Estimation of spatially correlated random scaling factors based on Markov random field priors

Alexander Razen, Stefan Lang, Judith Santer

Working Papers in Economics and Statistics

2016-33

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.

Estimation of Spatially Correlated Random Scaling Factors based on Markov Random Field Priors

Alexander Razen
University of Innsbruck

Stefan Lang
University of Innsbruck

Judith Santer
University of Innsbruck

Abstract

Multiplicative random effects allow for cluster-specific scaling of covariate effects. In many applications with spatial clustering, however, the random effects additionally show some geographical pattern, which usually can not sufficiently be captured with existing estimation techniques. Relying on Markov random fields, we present a fully Bayesian inference procedure for spatially correlated scaling factors. The estimation is based on highly efficient Markov Chain Monte Carlo (MCMC) algorithms and is smoothly incorporated into the framework of distributional regression.

We run a comprehensive simulation study for different response distributions to examine the statistical properties of our approach. We also compare our results to those of a general estimation procedure for independent random scaling factors. Furthermore, we apply the method to German real estate data and show that exploiting the spatial correlation of the scaling factors further improves the performance of the model.

Keywords: distributional regression, iteratively weighted least squares proposals, MCMC, multiplicative random effects, spatial smoothing, structured additive predictors

This work was supported by funds of the Oesterreichische Nationalbank (Oesterreichische Nationalbank, Anniversary Fund, project number: 15309).

1 Introduction

Distributional regression as introduced by Klein et al. (2015) assumes a parametric distribution for a response variable y and models some or all of its parameters in dependence of covariates. By linking the parameters to structured additive predictors the framework simultaneously can capture nonlinear covariate effects and time trends, unit- or cluster-specific heterogeneity, spatial heterogeneity and complex interactions between covariates of different type.

In Razen and Lang (2016), we embedded the concept of multiplicative random effects into this framework, which allows for cluster-specific scaling of the covariates' effects. The idea is applicable in various fields, see e.g. Brunauer et al. (2010) or Weber et al. (2016), and can considerably improve the predictive ability of a model.

In applications with spatial clustering, the scaling factors are often correlated. For example, when analyzing German house price data in Razen and Lang (2016), considerable differences in the scaling of the covariate effects between Eastern and Western Germany have been found. The current estimation procedure, however, does not incorporate such correlations ex ante. Therefore, we usually do not get smooth spatial effects, although in general some kind of smoothness seems to be reasonable. On the other hand, ignoring spatial correlation may cause problems in getting reliable estimation results for clusters where only a few observations are available.

A widely used concept for modeling spatial correlation is given by Markov random fields, see e.g. Besag et al. (1991), Fahrmeir and Lang (2001) or Rue and Held (2005). Here, based on some definition of neighborhood, the geographical vicinity of the clusters are taken into account when estimating a spatial effect, smoothing the respective results.

The aim of this paper is to apply Markov random field priors to multiplicative random effects within the framework of distributional regression in order to model spatially correlated random scaling factors. The proposed estimation procedure is fully Bayesian and relies on highly efficient Markov Chain Monte Carlo (MCMC) algorithms.

We run extensive simulation experiments for different response distributions and evaluate the performance of our method. Furthermore, we illustrate the benefits of this approach by applying it to a German real estate dataset and compare the results to those of our previous study in Razen and Lang (2016).

The paper is structured as follows: The methodology is introduced in Section 2. Section 3 attends to the simulation experiments before we present the application to the real estate data in Section 4. Finally, we conclude in Section 5.

2 Methodology

2.1 Correlated scaling in distributional regression

Suppose we are given data on n observations in the form $(y_i, \boldsymbol{z}_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$, with response y and a number of covariates $\boldsymbol{z}$ and $\boldsymbol{x}$. Assume further that the data is grouped into C clusters. Then, Bayesian distributional regression assumes an L-parametric distribution of the response y, given the covariates, and links its parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L$ to structured additive predictors $\boldsymbol{\eta}_l$ via known response functions h_l,

$$\boldsymbol{\theta}_l = h_l(\boldsymbol{\eta}_l),$$

$l = 1, \ldots, L$, see Klein et al. (2015), for details. Incorporating cluster-specific random scaling factors, the predictors are given by

$$\boldsymbol{\eta}_l = \boldsymbol{D}_{1l}\boldsymbol{f}_{1l}(\boldsymbol{z}_{1l}) + \ldots + \boldsymbol{D}_{q_l l}\boldsymbol{f}_{q_l l}(\boldsymbol{z}_{q_l l}) + \boldsymbol{X}_l\boldsymbol{\gamma}_l, \tag{1}$$

where the functions $\boldsymbol{f}_{jl}$ are possibly nonlinear functions of the covariates $\boldsymbol{z}_{jl}$, $\boldsymbol{D}_{jl} = diag(1 + \alpha_{jc_1 l}, \ldots, 1 + \alpha_{jc_n l})$, $c_i \in 1, \ldots, C$ are $n \times n$ diagonal matrices including random scaling factors for the C clusters in the main diagonal and the term $\boldsymbol{X}_l\boldsymbol{\gamma}_l$ comprises the linear effects of the model. For the sake of simplicity, we will suppress the index discriminating between the L parameters in the following whenever possible.

Using known basis functions B_k, a particular function f can be approximated by

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

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)'$ is a vector of unknown regression coefficients to be estimated. A standard choice for continuous covariates are B-spline basis functions, see below.

Defining the $n \times K$ design matrix $\boldsymbol{Z}$ with elements $\boldsymbol{Z}[i,k] = B_k(z_i)$, the vector $\boldsymbol{f} = (f(z_1), \ldots, f(z_n))'$ of function evaluations can be written in matrix notation as $\boldsymbol{f} = \boldsymbol{Z}\boldsymbol{\beta}$. Accordingly, the predictors in (1) can be written as

$$\boldsymbol{\eta} = \boldsymbol{D}_1\boldsymbol{Z}_1\boldsymbol{\beta}_1 + \ldots + \boldsymbol{D}_q\boldsymbol{Z}_q\boldsymbol{\beta}_q + \boldsymbol{X}\boldsymbol{\gamma}. \tag{2}$$

In a Bayesian framework, overfitting of a particular function $\boldsymbol{f}$ usually is avoided by employing a suitable smoothness prior for the regression coefficients $\boldsymbol{\beta}$, see e.g. Fahrmeir et al. (2013). A standard choice is a (possibly improper) Gaussian prior of the form

$$p(\boldsymbol{\beta}|\tau^2) \propto \left(\frac{1}{\tau^2}\right)^{\mathrm{rk}(\boldsymbol{K})/2} \exp\left(-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{K}\boldsymbol{\beta}\right) \cdot I(\boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{0}), \tag{3}$$

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. $\mathrm{rk}(\boldsymbol{K}) < K$, resulting in a partially improper prior. The specific structure of $\boldsymbol{K}$ depends on the covariate type and on prior assumptions about the smoothness of $\boldsymbol{f}$.

Priors for continuous covariate effects

We apply, for example, a Bayesian version of P-splines when modeling a smooth function $\boldsymbol{f}$ that depends on a continuous covariate $\boldsymbol{z}$, see Eilers and Marx (1996), Lang and Brezger (2004) and Eilers et al. (2015). Here, the columns of the design matrix $\boldsymbol{Z}$ are given by B-spline basis functions evaluated at the observations z_i and we use first or second order random walks as smoothness priors for the regression coefficients, i.e. $\beta_k = \beta_{k-1} + u_k$, or $\beta_k = 2\beta_{k-1} - \beta_{k-2} + u_k$, with Gaussian errors $u_k \sim \mathcal{N}(0, \tau^2)$ 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 (3) with the penalty matrix given by $\boldsymbol{K} = \boldsymbol{\mathcal{D}}'\boldsymbol{\mathcal{D}}$, where $\boldsymbol{\mathcal{D}}$ is a first or second order difference matrix.

The amount of smoothness is governed by the variance parameter τ^2. A conjugate inverse Gamma prior is employed for τ^2, i.e. $\tau^2 \sim IG(a, b)$ with small values for the hyperparameters a and b resulting in an uninformative prior on the log scale. As a default we choose $a = b = 0.001$.

The term $I(\boldsymbol{A\beta} = \boldsymbol{0})$ 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.

Priors for spatially correlated random effects

In a general setting, the random effects α_c in the scaling matrix $\boldsymbol{D} = diag(1 + \alpha_{c_1}, \ldots, 1 + \alpha_{c_n})$ of a particular function $\boldsymbol{f}$ are supposed to be independent and normally distributed with mean 0 and variance $\tilde{\tau}^2$. If we assume them to be spatially correlated, however, we instead propose the use of a Markov random field prior:

$$\alpha_c \,|\, \alpha_r, \tilde{\tau}^2, r \in N(c) \sim \mathcal{N}\left(\frac{1}{|N(c)|}\sum_{r \in N(c)} \alpha_r \,,\, \frac{\tilde{\tau}^2}{|N(c)|}\right),$$

where $|N(c)|$ is the number of the neighbors $N(c)$ of region c. The joint distribution of all random effects $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_C)'$ then is given by

$$p(\boldsymbol{\alpha} \,|\, \tilde{\tau}^2) \propto \left(\frac{1}{\tilde{\tau}^2}\right)^{(C-1)/2} \exp\left(-\frac{1}{2\tilde{\tau}^2}\boldsymbol{\alpha}'\tilde{\boldsymbol{K}}\boldsymbol{\alpha}\right),$$

with the precision matrix $\tilde{\boldsymbol{K}}$ being defined by

$$\tilde{\boldsymbol{K}}[c, r] = \begin{cases} -1 & \text{if } c \neq r \text{ and } r \in N(c), \\ 0 & \text{if } c \neq r \text{ and } r \notin N(c), \\ |N(c)| & \text{if } c = r. \end{cases}$$

For the variance parameter $\tilde{\tau}^2$ we assign the usual inverse Gamma prior $\tilde{\tau}^2 \sim IG(\tilde{a}, \tilde{b})$. Furthermore, we center the random effects around zero, i.e.

$$\sum_{c=1}^{C} \alpha_c = 0.$$

Thus, the prior for the random effects again is a smoothing prior of the form (3).

2.2 Inference

For the sake of illustration, we consider a Gaussian model with a single predictor for the mean parameter of the form (2):

$$\boldsymbol{y} = \boldsymbol{D}_1\boldsymbol{Z}_1\boldsymbol{\beta}_1 + \ldots + \boldsymbol{D}_q\boldsymbol{Z}_q\boldsymbol{\beta}_q + \boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}. \tag{4}$$

Rewriting the model in terms of the random effects $\boldsymbol{\alpha}$ yields

$$\boldsymbol{y} = \boldsymbol{f}_1 + \tilde{\boldsymbol{D}}_1\tilde{\boldsymbol{Z}}_1\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{f}_q + \tilde{\boldsymbol{D}}_q\tilde{\boldsymbol{Z}}_q\boldsymbol{\alpha}_q + \boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{5}$$

with $\boldsymbol{f}_j = (f(z_{j1}), \ldots, f(z_{jn}))$, $\tilde{\boldsymbol{D}}_j = diag(f(z_{j1}), \ldots, f(z_{jn}))$ and $\tilde{\boldsymbol{Z}}_j$ being a $n \times C$ matrix indicating if observation i belongs to cluster c (in this case $\tilde{\boldsymbol{Z}}_j(i, c) = 1$, otherwise it equals 0).

Following Razen and Lang (2016), we can interpret equations (4) and (5) as varying coefficient models and alternately employ Gibbs updates for the regression coefficients and the random effects.

The full conditionals of the regression parameters $\boldsymbol{\beta}_j$ are derived from (4) and are multivariate Gaussian $\boldsymbol{\beta}_j \,|\, \cdot \sim \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ with

$$\boldsymbol{\Sigma}_j^{-1} = \frac{1}{\sigma^2}\left(\boldsymbol{Z}_j'\boldsymbol{D}_j^2\boldsymbol{Z}_j + \frac{\sigma^2}{\tau_j^2}\boldsymbol{K}_j\right), \qquad \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j = \frac{1}{\sigma^2}\boldsymbol{Z}_j'\boldsymbol{D}_j(\boldsymbol{y} - \boldsymbol{\eta}_j),$$

where $\boldsymbol{\eta}_j$ contains the current predictor except the j-th term.

The full conditionals of the random effects $\boldsymbol{\alpha}_j$ are derived from (5), now with $\boldsymbol{\alpha}_j \,|\, \cdot \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_j, \tilde{\boldsymbol{\Sigma}}_j)$ and

$$\tilde{\boldsymbol{\Sigma}}_j^{-1} = \frac{1}{\sigma^2}\left(\tilde{\boldsymbol{Z}}_j'\tilde{\boldsymbol{D}}_j^2\tilde{\boldsymbol{Z}}_j + \frac{\sigma^2}{\tilde{\tau}_j^2}\tilde{\boldsymbol{K}}_j\right), \qquad \tilde{\boldsymbol{\Sigma}}_j^{-1}\tilde{\boldsymbol{\mu}}_j = \frac{1}{\sigma^2}\tilde{\boldsymbol{Z}}_j'\tilde{\boldsymbol{D}}_j(\boldsymbol{y} - \boldsymbol{f}_j - \boldsymbol{\eta}_j).$$

The full conditionals of the variance parameters are inverse Gamma and are given by

$$\begin{aligned}
\tau_j^2 &\sim IG(a', b'), \quad a' = a + 0.5\,\mathrm{rk}(\boldsymbol{K}_j), \quad b' = b + 0.5\,\boldsymbol{\beta}_j'\boldsymbol{K}_j\boldsymbol{\beta}_j, \\
\tilde{\tau}_j^2 &\sim IG(\tilde{a}', \tilde{b}'), \quad \tilde{a}' = \tilde{a} + 0.5\,(C-1), \quad \tilde{b}' = \tilde{b} + 0.5\,\boldsymbol{\alpha}_j'\tilde{\boldsymbol{K}}_j\boldsymbol{\alpha}_j.
\end{aligned}$$

For those distributional regression models where the full conditionals of $\boldsymbol{\beta}_j$ and $\boldsymbol{\alpha}_j$ are no longer Gaussian, we employ Metropolis Hastings updates instead of Gibbs updates, see Klein et al. (2015) or Klein et al. (2014) for details.

3 Simulation

3.1 Simulation design

In this section we present an extensive simulation study including models with Gaussian, gamma or binomial responses. Each parameter of the respective distribution is linked to a separate predictor via a suitable link function. The precise settings are summarized in Table 1.

Response distribution	Parameter	Predictor	Link function
Gaussian	μ	η_1	identity function
	σ	η_2	exponential function
Gamma	μ	η_1	exponential function
	σ	η_2	exponential function
Binomial	p	η_1	logistic function

Table 1: *Model settings*

For $l = 1, 2$ the parameters are constructed as follows:

$$\eta_l = (1 + \alpha_{cl})\, f_l\,(z)\,,$$

where $f_l(z) = \sin(z)$ in the interval $[-\pi, \pi]$ and $(1 + \alpha_{cl})$ are spatially correlated random scaling factors for the districts of Western Germany. The random effects α_{cl} have mean 0 and variance $\tilde{\tau}_l^2$ and are defined by the normalized centroid coordinates x_c and y_c of the individual districts c as follows:

$$\begin{aligned} \alpha_{c1} &= 0.4 \cdot (x_c + y_c), \\ \alpha_{c2} &= 2.5 \cdot x_c + 0.5 \cdot y_c. \end{aligned}$$

We additionally scale them in three different ways, so that their variance $\tilde{\tau}_l^2$ is either 0.1^2, 0.5^2 or 1.0^2, in order to evaluate the influence of $\tilde{\tau}_l^2$. Furthermore, we vary the number of observations per district. In total, for each response distribution we analyze five different models, whose specifications are summarized in Table 2.

Model	Obs. per district	Variance of α_{cl}
Model 1	10	0.5^2
Model 2	25	0.5^2
Model 3	50	0.5^2
Model 4	25	0.1^2
Model 5	25	1.0^2

Table 2: *Model specifications*

For illustration, the geographical maps of the random effects α_{cl} of Model 1 are depicted in Figure 1, the corresponding effects $(1 + \alpha_{cl}) f_l(z)$ are shown in Figure 2.

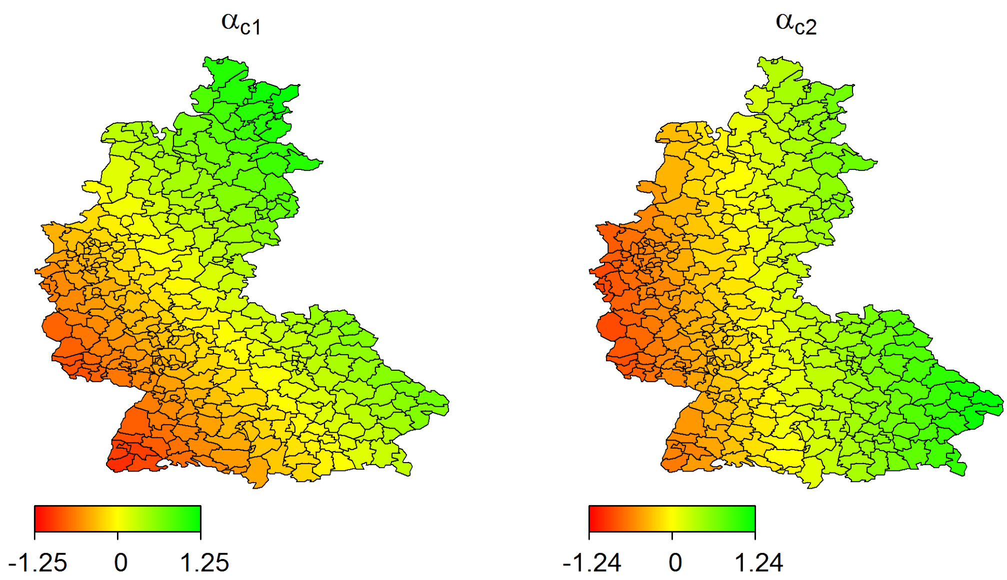

Figure 1: Maps of the random effects α_{cl} with variance $\tilde{\tau}_l^2 = 0.5^2$.

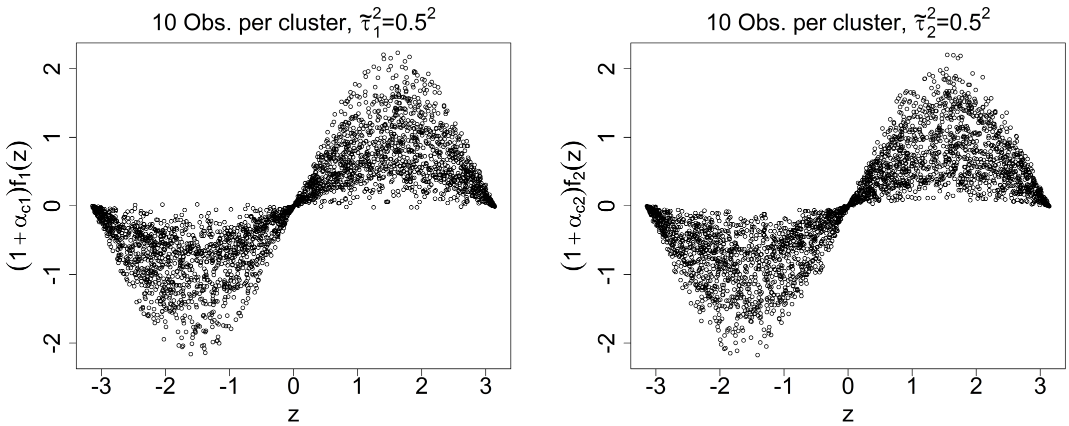

Figure 2: The functions $f_l(z)$, multiplied with the respective random scaling factors.

3.2 Results

For each model, we generate 250 replications and carry out the estimation procedure described in Section 2 based on a final MCMC run with $120,000$ iterations and a burn in period of $20,000$ iterations. We store every 100th iteration in order to obtain a sample of $1,000$ draws from the posterior. For illustration, Figure 3 depicts the sampling paths of the random effects α_{cl} of the district of Munich in one of the replications of the Gaussian Model 1, the corresponding autocorrelation functions are shown in Figure 4. As we can see, the draws are practically independent, indicating a good mixing.

We then calculate the arithmetic mean from the 250 replications. In the following, we present the respective results for each response distribution separately.

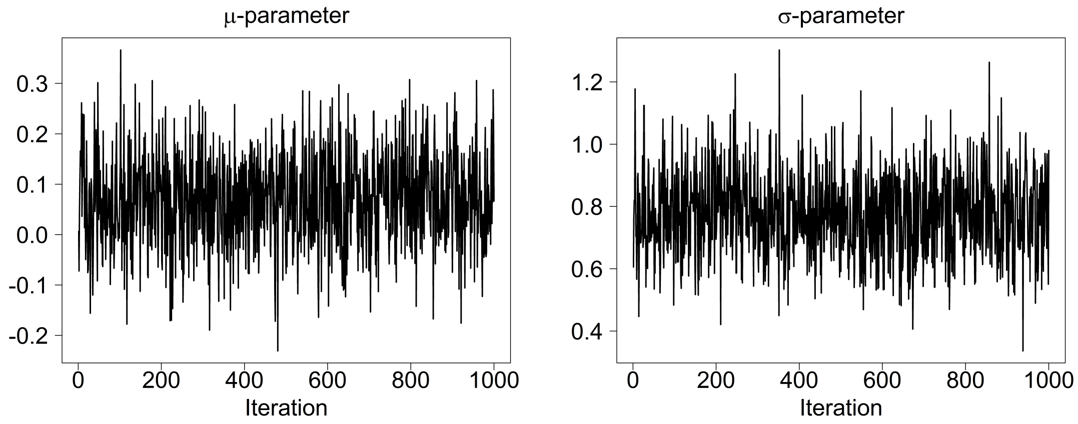

Figure 3: Sampling paths of the random effects α_{cl} of the district of Munich in one of the replications of the Gaussian Model 1.

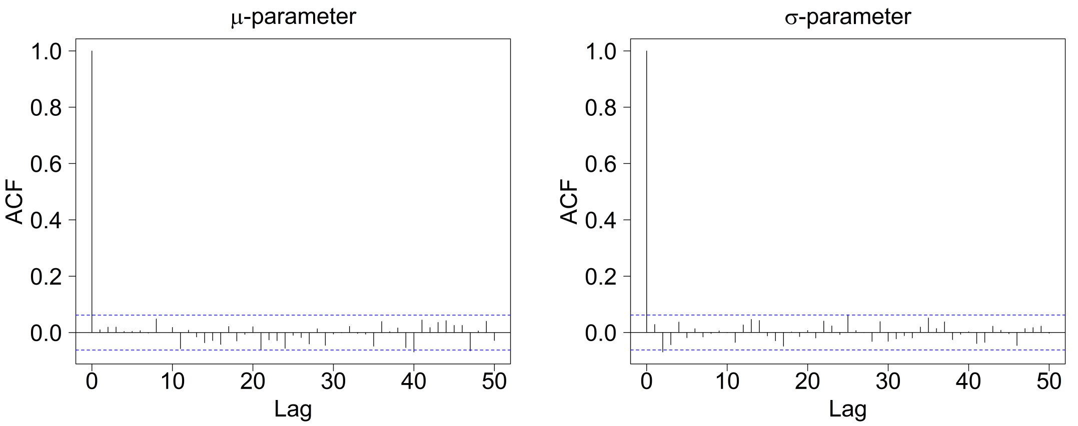

Figure 4: Autocorrelation functions of the random effects α_{cl} of the district of Munich in one of the replications of the Gaussian Model 1.

Gaussian Models

Figures 5 and 6 show the average estimates of the effects $f_l(z)$ as well as of the district-specific effects $(1 + \alpha_{cl})\, f_l(z)$ for the smallest and largest random effects α_{cl} (solid) of the Gaussian models. The true effects also are plotted (dashed) in order to facilitate comparison. The first column refers to the μ-parameter, the second column to the σ-parameter.

As we can see from Figure 5, the scaled effects almost perfectly can be estimated for both parameters, even with just 10 observations per district. With respect to the variance of the random effects, Figure 6 shows that we get good estimation results even for a small variance.

For illustration, Figure 7 shows the geographical maps of the estimated random effects $\widehat{\alpha_{cl}}$ of the Gaussian Model 4 with 25 observations per district and the smallest variance of the random effects $\tilde{\tau}_l^2 = 0.1^2$ (left column). The first row refers to the μ-parameter, the second row to the σ-parameter. For comparison, the true effects α_{cl} are plotted in the middle column. We only detect minor differences so that the biases, defined as $\widehat{\alpha_{cl}} - \alpha_{cl}$ and depicted in the right column, are rather small. This reflects the great performance of our method in identifying correlated random effects, even if their variance is small.

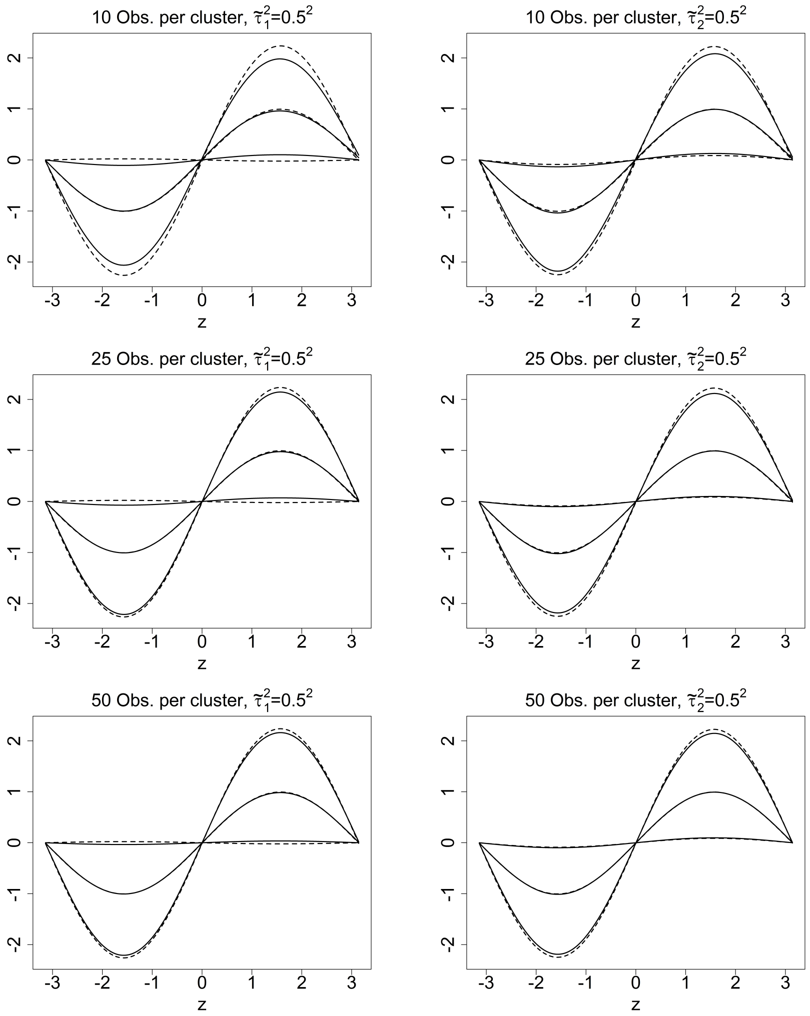

Figure 5: The average estimates of the functions $f_l(z)$ as well as of the smallest and largest cluster-specific effects $(1 + \alpha_{cl})\, f_l(z)$ (solid) and the respective true effects (dashed) for the Gaussian Models 1 to 3. The first column shows the effects of the μ-parameter, the second column those of the σ-parameter.

Gamma Models

The main findings for the Gamma models are very similar to those of the Gaussian models. Even with only 10 observations per district the estimation results for both parameters are still very good, see Figure 8. If we increase the number of observations, then we are again able to estimate the scaled effects almost perfectly. Furthermore, as we can see from Figure 9, we also get reasonable results even for a small variance of the random effects.

The geographical maps of the estimated random effects $\widehat{\alpha_{cl}}$ of Model 4 shown in Figure 10 confirm this finding. Despite the low variance of the random effects in this model, their spatial structure can be identified very well, even though the effects themselves are somewhat underestimated in their magnitude, leading to slightly larger biases.

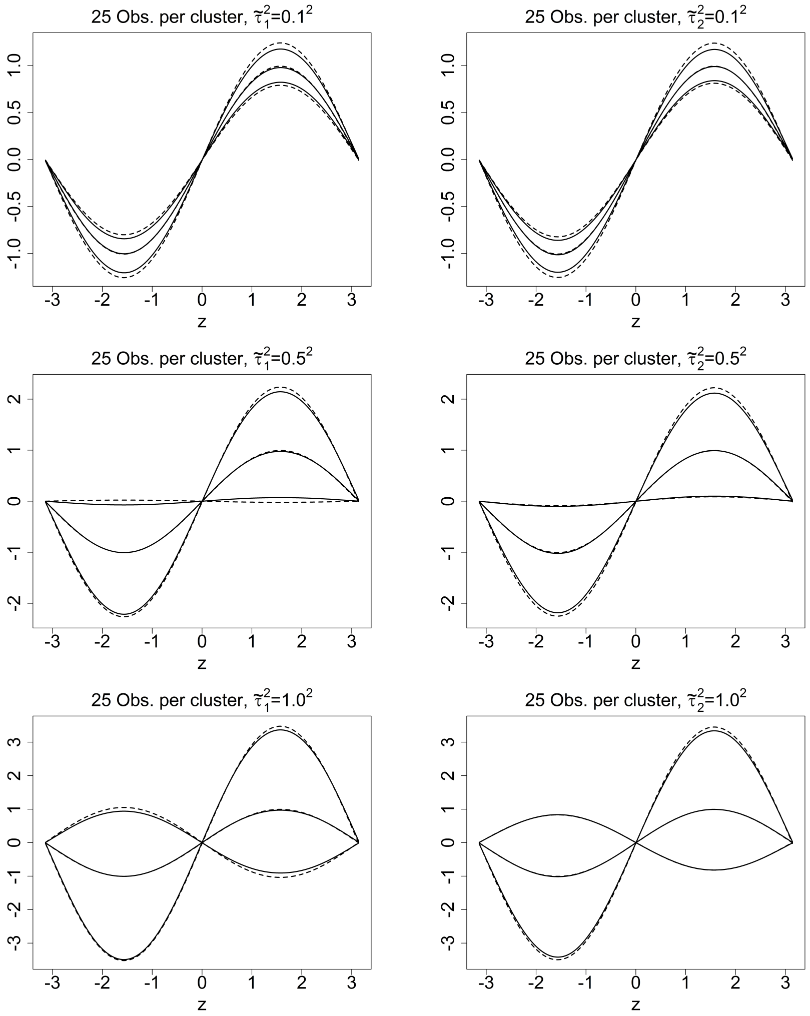

Figure 6: The average estimates of the functions $f_l(z)$ as well as of the smallest and largest district-specific effects $(1 + \alpha_{cl})\, f_l(z)$ (solid) and the respective true effects (dashed) for the Gaussian Models 4, 2 and 5. The first column shows the effects of the μ-parameter, the second column those of the σ-parameter.

Binomial Models

Even in Binomial models we already get reasonable estimation results for the scaled effects with just 10 observations per district, see Figure 11. Again, the results further improve if we increase the number of observations. With respect to the variance of the random effects, we underestimate the scaling of the effects for a small variance but get appropriate results if the variance is higher (Figure 12).

The first row of Figure 13 reflects the difficulty in estimating the magnitude of the random effects properly if their variance is too small, even though the spatial structure is already recognizable. For larger variances (e.g. $\tilde{\tau}_l^2 = 0.5^2$), the results considerably improve and we only detect minor biases between the estimated and the true random effects (second row of Figure 13).

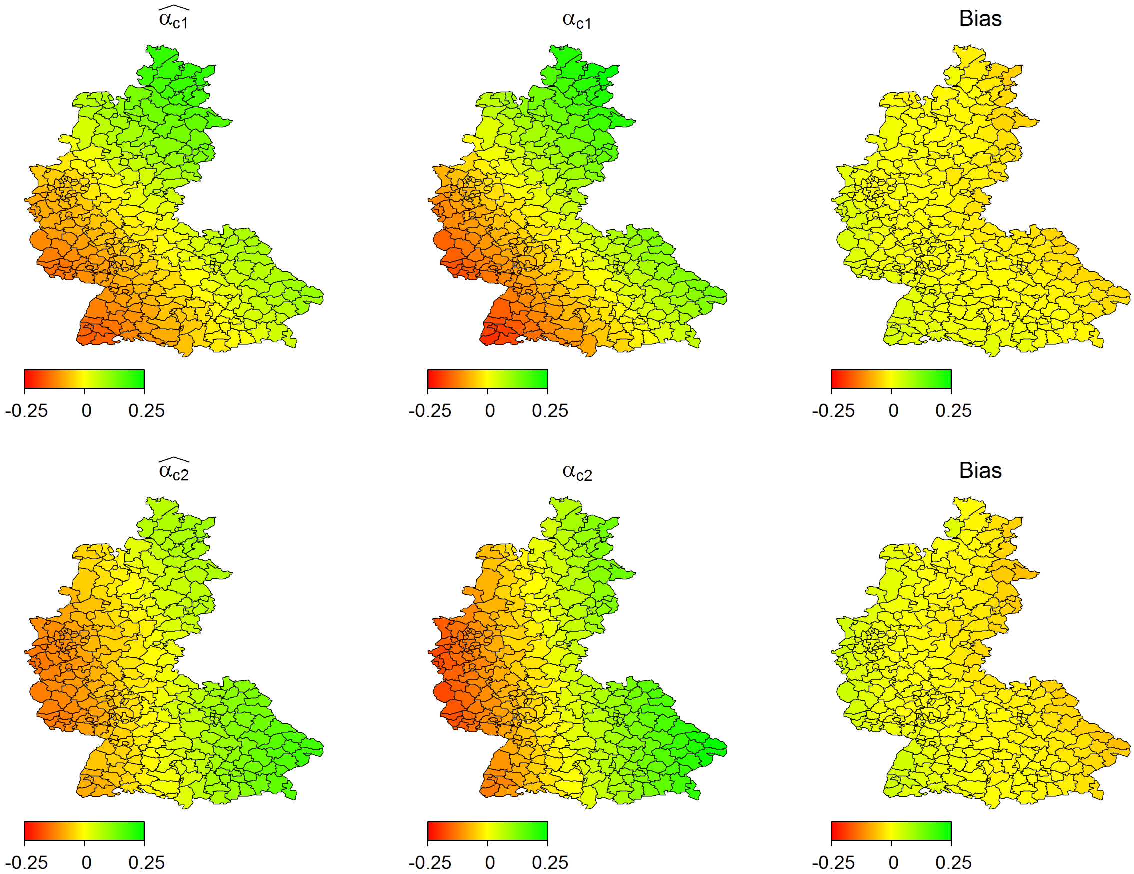

Figure 7: Maps of the random effects α_{cl} of the Gaussian Model 4. The first column shows the estimated effects, the second column shows the true effects, the third column shows the biases.

The results of the Binomial models perfectly illustrate the advantage of our method for estimating correlated scaling factors based on Markov random field priors, which we will call *Correlated Method*. If we would neglect their spatial correlation and instead apply the estimation procedure for independent random scaling factors described in Razen and Lang (2016), referred to as *Uncorrelated Method*, then we would hardly be able to identify any random effects in the Binomial Models 1 or 4, i.e. in models where we only have a small number of observations per district or a low variance of the random effects. Even in Model 3 with 50 observations per district and a moderate variance of the random effects, the estimates of the Uncorrelated Method still would be much worse. Figure 14 compares the random effects estimated with and without Markov random field priors for the Models 1 (first row), 4 (second row) and 3 (third row), respectively.

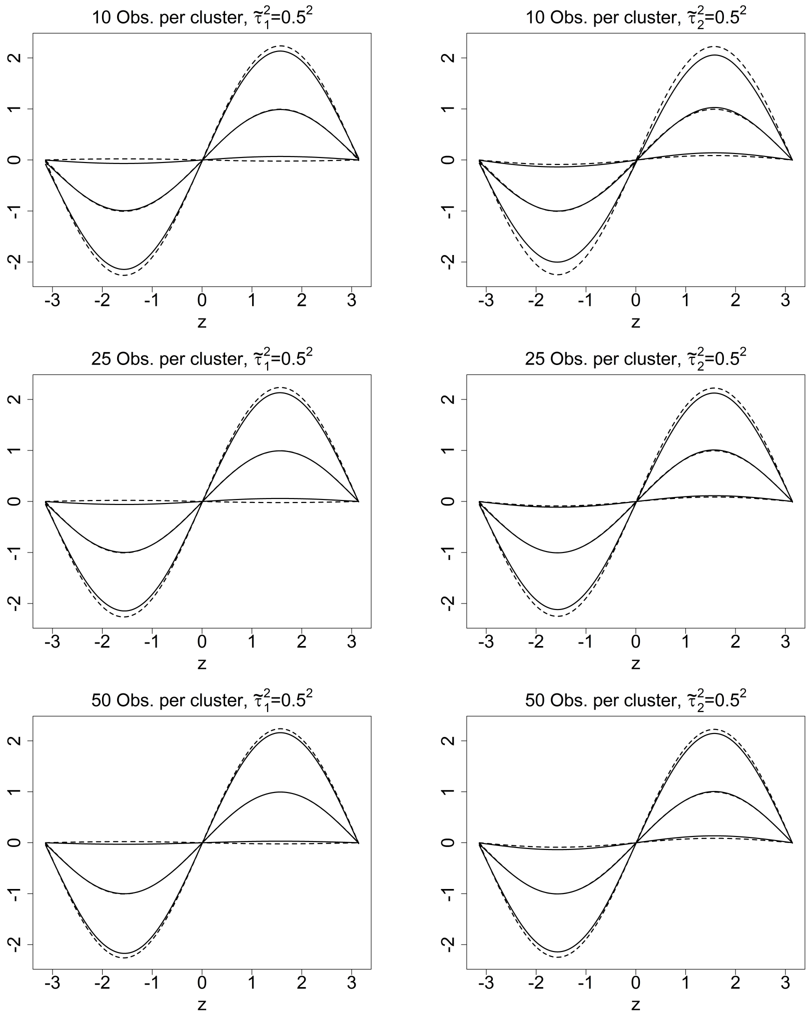

Figure 8: The average estimates of the functions $f_l(z)$ as well as of the smallest and largest district-specific effects $(1 + \alpha_{cl})\, f_l(z)$ (solid) and the respective true effects (dashed) for the Gamma Models 1 to 3. The first column shows the effects of the μ-parameter, the second column those of the σ-parameter.

3.3 Model evaluation

We evaluate the performance of our approach, the Correlated Method, using mean squared errors (MSEs). In each replication of our models, we calculate the MSE for the l different parameters as follows

$$\text{MSE}_l = \frac{1}{n} \left(\hat{\boldsymbol{\eta}}_l - \boldsymbol{\eta}_l\right)' \left(\hat{\boldsymbol{\eta}}_l - \boldsymbol{\eta}_l\right).$$

We then reestimate all replications using the estimation approach for independent random scaling factors described in Razen and Lang (2016), the Uncorrelated Method, and again calculate the corresponding MSEs.

The first row of Figure 15 depicts the boxplots of the MSEs of the 250 replications for the Gaussian Models 1-3 with 10, 25 and 50 observations per district and a variance of the random effects $\tilde{\tau}_l^2 = 0.5^2$. The second row shows the corresponding boxplots for the Gaussian Models 4, 2 and 5 with variances $\tilde{\tau}_l^2 = 0.1^2$, $\tilde{\tau}_l^2 = 0.5^2$ and $\tilde{\tau}_l^2 = 1.0^2$, each with 25 observations per cluster. We can see that our estimation approach for correlated random

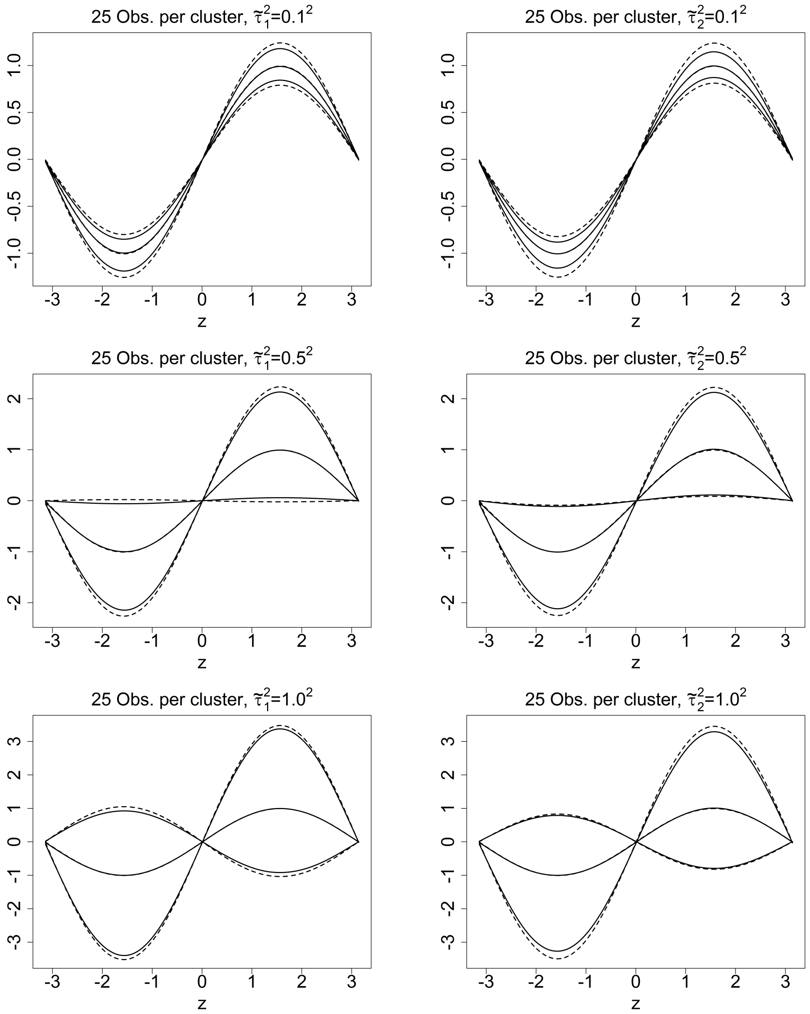

Figure 9: The average estimates of the functions $f_l(z)$ as well as of the smallest and largest district-specific effects $(1 + \alpha_{cl})\, f_l(z)$ (solid) and the respective true effects (dashed) for the Gamma Models 4, 2 and 5. The first column shows the effects of the μ-parameter, the second column those of the σ-parameter.

scaling factors based on Markov random field priors consistently reduces the MSEs. With respect to the number of observations per district, we find that the Correlated Method is even better with just 10 observations per district than the Uncorrelated Method with 50 observations per district. Regarding the variance of the random effects, we see only a moderate increase in the MSEs for higher variances when using the Correlated Method, whereas the increase is much larger when using the Uncorrelated Method.

Our results are consistent over the different response distributions. For the Gamma models (Figure 16) and the Binomial models (Figure 17) we are also able to considerably reduce the MSEs for all model specifications, confirming the better results of our new estimation approach that we have seen in Figure 14. The largest improvements with respect to the MSEs again can be achieved in settings with either a low number of observations per district or a high variance of the random effects.

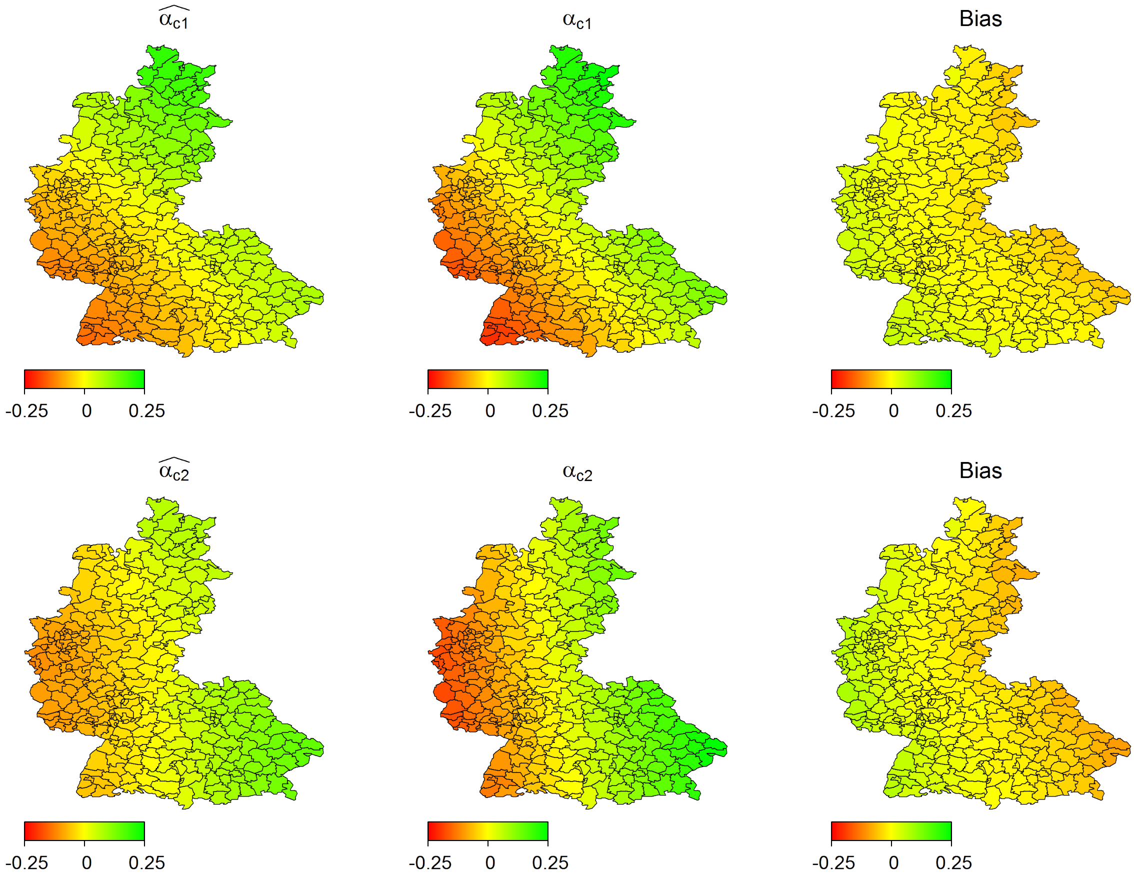

Figure 10: Maps of the random effects α_{cl} of the Gamma Model 4. The first column shows the estimated effects, the second column shows the true effects, the third column shows the biases.

4 Application

4.1 Model specification

We demonstrate the benefits of our method in an application to real estate valuation. Our dataset consists of almost $100,000$ single family homes all over Germany and includes information about the buildings and their location. The buildings are characterized by their floor area (*area*), the plot area where they are built on (*plot_area*), the year of construction (*year*) and their equipment, which is classified by four categories. Regarding the location, we know the districts and the states where the houses are located in and have an expert rating (*rating*) available that evaluates the respective surrounding. Of course, we are also given the house price per square meter (p_{qm}), which will be the dependent variable in our analysis. A more detailed description of the data can be found in Razen and Lang (2016).

Based on our findings in Razen and Lang (2016), we set up a Gamma model for the house price per square meter with mean parameter μ and shape parameter σ, which we link to predictors η_1 and η_2 via

$$\mu = \exp(\eta_1),$$
$$\sigma = \exp(\eta_2).$$

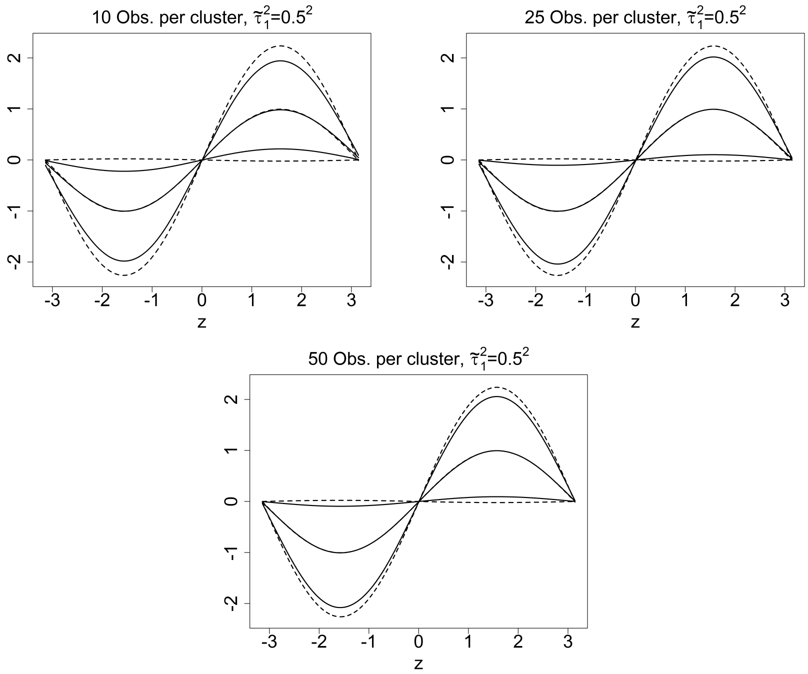

Figure 11: The average estimates of the functions $f_1(z)$ as well as of the smallest and largest cluster-specific effects $(1 + \alpha_{c1})\, f_1(z)$ (solid) and the respective true effects (dashed) for the Binomial Models 1 to 3.

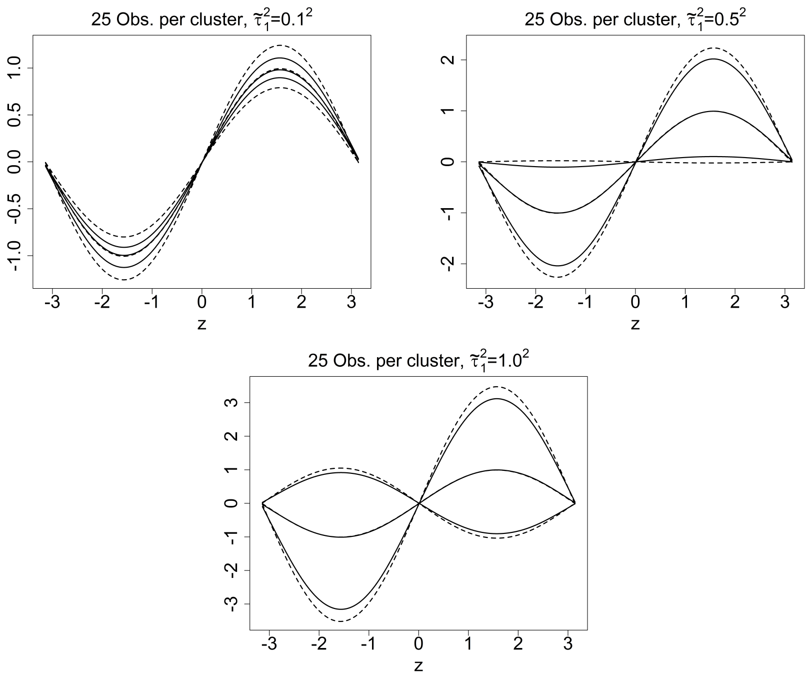

Figure 12: The average estimates of the functions $f_1(z)$ as well as of the smallest and largest cluster-specific effects $(1 + \alpha_{c1})\, f_1(z)$ (solid) and the respective true effects (dashed) for the Binomial Models 4, 2 and 5.

For $l = 1, 2$, the predictors are constructed by

$$\begin{aligned}\boldsymbol{\eta_l} &= \boldsymbol{D}_{1l}\boldsymbol{f}_{1l}(area) + \boldsymbol{D}_{2l}\boldsymbol{f}_{2l}(plot\_area) \\ &\quad + \boldsymbol{D}_{3l}\boldsymbol{f}_{3l}(year) + \boldsymbol{f}_{4l}(rating) + \boldsymbol{X}\boldsymbol{\gamma}_l,\end{aligned}$$

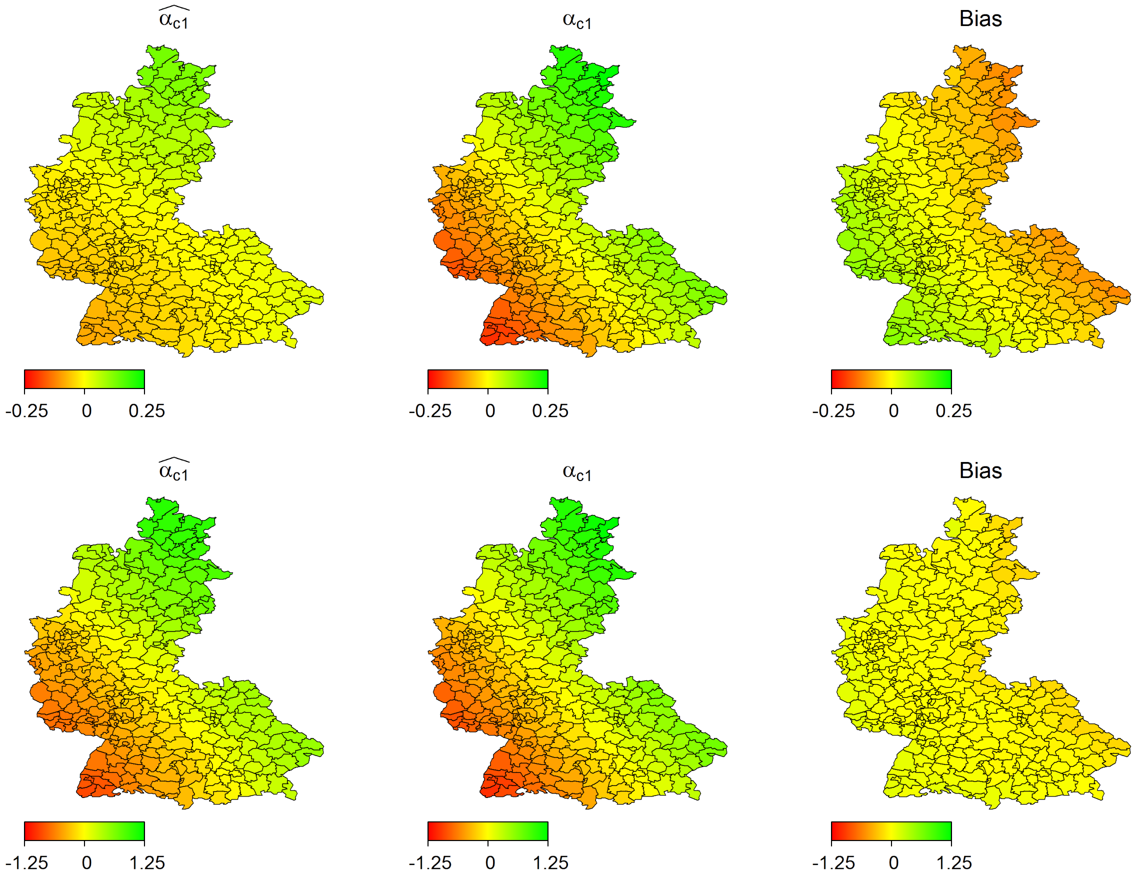

Figure 13: Maps of the random effects α_{c1} of the Binomial Models 4 (first row) and 2 (second row). The first column shows the estimated effects, the second column shows the true effects, the third column shows the biases.

where $\boldsymbol{f}_{1l}, \ldots, \boldsymbol{f}_{4l}$ are possibly nonlinear functions of the respective continuous covariates and will be modeled with P-splines. $\boldsymbol{D}_{jl} = \operatorname{diag}(1 + \alpha_{jc_1l}, \ldots, 1 + \alpha_{jc_nl})$, $c_i \in \{1, \ldots, C\}$, contain random scaling factors for the C districts that allow for regional heterogeneity in the respective price response functions and will be modeled with Markov random fields. The intercept as well as the dummy variables for the equipment of the houses and the states where the buildings are located in are subsumed in the design matrix $\boldsymbol{X}$ with parameters $\boldsymbol{\gamma}_l$.

4.2 Results

The estimation results are based on a final MCMC run with $270,000$ iterations and a burn in period of $20,000$ iterations. We stored every 250th iteration, leading to a sample of $1,000$ practically independent draws from the posterior.

For the sake of illustration, we compare the results of the estimation procedure proposed in Section 2 (Correlated Method) to those of the estimation procedure for independent random scaling factors (Uncorrelated Method) described in Razen and Lang (2016). As we can see from Figure 18, the mean effects of the covariates, averaged over all districts, almost coincide for the two methods. Here, the other continuous covariates are always held constant at mean level of attributes and the categorical variables are held at their mode level.

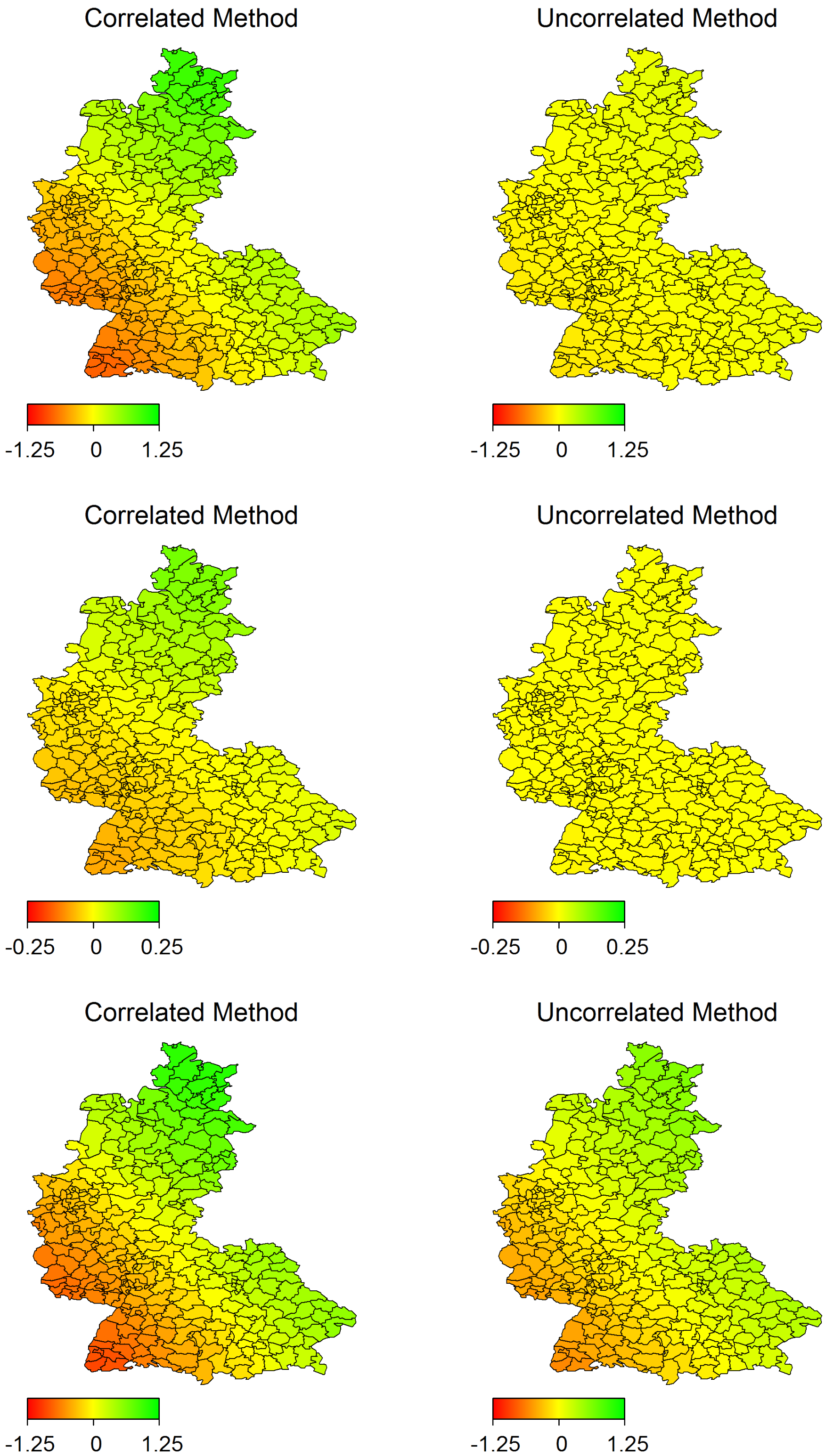

Figure 14: Maps of the random effects α_{c1} of the Binomial Model 1 with 10 observations per district and $\tilde{\tau}_l^2 = 0.5^2$ (first row), of Model 4 with 25 observations per cluster and $\tilde{\tau}_l^2 = 0.1^2$ (second row) and of Model 3 with 50 observations per district and $\tilde{\tau}_l^2 = 0.5^2$ (third row). The first column shows the estimates of the Correlated Method, the second column those of the Uncorrelated Method.

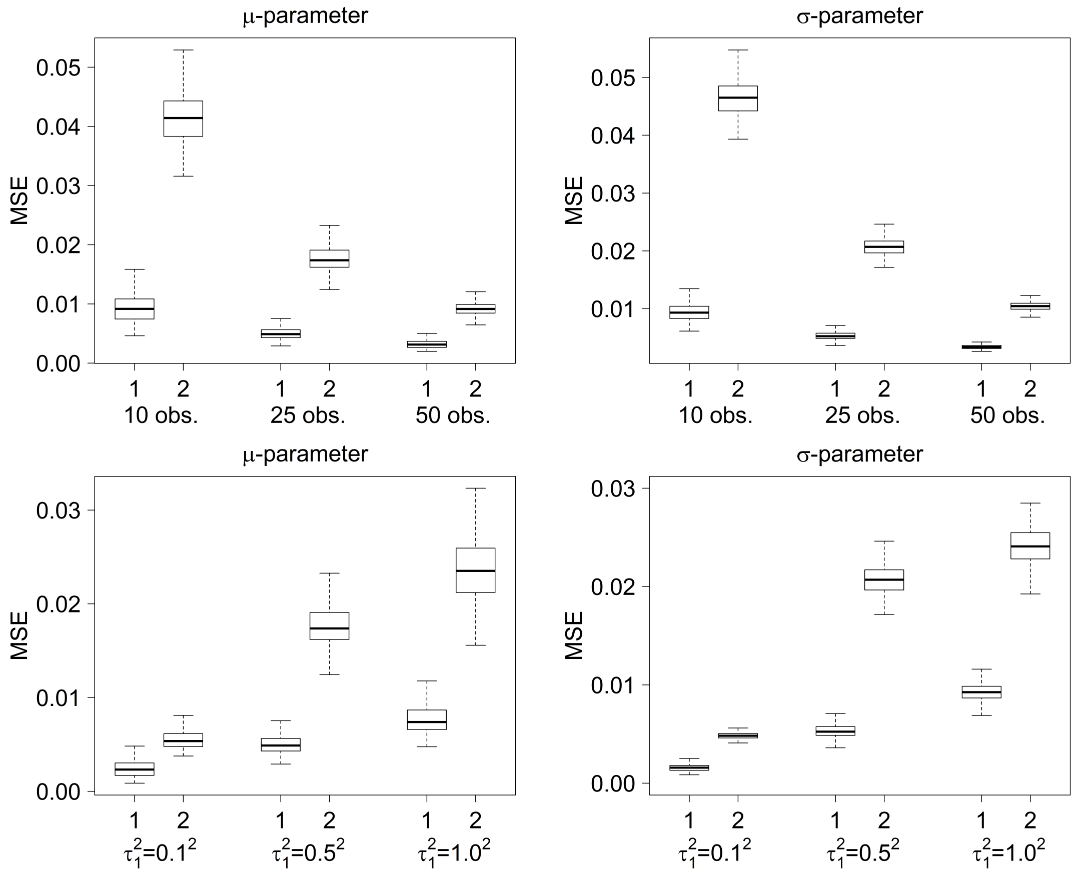

Figure 15: MSEs for the μ- and σ-parameter in the Gaussian models using the Correlated Method ("1") and the Uncorrelated Method ("2") for estimation. The first row shows the results for the different numbers of observations, the second row those for the different variances.

With respect to the scaling factors, however, the results slightly deviate. On the one hand, their variation is somewhat higher if the estimation is based on Markov random fields (Correlated Method), as can be seen from Table 3, showing the summary statistics of the random scaling factors of the floor area for both the μ- and the σ-parameter. On the other hand, as expected, the spatial distribution of the scaling factors is much smoother when using the Correlated Method. The geographical maps of the scaling factors of the floor area are depicted in Figure 19. For the μ-parameter, we see a clear difference between Western and Eastern Germany that is more marked when using Markov random field priors for the random effects. For the σ-parameter, the Correlated Method provides a distinct north-south divide that is not that pronounced when using the Uncorrelated Method.

Parameter	Method	Min.	1st Quart.	Mean	3rd Quart.	Max.
μ	Correlated Method	0.04	0.88	1.00	1.14	1.56
μ	Uncorrelated Method	0.49	0.90	1.00	1.11	1.60
σ	Correlated Method	−4.00	0.48	1.00	1.75	4.01
σ	Uncorrelated Method	−2.16	0.58	1.00	1.50	3.04

Table 3: *Random scaling factors of the floor area for the μ- and the σ-parameter.*

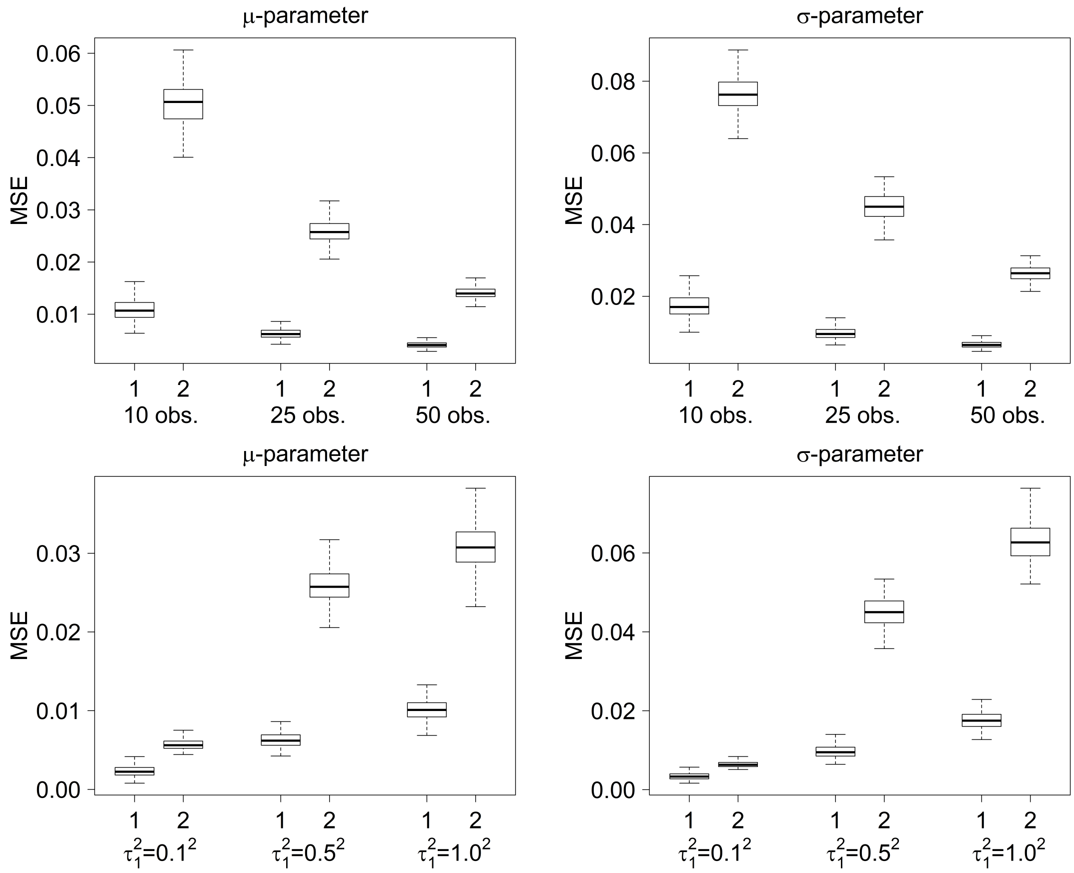

Figure 16: MSEs for the μ- and σ-parameter in the Gamma models using the Correlated Method ("1") and the Uncorrelated Method ("2") for estimation. The first row shows the results for the different numbers of observations, the second row those for the different variances.

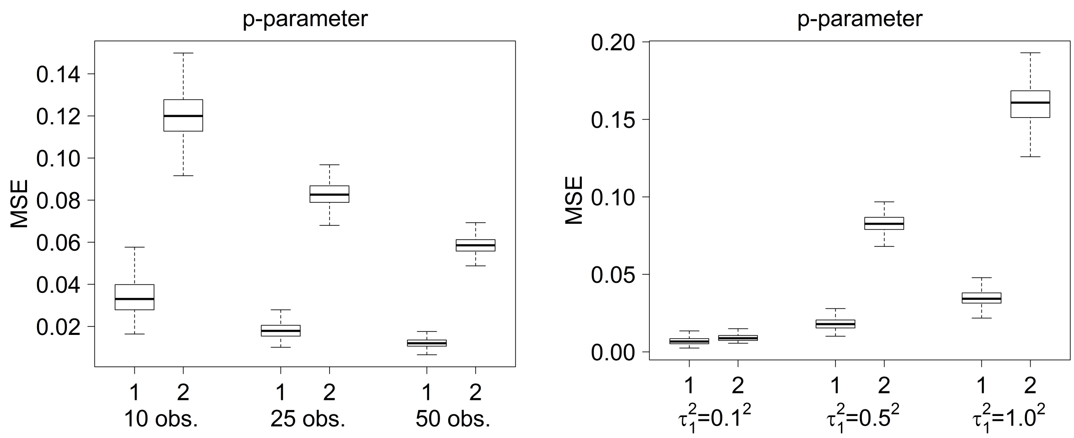

Figure 17: MSEs for the p-parameter in the Binomial models using the Correlated Method ("1") and the Uncorrelated Method ("2") for estimation. The left panel shows the results for the different numbers of observations, the right one those for the different variances.

For the plot area, we get similar results. The ranges of the scaling factors are slightly higher when estimating them using Markov random fields. For the μ-parameter, for example, the scaling factors go from -0.46 to 2.11 according to the Correlated Method, while they only range from -0.08 to 2.05 when using the Uncorrelated Method, see Table

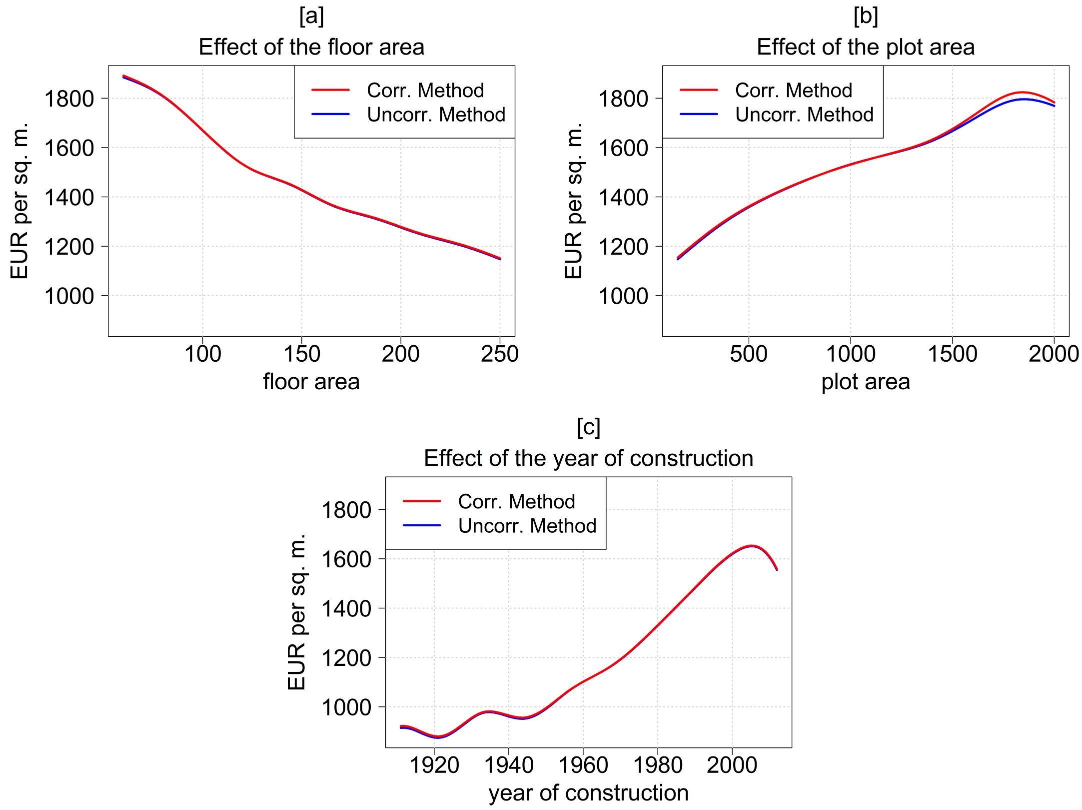

Figure 18: Mean effects of the covariates for the two estimation methods. [a]: Average effect of the floor area. [b]: Average effect of the plot area. [c]: Average effect of the year of construction.

4 for details. Again, the spatial distribution of the scaling factors is smoother when using the Correlated Method, which is particularly apparent for the σ-parameter, shown in the second row of Figure 20.

Parameter	Method	Min.	1^{st} Quart.	Mean	3^{rd} Quart.	Max.
μ	Correlated Method	-0.46	0.73	1.00	1.32	2.11
μ	Uncorrelated Method	-0.08	0.74	1.00	1.25	2.05
σ	Correlated Method	-1.91	0.51	1.00	1.46	3.19
σ	Uncorrelated Method	-0.50	0.58	1.00	1.38	3.30

Table 4: *Random scaling factors of the plot area for the μ- and the σ-parameter.*

Finally, the estimated scaling factors of the year of construction are summarized in Table 5. Once more, we see a slightly higher variation in the results coming from the Correlated Method, even though the differences are smaller than for the other covariates. With respect to the spatial distribution of the scaling factors, we again get much smoother effects when using Markov random fields, see Figure 21.

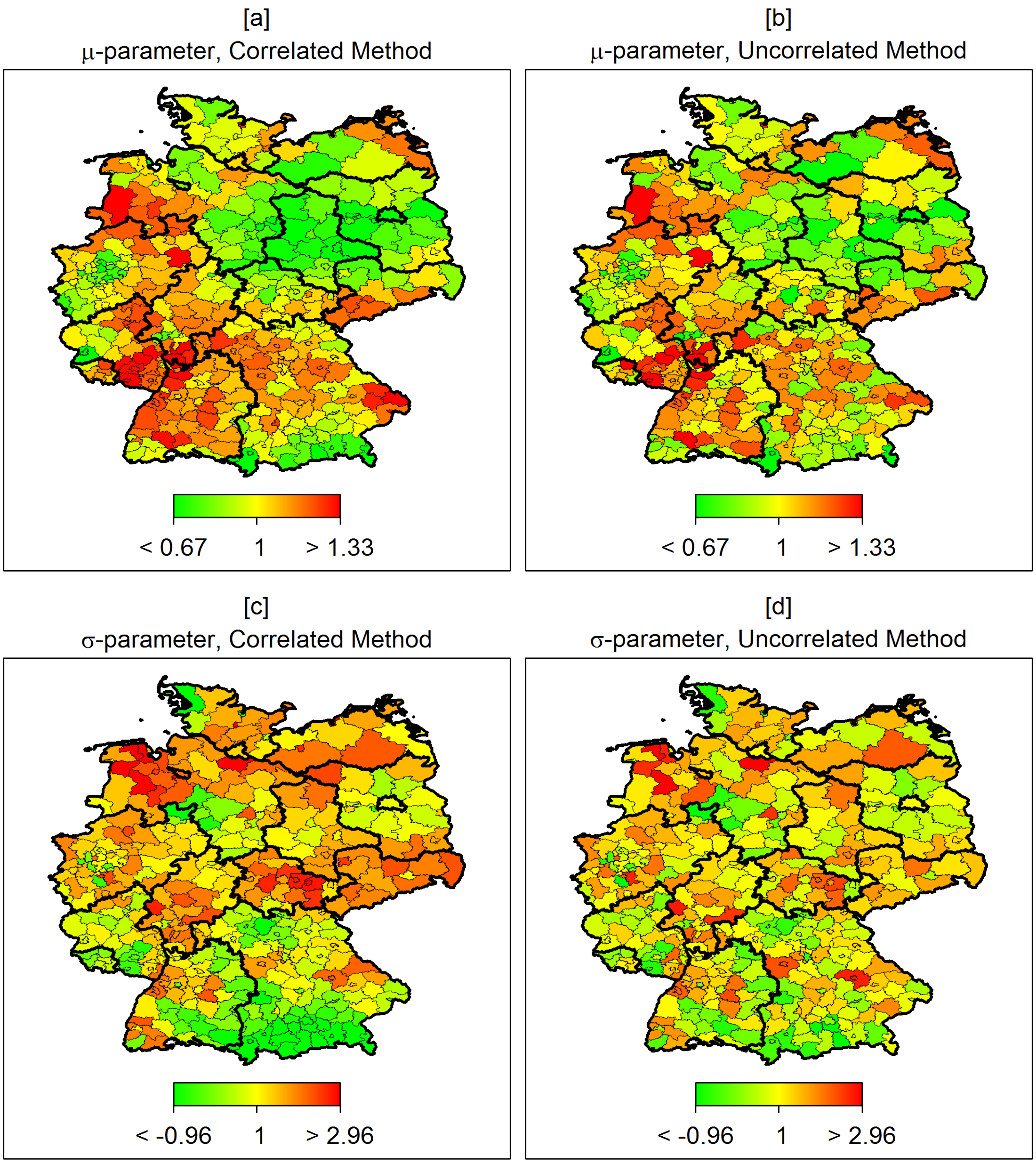

Figure 19: Random scaling factors of the floor area. [a]: Estimation results for the μ-parameter using the Correlated Method. [b]: Estimation results for the μ-parameter using the Uncorrelated Method. [c]: Estimation results for the σ-parameter using the Correlated Method. [d]: Estimation results for the σ-parameter using the Uncorrelated Method.

Model comparison

So far, we have seen that the estimation method for correlated random scaling factors based on Markov random field priors indeed leads to smoother spatial effects than the method for independent random scaling factors. Using different criteria, we now want to evaluate which results are preferable.

Accounting for both the fit of the data and the model complexity, we first investigate the performance of the two estimation methods with respect to the deviance information criterion (DIC) of Spiegelhalter et al. (2002), which we easily can compute from the

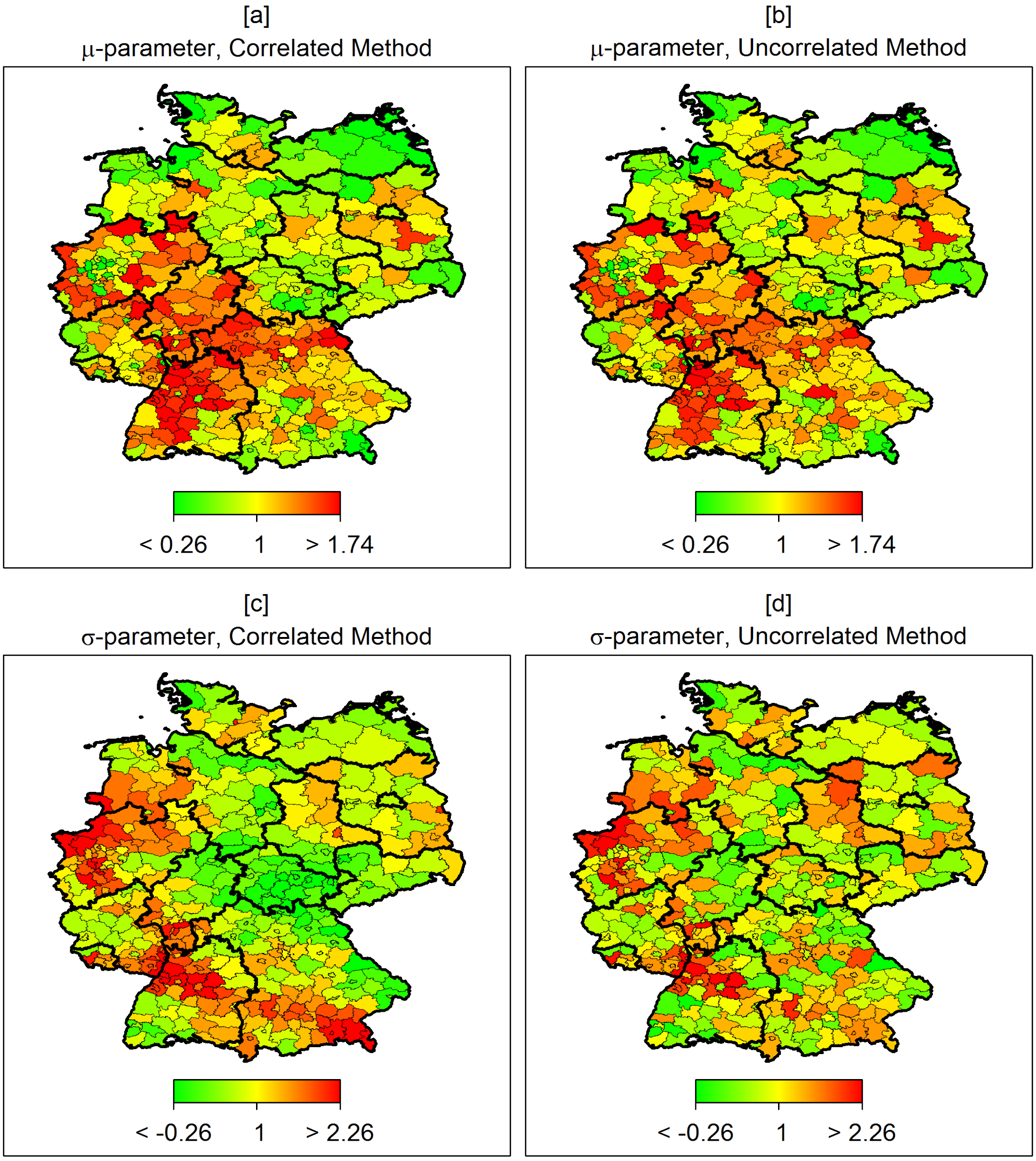

Figure 20: Random scaling factors of the plot area. [a]: Estimation results for the μ-parameter using the Correlated Method. [b]: Estimation results for the μ-parameter using the Uncorrelated Method. [c]: Estimation results for the σ-parameter using the Correlated Method. [d]: Estimation results for the σ-parameter using the Uncorrelated Method.

MCMC outputs. As we can see from Table 6, exploiting the spatial correlation of the scaling factors by Markov random fields increases the performance of the model.

In order to evaluate the predictive ability of the models, we refer to the scores proposed by Gneiting and Raftery (2007). In particular, we consider the logarithmic score, the quadratic score, the spherical score as well as the continuous ranked probability score (CRPS). According to Table 7, the model based on the estimation procedure with Markov random fields has consistently higher scores, confirming the better performance of the model.

Parameter	Method	Min.	1st Quart.	Mean	3rd Quart.	Max.
μ	Correlated Method	0.18	0.81	1.00	1.18	1.78
μ	Uncorrelated Method	0.14	0.84	1.00	1.17	1.66
σ	Correlated Method	-0.09	0.66	1.00	1.29	2.69
σ	Uncorrelated Method	-0.10	0.73	1.00	1.27	2.38

Table 5: *Random scaling factors of the year of construction for the μ- and the σ-parameter.*

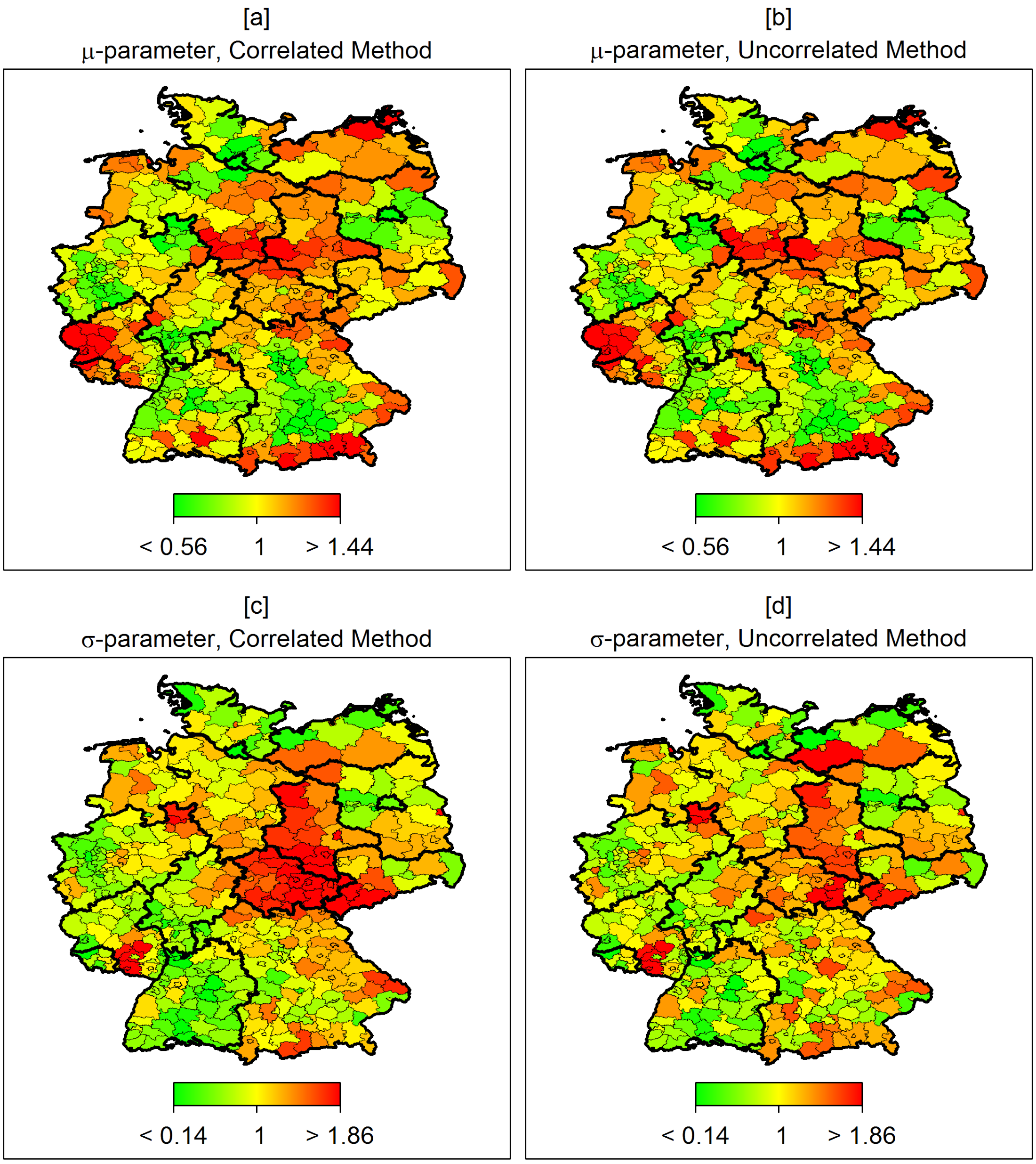

Figure 21: Random scaling factors of the year of construction. [a]: Estimation results for the μ-parameter using the Correlated Method. [b]: Estimation results for the μ-parameter using the Uncorrelated Method. [c]: Estimation results for the σ-parameter using the Correlated Method. [d]: Estimation results for the σ-parameter using the Uncorrelated Method.

Method	DIC
Correlated Method	$50,560$
Uncorrelated Method	$50,753$

Table 6: *DIC based on the two different estimation methods.*

Method	Log. score	Quadratic score	Spherical score	CRPS
Correlated Method	$-\mathbf{0.2608}$	**0.9778**	**0.9766**	$-\mathbf{0.1884}$
Uncorrelated Method	-0.2621	0.9759	0.9759	-0.1885

Table 7: *Comparison of average score contributions of the two estimation methods.*

5 Conclusion

Random scaling factors are a useful tool to model cluster-specific differences in covariate effects. In the case of spatial clustering, one often observes geographical patterns in the distribution of the scaling factors. For such situations, this paper provides inference based on Markov random field priors that allows for estimating spatially correlated scaling factors within the framework of distributional regression. In extensive simulation experiments we show that

- the estimation procedure works well for different response distributions,
- we need much less observations per cluster to get accurate results compared to estimating the scaling factors without incorporating their spatial correlation and
- our approach even allows for identifying scaling factors whose variance is small.

Applying the proposed methodology to German real estate data enhances the performance of the model with respect to both model fit and predictive ability.

For the future, several directions for further research are conceivable. First, we plan to provide other spatial smoothing techniques (e.g. Kriging) for estimating correlated scaling factors. Second, we aim to implement structured additive predictors for the scaling factors themselves, which would allow us to explain them, for example, by cluster-specific covariates. Third, in our application example, we have obtained negative scaling factors for some districts. Notwithstanding that they have occurred only occasionally, negative scaling factors seem rather unlikely in most applications from an economic perspective. Thus, we intend to provide alternative scaling factors that are positive by default, e.g. by defining them by $\exp(\alpha_c)$ or by implementing truncated factors.

References

Besag, J., J. York, and A. Mollié (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics 43*, 1–59.

Brunauer, W. A., S. Lang, P. Wechselberger, and S. Bienert (2010). Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna. *Journal of Real Estate Finance and Economics 40*, 390–410.

Eilers, P., B. Marx, and M. Durbán (2015). Twenty years of P-splines. *Statistics and Operations Research Transactions 39*, 149–186.

Eilers, P. H. C. and B. D. Marx (1996). Flexible Smoothing with B-splines and Penalties (with comments and rejoinder). *Statistical Science 11*, 89–121.

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

Fahrmeir, L. and S. Lang (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Applied Statistics 50*, 201–220.

Gneiting, T. and A. E. Raftery (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association 102*, 359–378.

Klein, N., T. Kneib, and S. Lang (2014). Bayesian generalized additive models for location, scale and shape for zero-inflated and overdispersed count data. *Journal of the American Statistical Association 10*, 405–419.

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

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

Razen, A. and S. Lang (2016). Random Scaling Factors in Bayesian Distributional Regression Models with an Application to Real Estate Data. Technical report, Department of Statistics, University of Innsbruck.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields*. Chapman & Hall/CRC, Boca Raton, FL.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society – Series B 64*, 583–639.

Weber, A., W. Steiner, and S. Lang (2016). A comparison of semiparametric and heterogeneous store sales models for optimal category pricing. *OR Spectrum (to appear)*.

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

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

2016-33 **Alexander Razen, Stefan Lang, Judith Santer:** Estimation of spatially correlated random scaling factors based on Markov random field priors

2016-32 **Meike Köhler, Nikolaus Umlauf, Andreas Beyerlein, Christiane Winkler, Anette-Gabriele Ziegler, Sonja Greven:** Flexible Bayesian additive joint models with an application to type 1 diabetes research

2016-31 **Markus Dabernig, Georg J. Mayr, Jakob W. Messner, Achim Zeileis:** Simultaneous ensemble post-processing for multiple lead times with standardized anomalies

2016-30 **Alexander Razen, Stefan Lang:** Random scaling factors in Bayesian distributional regression models with an application to real estate data

2016-29 **Glenn Dutcher, Daniela Glätzle-Rützler, Dmitry Ryvkin:** Don't hate the player, hate the game: Uncovering the foundations of cheating in contests

2016-28 **Manuel Gebetsberger, Jakob W. Messner, Georg J. Mayr, Achim Zeileis:** Tricks for improving non-homogeneous regression for probabilistic precipitation forecasts: Perfect predictions, heavy tails, and link functions

2016-27 **Michael Razen, Matthias Stefan:** Greed: Taking a deadly sin to the lab

2016-26 **Florian Wickelmaier, Achim Zeileis:** Using recursive partitioning to account for parameter heterogeneity in multinomial processing tree models

2016-25 **Michel Philipp, Carolin Strobl, Jimmy de la Torre, Achim Zeileis:** On the estimation of standard errors in cognitive diagnosis models

2016-24 **Florian Lindner, Julia Rose:** No need for more time: Intertemporal allocation decisions under time pressure

2016-23 **Christoph Eder, Martin Halla:** The long-lasting shadow of the allied occupation of Austria on its spatial equilibrium

2016-22 **Christoph Eder:** Missing men: World War II casualties and structural change

2016-21 **Reto Stauffer, Jakob Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Ensemble post-processing of daily precipitation sums over complex terrain using censored high-resolution standardized anomalies

2016-20 **Christina Bannier, Eberhard Feess, Natalie Packham, Markus Walzl:** Incentive schemes, private information and the double-edged role of competition for agents

2016-19 **Martin Geiger, Richard Hule:** Correlation and coordination risk

2016-18 **Yola Engler, Rudolf Kerschbamer, Lionel Page:** Why did he do that? Using counterfactuals to study the effect of intentions in extensive form games

2016-17 **Yola Engler, Rudolf Kerschbamer, Lionel Page:** Guilt-averse or reciprocal? Looking at behavioural motivations in the trust game

2016-16 **Esther Blanco, Tobias Haller, James M. Walker:** Provision of public goods: Unconditional and conditional donations from outsiders

2016-15 **Achim Zeileis, Christoph Leitner, Kurt Hornik:** Predictive bookmaker consensus model for the UEFA Euro 2016

2016-14 **Martin Halla, Harald Mayr, Gerald J. Pruckner, Pilar García-Gómez:** Cutting fertility? The effect of Cesarean deliveries on subsequent fertility and maternal labor supply

2016-13 **Wolfgang Frimmel, Martin Halla, Rudolf Winter-Ebmer:** How does parental divorce affect children's long-term outcomes?

2016-12 **Michael Kirchler, Stefan Palan:** Immaterial and monetary gifts in economic transactions. Evidence from the field

2016-11 **Michel Philipp, Achim Zeileis, Carolin Strobl:** A toolkit for stability assessment of tree-based learners

2016-10 **Loukas Balafoutas, Brent J. Davis, Matthias Sutter:** Affirmative action or just discrimination? A study on the endogenous emergence of quotas *forthcoming in Journal of Economic Behavior and Organization*

2016-09 **Loukas Balafoutas, Helena Fornwagner:** The limits of guilt

2016-08 **Markus Dabernig, Georg J. Mayr, Jakob W. Messner, Achim Zeileis:** Spatial ensemble post-processing with standardized anomalies

2016-07 **Reto Stauffer, Jakob W. Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model

2016-06 **Michael Razen, Jürgen Huber, Michael Kirchler:** Cash inflow and trading horizon in asset markets

2016-05 **Ting Wang, Carolin Strobl, Achim Zeileis, Edgar C. Merkle:** Score-based tests of differential item functioning in the two-parameter model

2016-04 **Jakob W. Messner, Georg J. Mayr, Achim Zeileis:** Non-homogeneous boosting for predictor selection in ensemble post-processing

2016-03 **Dietmar Fehr, Matthias Sutter:** Gossip and the efficiency of interactions

2016-02 **Michael Kirchler, Florian Lindner, Utz Weitzel:** Rankings and risk-taking in the finance industry

2016-01 **Sibylle Puntscher, Janette Walde, Gottfried Tappeiner:** Do methodical traps lead to wrong development strategies for welfare? A multilevel approach considering heterogeneity across industrialized and developing countries

2015-16 **Niall Flynn, Christopher Kah, Rudolf Kerschbamer:** Vickrey Auction vs BDM: Difference in bidding behaviour and the impact of other-regarding motives

2015-15 **Christopher Kah, Markus Walzl:** Stochastic stability in a learning dynamic with best response to noisy play

2015-14 **Matthias Siller, Christoph Hauser, Janette Walde, Gottfried Tappeiner:** Measuring regional innovation in one dimension: More lost than gained?

2015-13 **Christoph Hauser, Gottfried Tappeiner, Janette Walde:** The roots of regional trust

2015-12 **Christoph Hauser:** Effects of employee social capital on wage satisfaction, job satisfaction and organizational commitment

2015-11 **Thomas Stöckl:** Dishonest or professional behavior? Can we tell? A comment on: Cohn et al. 2014, Nature 516, 86-89, “Business culture and dishonesty in the banking industry”

2015-10 **Marjolein Fokkema, Niels Smits, Achim Zeileis, Torsten Hothorn, Henk Kelderman:** Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees

2015-09 **Martin Halla, Gerald Pruckner, Thomas Schober:** The cost-effectiveness of developmental screenings: Evidence from a nationwide programme *forthcoming in Journal of Health Economics*

2015-08 **Lorenz B. Fischer, Michael Pfaffermayr:** The more the merrier? Migration and convergence among European regions

2015-07 **Silvia Angerer, Daniela Glätzle-Rützler, Philipp Lergetporer, Matthias Sutter:** Cooperation and discrimination within and across language borders: Evidence from children in a bilingual city *forthcoming in European Economic Review*

2015-06 **Martin Geiger, Wolfgang Luhan, Johann Scharler:** When do Fiscal Consolidations Lead to Consumption Booms? Lessons from a Laboratory Experiment *forthcoming in Journal of Economic Dynamics and Control*

2015-05 **Alice Sanwald, Engelbert Theurl:** Out-of-pocket payments in the Austrian healthcare system - a distributional analysis

2015-04 **Rudolf Kerschbamer, Matthias Sutter, Uwe Dulleck:** How social preferences shape incentives in (experimental) markets for credence goods *forthcoming in Economic Journal*

2015-03 **Kenneth Harttgen, Stefan Lang, Judith Santer:** Multilevel modelling of child mortality in Africa

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

University of Innsbruck

Working Papers in Economics and Statistics

2016-33

Alexander Razen, Stefan Lang, Judith Santer

Estimation of spatially correlated random scaling factors based on Markov random field priors

Abstract
Multiplicative random effects allow for cluster-specific scaling of covariate effects. In many applications with spatial clustering, however, the random effects additionally show some geographical pattern, which usually can not sufficiently be captured with existing estimation techniques. Relying on Markov random fields, we present a fully Bayesian inference procedure for spatially correlated scaling factors. The estimation is based on highly efficient Markov Chain Monte Carlo (MCMC) algorithms and is smoothly incorporated into the framework of distributional regression.

We run a comprehensive simulation study for different response distributions to examine the statistical properties of our approach. We also compare our results to those of a general estimation procedure for independent random scaling factors. Furthermore, we apply the method to German real estate data and show that exploiting the spatial correlation of the scaling factors further improves the performance of the model.

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