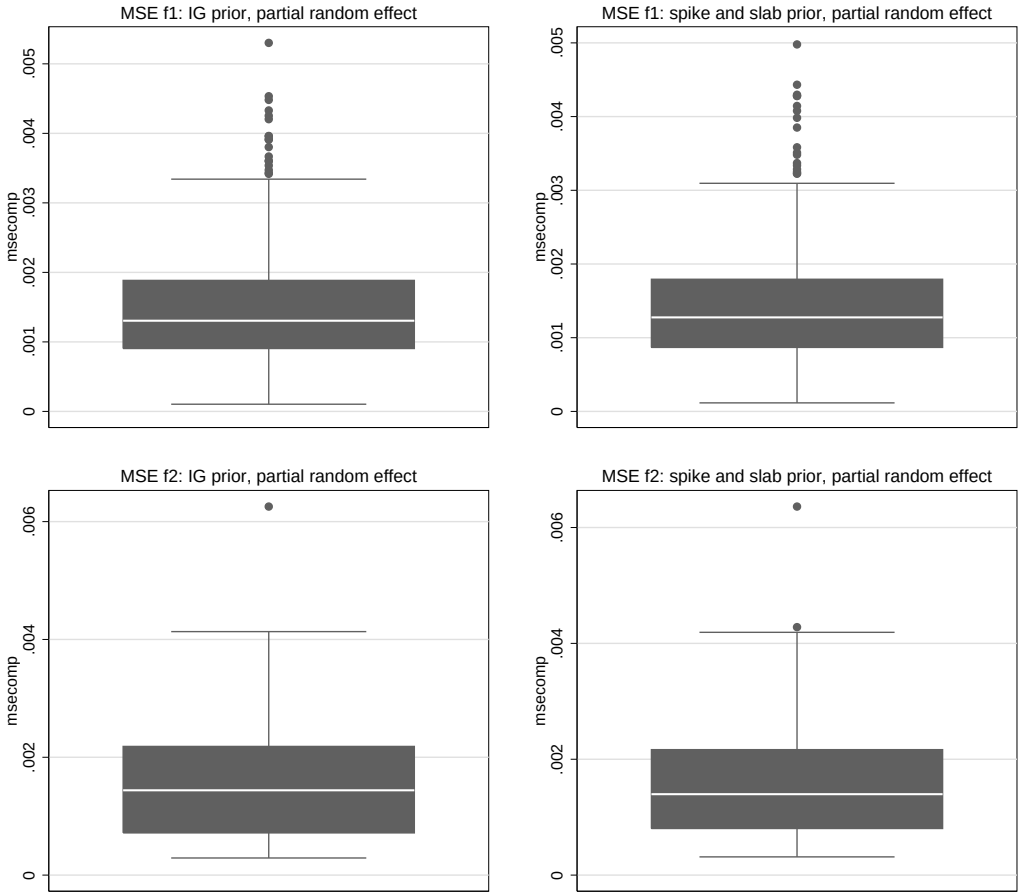

Figure 4: MSE for f_1 and f_2 in case of a partial random effect. The left panel corresponds to Inverse Gamma priors for the random effects variances and the right panel to spike and slab priors.

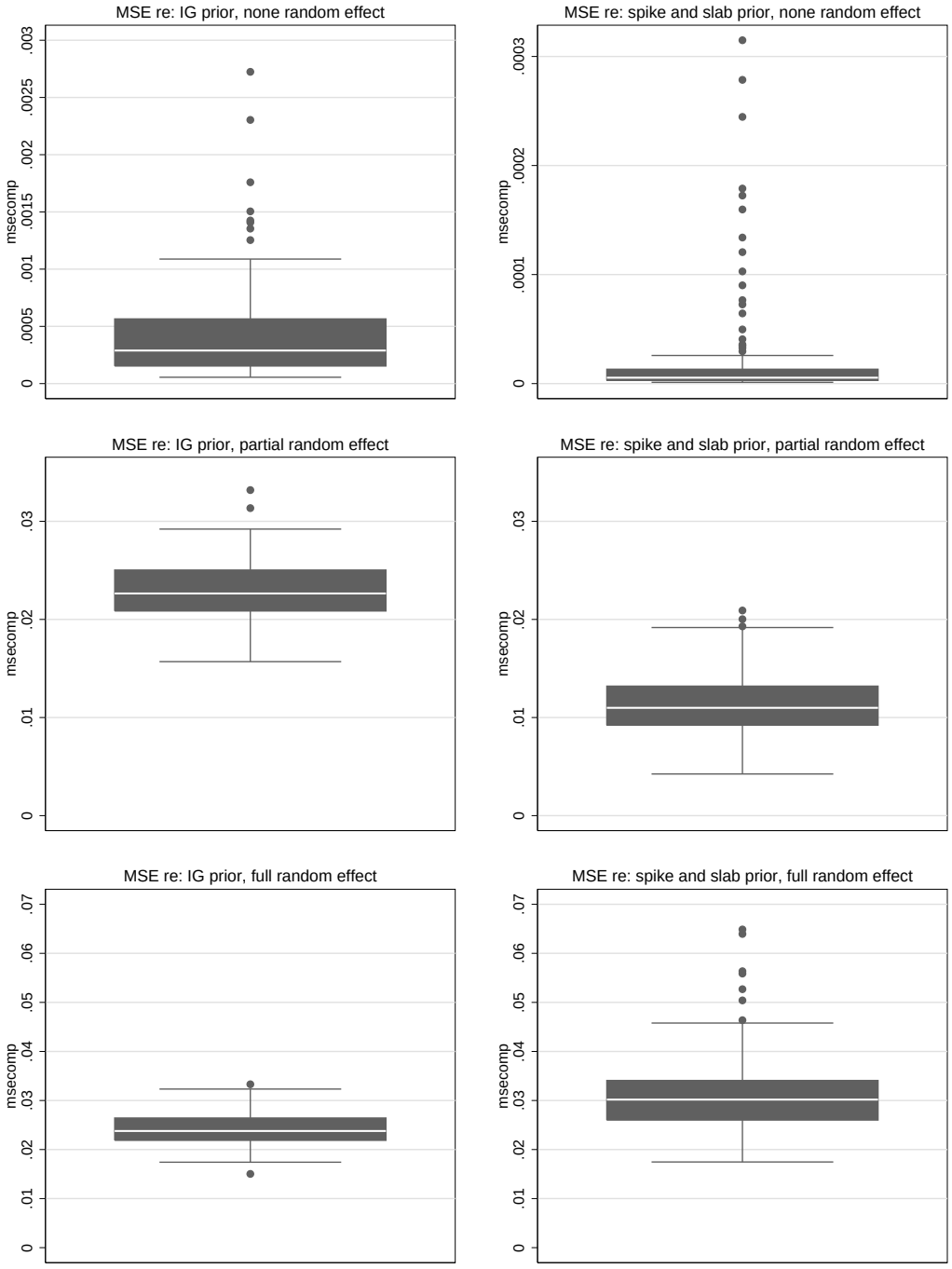

Figure 5: MSE for random effects. The left panel corresponds to Inverse Gamma priors for the random effects variances and the right panel to spike and slab priors.

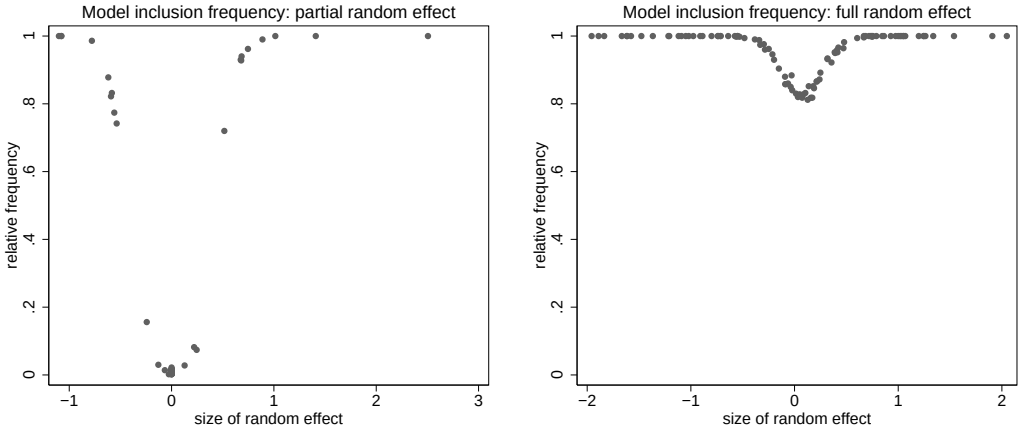

Figure 6: Relative inclusion frequencies depending on the size of the random effect.

5 Application: Malnutrition in Zambia

In this section we illustrate the application of spike and slab priors for random intercepts using DHS (demographic and health survey) data from 1992 on undernutrition in Zambia. The data set has already been used in various illustrations, see among others the detailed case study on structured additive regression in Fahrmeir, Kneib, Lang, and Marx (2013). The book also contains a variable description and summary statistics of all variables. Here we build on Fahrmeir, Kneib, Lang, and Marx (2013) and use their model

$$\begin{aligned} zscore_i &= \boldsymbol{\gamma}'\boldsymbol{x}_i + f_1(m\_agebirth_i) + f_2(m\_height_i) + f_3(m\_bmi_i) + \\ &\quad f_4(c\_breastf_i \cdot c\_age_i) + \beta_{district_i} + \varepsilon_i, \end{aligned}$$

where $zscore_i$ is a zscore measuring chronic undernutrition, $f_1, \ldots, f_4$ are possibly nonlinear one or two dimensional functions of the continuous covariates $m\_agebirth$ (mothers age at birth), m_height (mothers height), m_bmi (mothers body mass index), $c\_breastf$ (duration of breastfeeding) and c_age (childs age). The term $\beta_{district_i}$ is a random intercept to cope with district specific heterogeneity. The one dimensional functions f_1, f_2, f_3 are modeled by cubic P-splines with 20 inner knots, the two dimensional function f_4 is modeled through a Gaussian field (kriging term). For the district specific random effect we applied our spike and slab prior with hyperparameters $\nu = 5, Q = 25$ for the inverse gamma prior on τ^2 and $a_0 = b_0 = 1$ for the beta prior on ω (uniform prior). As a kind of sensitivity check we also used the hyperparameter combinations $\nu = 1, Q = 25$ and $\nu = 10, Q = 10$. We also compared the results with a standard inverse gamma prior with hyperparameters $a = b = 0.001$ for the variance τ^2 of the random effects.

As could have been expected from the simulation study, the results for the linear effects $\boldsymbol{\gamma}'\boldsymbol{x}$ and the nonlinear effects $f_1, \ldots, f_4$ are almost unaffected from the random effects modeling and practically identical to Fahrmeir, Kneib, Lang, and Marx (2013). We therefore report only on the results for the random effects. Figure 7 displays maps of the estimated random effects, where the left panel corresponds to a spike and slab prior with hyperparameters

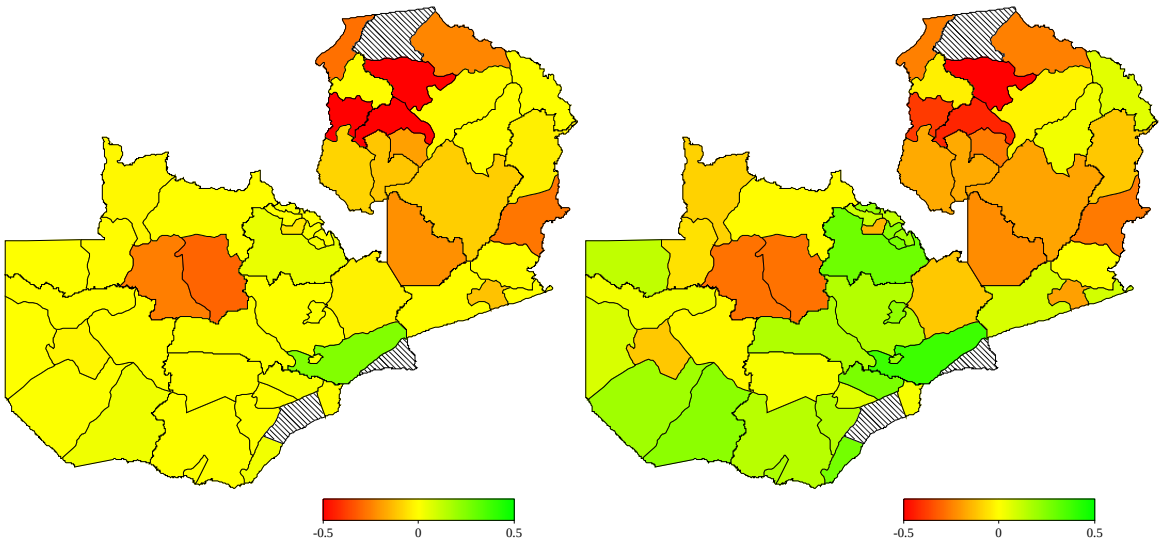

Figure 7: Posterior means of district specific random effects. The left panel displays results for the spike and slab prior with hyperparameters $\nu = 5, Q = 25$, the right panel corresponds to the standard IG prior.

$\nu = 5, Q = 25$ and the right panel to an IG prior. To ease comparison with the IG prior Figure 8 displays the random effects in ascending order of size. Additionally the districts are marked where $\delta_k > 0.5$ for the posterior mean of the inclusion parameters δ_k. A key finding of the analysis is that the spike and slab prior shrinks small effects towards zero while large effects appear to be more pronounced than with an IG prior. Using the IG prior it seems that small effects are overestimated and large effects are underestimated as an effect of the overall (nonlocal) variance parameter. In contrast the spike and slab prior shrinks unimportant effects towards zero while important effects are not affected. Figure 8 also shows that the districts with inclusion probability larger than 0.5 correspond largely to the districts where the 95% credible interval of the effect based on an IG prior does not cover zero.

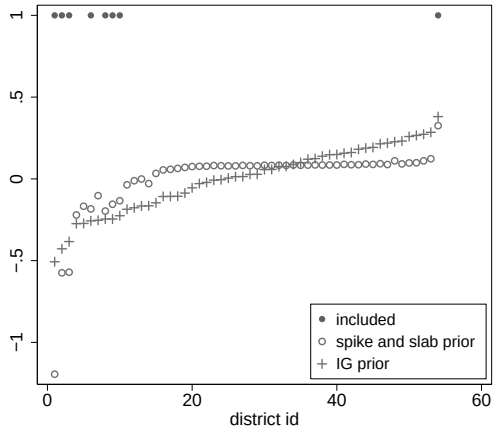

Figure 8: Posterior means of estimated random effects in ascending order of size. Results for spike and slab priors are denoted by hollow dots, results for IG priors by plus signs. The black dots indicate a posterior mean above 0.5 for the corresponding inclusion parameter δ_k.

6 Conclusion

This paper develops random intercept selection within the context of models with a structured additive predictor. The selection of random intercept coefficients is based on spike and slab priors for the variance of the random effects distribution. Our extensive simulation experiments show that the proposed prior performs superior to the usual Inverse Gamma prior if some or all random effects coefficients are in fact zero. In cases where all random effects coefficients are nonzero spike and slab priors and the standard Inverse Gamma prior perform almost equally well. The distinction between spike and slab priors and Inverse Gamma priors can be done in practice using the deviance information criterion.

A Marginal prior inclusion probability given the effect size

The marginal prior inclusion probability given the effect β_k can formally be computed as

$$p(\delta_k|\beta_k) = \frac{p(\beta_k|\delta_k)p(\delta_k)}{p(\beta_k)}.$$

From the joint distribution

$$p\left(\beta_k, \tau^2, \delta_k, \omega\right) = p\left(\beta_k|\tau^2, \delta_k\right) p\left(\tau^2\right) p\left(\delta_k|\omega\right) p\left(\omega\right),$$

we get

$$p\left(\beta_k|\delta_k\right) = \int p\left(\beta_k|\tau^2, \delta_k\right) p(\tau^2) d\tau^2 = p_{2\nu}\left(\beta_k|\delta_k; r\left(\delta_k\right)\frac{Q}{\nu}\right)$$

where $p_{2\nu}\left(\beta_k|\delta_k; r\left(\delta_k\right)\frac{Q}{\nu}\right)$ is the density of a scaled Student-t distribution with 2ν degrees of freedom, i.e.,

$$\sqrt{\frac{\nu}{r\left(\delta_i\right)Q}}\beta_k|\delta_k \sim t_{2\nu}.$$

Also

$$p(\delta_k) \approx \int p\left(\delta_k|\omega\right) p\left(\omega\right) d\omega = \begin{cases} \frac{a_0}{a_0+b_0} & \delta_k = 1 \\ \\ \frac{b_0}{a_0+b_0} & \delta_k = 0 \end{cases},$$

such that the random effect β_k follows a mixture of scaled Student-t distributions with 2ν degrees of freedom

$$\beta_k \sim \frac{b_0}{a_0+b_0} t_{2\nu}\left(r\frac{Q}{\nu}\right) + \frac{a_0}{a_0+b_0} t_{2\nu}\left(\frac{Q}{\nu}\right),$$

where the mixture weights are the prior expected exclusion and inclusion probability, respectively.

Hence finally

$$p\left(\delta_k = 1|\beta_k\right) = \frac{\frac{a_0}{a_0+b_0} p_{2\nu}\left(\beta_k; \frac{Q}{\nu}\right)}{\frac{b_0}{a_0+b_0} p_{2\nu}\left(\beta_k; r\frac{Q}{\nu}\right) + \frac{a_0}{a_0+b_0} p_{2\nu}\left(\beta_k; \frac{Q}{\nu}\right)} \tag{7}$$

$$= \frac{1}{1 + \frac{b_0}{a_0}\frac{p_{2\nu}\left(\beta_k; r\frac{Q}{\nu}\right)}{p_{2\nu}\left(\beta_k; \frac{Q}{\nu}\right)}}. \tag{8}$$

From (7) we can see that the prior inclusion probability depends on the share of the prior probability in the slab component of the total marginal prior effect probability.

References

Belitz, C., A. Brezger, T. Kneib, S. Lang, and N. Umlauf (2013). *BayesX: Software for Bayesian Inference in Structured Additive Regression Models*. Version 2.1.

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

Diggle, P. J., P. Heagerty, K.-L. Liang, and S. L. Zeger (2002). *Analysis of Longitudinal Data (2. Auflage)*. Oxford University Press, Oxford.

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

Fahrmeir, L., T. Kneib, and S. Konrath (2010). Bayesian regularisation in structured additive regression: a unifying perspective on shrinkage, smoothing and predictor selection. *Statistics and Computing 20*, 203–219.

Fahrmeir, L., T. Kneib, and S. Lang (2004). Penalized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica 14*, 731–761.

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

Frühwirth-Schnatter, S. and H. Wagner (2011). Bayesian variable selection for random intercept modelling of Gaussian and non-Gaussian data. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West (Eds.), *Bayesian Statistics 9*, pp. 165–200. Oxford University Press.

George, E. I. and R. McCulloch (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association 88*, 881–889.

Ishwaran, H. and S. Rao (2003). Detecting differentially expressed genes in microarrays using Bayesian model selection. *Journal of the Americal Statistical Association 98*, 438–455.

Ishwaran, H. and S. Rao (2005). Spike and slab variable selection: frequentist and Bayesian strategies. *Annals of Statistics 33*, 730–773.

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

Malsiner-Walli, G. and H. Wagner (2011). Comparing spike and slab priors for Bayesian variable selection. *Austrian Journal of Statistics 40*, 241–264.

Mitchell, T. and J. J. Beauchamp (1988). Bayesian variable selection in linear regression. *83*, 1023 – 1032.

Rigby, R. A. and D. M. Stasinopoulos (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society C 54*, 507–554.

Rue, H., S. Martino, and C. Nicolas (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society B 71*, 319–392.

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

Scheipl, F., L. Fahrmeir, and T. Kneib (2012). Function selection in structured additive regression models based on spike-and-slab priors. *Journal of the American Statistical Association*. To appear.

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

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B 58*, 267–288.

Umlauf, N., D. Adler, T. Kneib, S. Lang, and A. Zeileis (2014). Structured additive regression models: An R interface to bayesx. *Journal of Statistical Software*.

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

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

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

2015-02 **Helene Roth, Stefan Lang, Helga Wagner:** Random intercept selection in structured additive regression models

2015-01 **Alice Sanwald, Engelbert Theurl:** Out-of-pocket expenditures for pharmaceuticals: Lessons from the Austrian household budget survey

2014-32 **Esther Blanco, Tobias Haller, James M. Walker:** Externalities in appropriation: Responses to probabilistic losses

2014-31 **Michael Kirchler, Stefan Palan:** Friendliness pays off! Monetary and immaterial gifts in consumer-salesperson interactions

2014-30 **Alice Sanwald, Engelbert Theurl:** Out-of-pocket expenditures of private households for dental services - Empirical evidence from Austria

2014-29 **Stefan Borsky, Esther Blanco:** Setting one voluntary standard in a heterogeneous Europe - EMAS, corruption and stringency of environmental regulations

2014-28 **Eberhard Feess, Christian Grund, Markus Walzl, Ansgar Wohlschlegel:** Competing trade mechanisms and monotone mechanism choice

2014-27 **Esther Blanco, E. Glenn Dutcher, Tobias Haller:** To mitigate or to adapt? Collective action under asymmetries in vulnerability to losses

2014-26 **Alice Sanwald, Thomas Schober:** Follow your heart: Survival chances and costs after heart attacks - An instrumental variable approach

2014-25 **Eric Mayer, Sebastian Rüth, Johann Scharler:** Total factor productivity and the propagation of shocks; Empirical evidence and implications for the business cycle

2014-24 **Susanne Berger, Herbert Stocker, Achim Zeileis:** Innovation and institutional ownership revisited: An empirical investigation with count data models

2014-23 **James C. Cox, Rudolf Kerschbamer, Daniel Neururer:** What is trustworthiness and what drives it?

2014-22 **Michael Kirchler, Caroline Bonn, Jürgen Huber, Michael Razen:** The "Inflow-Effect" - Trader inflow and bubble formation in asset markets

2014-21 **Silvia Angerer, Daniela Glätzle-Rützler, Philipp Lergetporer, Matthias Sutter:** Donations, risk attitudes and time preferences: A study on altruism in primary school children *forthcoming in Journal of Economic Behavior and Organization*

2014-20 **Christian Kleiber, Achim Zeileis:** Visualizing count data regressions using rootograms

2014-19 **Matthias Siller, Christoph Hauser, Janette Walde, Gottfried Tappeiner:** The multiple facets of regional innovation

2014-18 **Carmen Arguedas, Esther Blanco:** On fraud and certification of corporate social responsibility

2014-17 **Achim Zeileis, Christoph Leitner, Kurt Hornik:** Home victory for Brazil in the 2014 FIFA World Cup

2014-16 **Andreas Exenberger, Andreas Pondorfer, Maik H. Wolters:** Estimating the impact of climate change on agricultural production: accounting for technology heterogeneity across countries

2014-15 **Alice Sanwald, Engelbert Theurl:** Atypical employment and health: A meta-analysis

2014-14 **Gary Charness, Francesco Feri, Miguel A. Meléndez-Jiménez, Matthias Sutter:** Experimental games on networks: Underpinnings of behavior and equilibrium selection *slightly revised version forthcoming in Econometrica*

2014-13 **Uwe Dulleck, Rudolf Kerschbamer, Alexander Konovalov:** Too much or too little? Price-discrimination in a market for credence goods

2014-12 **Alexander Razen, Wolgang Brunauer, Nadja Klein, Thomas Kneib, Stefan Lang, Nikolaus Umlauf:** Statistical risk analysis for real estate collateral valuation using Bayesian distributional and quantile regression

2014-11 **Dennis Dlugosch, Kristian Horn, Mei Wang:** Behavioral determinants of home bias - theory and experiment

2014-10 **Torsten Hothorn, Achim Zeileis:** partykit: A modular toolkit for recursive partytioning in R

2014-09 **Rudi Stracke, Wolfgang Höchtl, Rudolf Kerschbamer, Uwe Sunde:** Incentives and selection in promotion contests: Is it possible to kill two birds with one stone? *forthcoming in Managerial and Decision Economics*

2014-08 **Rudi Stracke, Wolfgang Höchtl, Rudolf Kerschbamer, Uwe Sunde:** Optimal prizes in dynamic elimination contests: Theory and experimental evidence *forthcoming in Journal of Economic Behavior and Organization*

2014-07 **Nikolaos Antonakakis, Max Breitenlechner, Johann Scharler:** How strongly are business cycles and financial cycles linked in the G7 countries?

2014-06 **Burkhard Raunig, Johann Scharler, Friedrich Sindermann:** Do banks lend less in uncertain times?

2014-05 **Julia Auckenthaler, Alexander Kupfer, Rupert Sendlhofer:** The impact of liquidity on inflation-linked bonds: A hypothetical indexed bonds approach

2014-04 **Alice Sanwald, Engelbert Theurl:** What drives out-of pocket health expenditures of private households? - Empirical evidence from the Austrian household budget survey

2014-03 **Tanja Hörtnagl, Rudolf Kerschbamer:** How the value of information shapes the value of commitment or: Why the value of commitment does not vanish

2014-02 **Adrian Beck, Rudolf Kerschbamer, Jianying Qiu, Matthias Sutter:** Car mechanics in the lab - Investigating the behavior of real experts on experimental markets for credence goods *forthcoming in Journal of Economic Behavior and Organization*

2014-01 **Loukas Balafoutas, Adrian Beck, Rudolf Kerschbamer, Matthias Sutter:** The hidden costs of tax evasion - Collaborative tax evasion in markets for expert services

University of Innsbruck

Working Papers in Economics and Statistics

2015-02

Helene Roth, Stefan Lang, Helga Wagner

Random intercept selection in structured additive regression models

Abstract
This paper discusses random intercept selection within the context of semiparametric regression models with structured additive predictor (STAR). STAR models can deal simultaneously with nonlinear covariate effects and time trends, unit- or cluster-specific heterogeneity, spatial heterogeneity and complex interactions between covariates of different type. The random intercept selection is based on spike and slab priors for the variances of the random intercept coefficients. The aim is to achieve shrinkage of small random intercept coefficients to zero similar as for the LASSO in frequentist linear models. The mixture structure of the spike and slab prior allows for selective shrinkage, as coefficients are either heavily shrunk under the spike component or left almost unshrunk under the slab component. The hyperparameters of the spike and slab prior are chosen by theoretical considerations based on the prior inclusion probability of a particular random coefficient given the true effect size. Using extensive simulation experiments we compare random intercept models based on spike and slab priors for variances with the usual Inverse Gamma priors. A case study on malnutrition of children in Zambia illustrates the methodology in a real data example.

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