University of Innsbruck

Random scaling factors in Bayesian distributional regression models with an application to real estate data

Alexander Razen, Stefan Lang

Working Papers in Economics and Statistics

2016-30

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 Scaling Factors in Bayesian Distributional Regression Models with an Application to Real Estate Data

Alexander Razen
University of Innsbruck

Stefan Lang
University of Innsbruck

Abstract

Distributional structured additive regression provides a flexible framework for modeling each parameter of a potentially complex response distribution in dependence of covariates. Structured additive predictors allow for an additive decomposition of covariate effects with nonlinear effects and time trends, unit- or cluster-specific heterogeneity, spatial heterogeneity and complex interactions between covariates of different type. Within this framework, we present a simultaneous estimation approach for multiplicative random effects that allow for cluster-specific heterogeneity with respect to the scaling of a covariate's effect. More specifically, a possibly nonlinear function $f(z)$ of a covariate z may be scaled by a multiplicative cluster-specific random effect $(1 + \alpha_c)$. Inference is fully Bayesian and is based on highly efficient Markov Chain Monte Carlo (MCMC) algorithms.

We investigate the statistical properties of our approach within extensive simulation experiments for different response distributions. Furthermore, we apply the methodology to German real estate data where we identify significant district-specific scaling factors. According to the deviance information criterion, the models incorporating these factors perform significantly better than standard models without random scaling factors.

Keywords: iteratively weighted least squares proposals, MCMC, multiplicative random effects, structured additive predictors

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

1 Introduction

Classical regression models, such as generalized linear models (GLMs, see McCullagh and Nelder, 1989), generalized additive models (GAMs, see Hastie and Tibshirani, 1990, or Wood, 2006) or structured additive regression models (STAR models, see Brezger and Lang, 2006, or Fahrmeir et al., 2013), assume that the distribution of a response variable y belongs to an exponential family and relate its mean to a number of covariates. A potential dependence of other moments of the response distribution, however, is neglected.

Generalized additive models for location, scale and shape (GAMLLS, see Rigby and Stasinopoulos, 2005) and its Bayesian version of distributional regression (see Klein et al., 2015) provide a more flexible framework. On the one hand, it is no longer restricted to the exponential family and on the other hand, it allows for modeling each parameter of the response distribution – and not only the mean – in dependence of a set of covariates.

In many applications, the data consists of a number of different clusters. In general, it is not guaranteed that a covariate's effect on a parameter of the response distribution – be it its mean or another parameter – is homogeneous across these clusters. In real estate data, for example, a frequently observed phenomenon is that the price effects of covariates vary from one spatial unit to another. However, completely different functional forms in each cluster are not common.

In order to deal with this challenge, Wechselberger et al. (2008) suggest the use of cluster-specific random scaling factors. In doing so, one still assumes homogeneity for the functional form of the response function but allows for heterogeneity with respect to its scaling. Lang et al. (2015) and Weber et al. (2015) successfully have applied this approach to store sales models, considerably improving the predictive validity of the models. In a real estate context, Brunauer et al. (2010) introduced spatial scaling factors to account for district-specific heterogeneity in rent prices.

Nevertheless, the present method has two drawbacks. On the one hand, the analyses so far are restricted to modeling the mean in Gaussian response distributions, ignoring alternative response distributions and covariate effects on other moments. On the other hand, due to identifiability reasons, one currently has to assume monotonicity for the response functions of the covariates, which in many applications is not justified a priori.

The aim of this paper is to provide a general framework that allows the use of random scaling factors for arbitrary response functions in applications that go beyond the modeling of the response distribution's mean. For this purpose, we extend the idea of random scaling factors to distributional regression models and develop identifiability constraints that do no longer restrict the response functions to be monotone.

We investigate the properties of our approach in comprehensive simulation scenarios including Gaussian, Gamma and Binomial models. We then apply the method to a German real estate dataset with almost $100,000$ observations. We consider and compare different distributional regression models where house-specific attributes flexibly are estimated using P-splines that at the same time are scaled according to district-specific random scaling factors. The results are compared to standard models without scaling factors.

The remainder of the paper is structured as follows: In Section 2 we present the methodology that subsequently is tested in different simulation scenarios in Section 3. Section 4 attends to the real estate data and the model specification before we present the results in Section 5. We conclude in Section 6 with an outlook on future research perspectives.

2 Methodology

2.1 Distributional regression models

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 $\boldsymbol{x}$. Bayesian distributional regression as introduced in Klein et al. (2015) now 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,

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

$l = 1, \ldots, L$. The predictors are defined in terms of (potentially different) subsets of the covariates,

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

where the functions $\boldsymbol{f}_{jl}$ are possibly nonlinear functions of the covariates $\boldsymbol{z}_{jl}$ 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

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

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

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

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 $\boldsymbol{f}$.

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 Lang and Brezger (2004). 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 $u_k \sim N(0, \tau^2)$ 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 (2) with penalty matrix given by $\boldsymbol{K} = \boldsymbol{D}'\boldsymbol{D}$, where $\boldsymbol{D}$ 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 $I(\boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{0})$ 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.

2.2 Multiplicative random effects

As outlined in the introduction, in many applications the data is clustered. Real estate data, for example, typically is clustered in spatial units (e.g. districts, states, etc.). Usually, there is no economic reason to assume homogeneous covariate effects across these units. In contrast, different consumer price sensitivities originating from varying levels of income, diverse value of land or different ways of construction suggest spatial heterogeneity in price response. Indeed, it is reasonable to assume the effects to have the same functional form but to vary with respect to the scaling of the function. Thus, in order to account for this kind of heterogeneity, we allow for cluster-specific random scaling factors for some or all of the nonlinear functions $\boldsymbol{f}_j$ in (1). This leads to predictors of the form

$$\eta_i = (1 + \alpha_{1c_i})\, f_1(z_{1i}) + \ldots + (1 + \alpha_{qc_i})\, f_q(z_{qi}) + \boldsymbol{x}_i'\boldsymbol{\gamma}, \tag{3}$$

$i = 1, \ldots n$, where $c_i \in \{1, \ldots, C\}$ is the cluster index of the respective observation and the α_{jc}, $j = 1, \ldots, q$, are independent and normally distributed random effects with mean 0 and variance $\tilde{\tau}_j$, i.e.

$$\alpha_{jc} \,|\, \tilde{\tau}_j^2 \sim \mathcal{N}\left(0, \tilde{\tau}_j\right), \quad c = 1, \ldots, C.$$

A positive random effect $\alpha_{jc} > 0$ leads to a scaling up of the function f_j indicating an increased price sensitivity while a negative random effect $\alpha_{jc} < 0$ refers to weaker price sensitivity. For the variance parameters $\tilde{\tau}_j^2$ we assign the usual inverse Gamma priors $\tilde{\tau}_j^2 \sim IG(\tilde{a}, \tilde{b})$.

A priori, the parameters are not identifiable since there is an arbitrary multiplicative constant for the functions f_j. Thus, previous works (e.g. Lang et al., 2015, or Weber et al., 2015) assumed the response functions to be monotone and restricted their spread by assuming

$$\sum_{k=1}^{K} \beta_{jk}^2 = d_j.$$

Typically, the constant d_j is chosen such that the squared sum of the coefficients is identical to that of the corresponding model without scaling factors.

Instead of making assumptions about the coefficients β_{jk}, we propose to assume

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

This assumption is preferable for at least two reasons: First, we do no longer need monotonicity constraints for the response functions f_j. Second, the unscaled functions now can be interpreted as the average effect over all clusters.

2.3 Inference

For the sake of illustration, we consider a Gaussian model with a single predictor for the mean parameter of the form (3).

The description of posterior inference is facilitated by rewriting the model equation in matrix notation. We obtain

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

where $\boldsymbol{Z}_j$, $j = 1, \ldots, q$ is the usual design matrix for the j-th nonparametric term (e.g. P-spline), $\boldsymbol{D}_j = diag(1 + \alpha_{jc_1}, \ldots, 1 + \alpha_{jc_n})$ is an $n \times n$ diagonal matrix with the random scaling factors in the main diagonal, and $\boldsymbol{\beta}_j$ is the j-th vector of regression coefficients.

Each of the q terms $\boldsymbol{D}_j\boldsymbol{Z}_j\boldsymbol{\beta}_j$ is formally in the form of a varying coefficient term (Hastie and Tibshirani, 1993) with effect modifier matrix $\boldsymbol{Z}_j$ and (pseudo) values of the interacting variable stored in the diagonal matrix $\boldsymbol{D}_j$.

An alternative formulation in terms of the scaling parameter vectors $\boldsymbol{\alpha}_j = (\alpha_{j1}, \ldots, \alpha_{jC})'$ is given by

$$\boldsymbol{y} = \cdots + \boldsymbol{f}_j + \tilde{\boldsymbol{D}}_j\tilde{\boldsymbol{Z}}_j\boldsymbol{\alpha}_j + \cdots, \tag{5}$$

where $\boldsymbol{f}_j$ is a vector of function evaluations at the observed covariate values, $\tilde{\boldsymbol{D}}_j = diag(f(z_{j1}), \ldots, f(z_{jn}))$ is a $n \times n$ diagonal matrix with diagonal elements now given by the function evaluations of the nonlinear effects, and $\tilde{\boldsymbol{Z}}_j$ is 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). Again, the expressions $\tilde{\boldsymbol{D}}_j\tilde{\boldsymbol{Z}}_j\boldsymbol{\alpha}_j$ in (5) are in the form of a varying coefficient term now with effect modifier matrix $\tilde{\boldsymbol{Z}}_j$ and values of the interacting variable given in $\tilde{\boldsymbol{D}}_j$.

The varying coefficient representation (5) suggests the following two-stage estimation procedure:

Stage 1: Assume the covariate effects to be homogeneous over all clusters, i.e. set the random effects $\boldsymbol{\alpha}_j$ to zero, and estimate the model $\boldsymbol{y} = \boldsymbol{f}_1 + \ldots + \boldsymbol{f}_q + \boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ as usual.

Stage 2: Treat the estimated functions $\hat{\boldsymbol{f}}_j$ from stage 1 as a fixed offset and estimate the random effects from the varying coefficient representation (5).

Conceptually, this two-stage procedure works fine. However, simulations show that ignoring the random effects in stage 1 can raise difficulties in properly estimating the functions $\boldsymbol{f}_j$, particularly if the variances of the random effects are large. Inappropriate estimations of the functions $\boldsymbol{f}_j$ may then lead to peculiar results for the random effects in stage 2. Thus, instead of this two-stage procedure, we propose a simultaneous estimation approach, where Gibbs updates for the random scaling terms are employed by alternately obeying the two different varying coefficient representations (4) and (5).

The full conditionals of the regression parameters $\boldsymbol{\beta}_j$ are easily derived from (4) and are multivariate Gaussian $\boldsymbol{\beta}_j \mid \cdot \sim \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ 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 scaling factors are derived from (5) with $\boldsymbol{\alpha}_j \,|\cdot \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_j, \tilde{\boldsymbol{\Sigma}}_j)$ and

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

In contrast to "usual" varying coefficients terms, the diagonal matrices $\boldsymbol{D}_j$ and $\tilde{\boldsymbol{D}}_j$ of the (pseudo) interacting variables are not constant during the Gibbs sampler. Hence cross products and other quantities can not be computed and stored in advance and numerical efficient updating is considerably complicated. However, the methodology for highly efficient Gibbs updates described in Lang et al., 2014, can be applied.

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

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

For most updates of regression parameters $\boldsymbol{\beta}_j$ and $\boldsymbol{\alpha}_j$ in distributional regression Gibbs steps are not available because the full conditionals are no longer Gaussian. Then we rely on the Metropolis Hastings updates with IWLS proposals as described in detail in Klein et al., 2015, and Klein et al., 2014. The key for updating $\boldsymbol{\beta}_j$ and $\boldsymbol{\alpha}_j$ is again the varying coefficient type notation of the random scaling terms given in (4) and (5).

3 Simulation

3.1 Gaussian models

Setup

In a first step, we simulate models with a normally distributed response y with mean μ and heteroscedastic variance σ^2. The parameters μ and σ are linked to predictors η_1 and η_2 via

$$\begin{aligned} \mu &= \eta_1, \\ \sigma &= \exp(\eta_2). \end{aligned}$$

The predictors are constructed as follows:

$$\begin{aligned} \eta_1 &= (1 + \alpha_{c1})\, f_1(x), \\ \eta_2 &= -0.5 + (1 + \alpha_{c2})\, f_2(x), \end{aligned}$$

where f_1 and f_2 both are the sine function $\sin(x)$ in the interval $[-\pi, \pi]$. These nonlinear functions are modified by random scaling factors $(1 + \alpha_{cl})$, $l = 1, 2$, with 20 clusters. The random effects α_{cl} are independent and normally distributed with mean 0 and variance $\tilde{\tau}_l^2$,

$$\alpha_{cl} \sim \mathcal{N}(0, \tilde{\tau}_l^2).$$

In order to evaluate the influence of both the number of observations per cluster and the variance of the random effects we analyze six different models, whose specifications are summarized in Table 1.

For illustration, the effects $(1 + \alpha_{cl}) f_l(x)$ of Model 2 are shown in Figure 1.

Model	Obs. per cluster	Variance of α_{cl}
Model 1	10	0.5^2
Model 2	50	0.5^2
Model 3	100	0.5^2
Model 4	300	0.5^2
Model 5	50	0.1^2
Model 6	50	1.0^2

Table 1: *Model specifications*

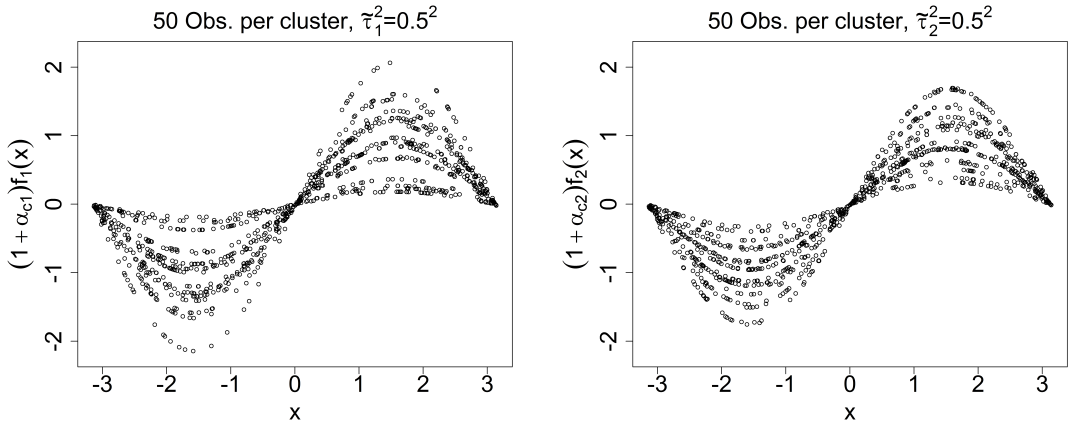

Figure 1: The functions f_l, multiplied with the respective random scaling factors.

Results

We generate 250 replications of the six models and carry out the simultaneous estimation procedure described in the previous section 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. Figure 2 shows the sampling paths of the random effects α_{cl} of the first cluster in one of the replications of model 2, Figure 3 shows the corresponding autocorrelation functions. As we can see, the draws are practically independent, indicating a good mixing.

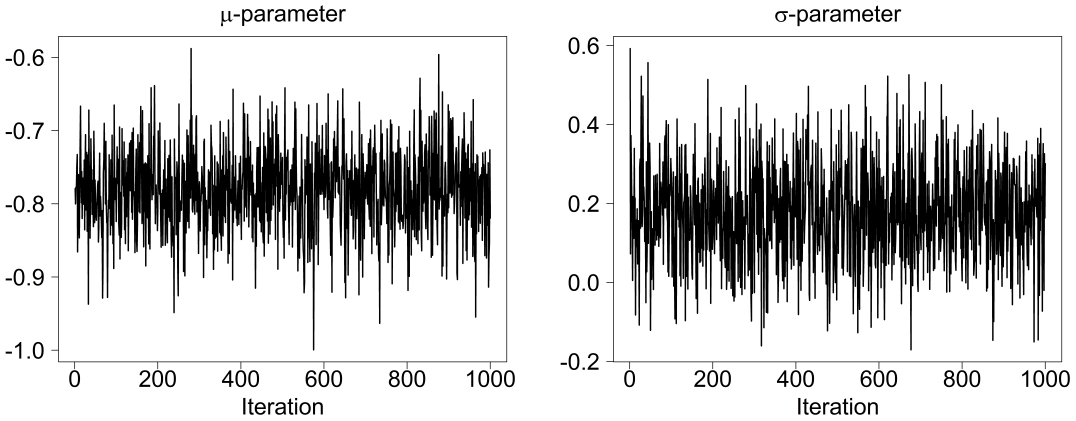

Figure 2: Sampling paths of the random effects α_{cl} of the first cluster in one of the replications of the Gaussian model 2.

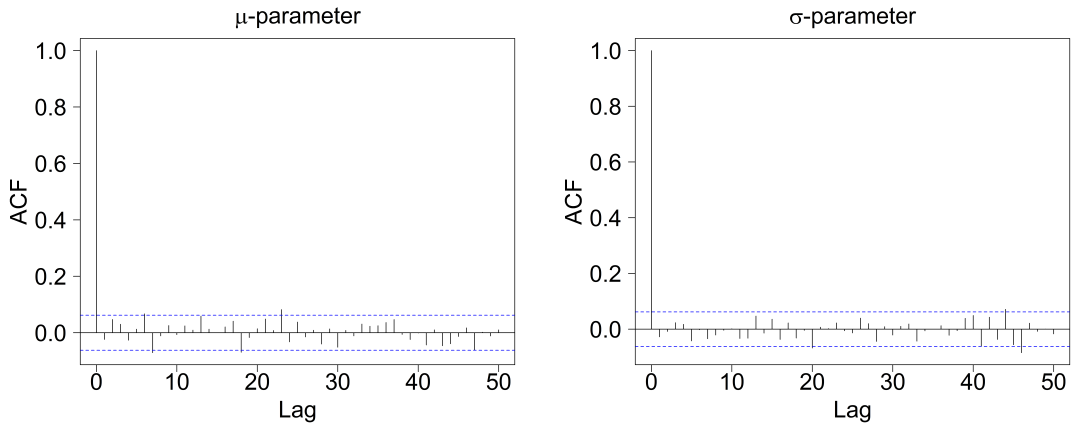

Figure 3: Autocorrelation functions of the random effects α_{cl} of the first cluster in one of the replications of the Gaussian model 2.

We then calculate the arithmetic mean from the 250 replications. Figures 4 and 5 show the average estimates of the effects f_l as well as of the cluster-specific effects $(1 + \alpha_{cl})\, f_l(x)$ for the smallest and largest random effects α_{cl} (solid). The true effects also are plotted (dashed) in order to facilitate comparison. As we can see from Figure 4, the scaled effects almost perfectly can be estimated if we have 50 or more observations per cluster (rows 2-4). Even with only 10 observations per cluster the estimation results are quite well at least for the μ-parameter. The estimation results for the σ-parameter, in contrast, are more biased. Furthermore, Figure 5 shows that if the variance of the random effects is too small, the effects can hardly be separated from the overall noise in the data (first row). In contrast, the larger the variance gets, the better the estimation results are. Thus, the estimates are more biased if the random effect is weak, which is a well-known feature of random effects estimators, see e.g. Gelman and Hill (2007).

3.2 Gamma models

Setup

In a next step, we consider models with a gamma distributed response y with mean μ and shape σ. Here, both parameters are linked to predictors η_1 and η_2 via the exponential function:

$$\begin{aligned} \mu &= \exp(\eta_1), \\ \sigma &= \exp(\eta_2). \end{aligned}$$

We set up the same predictors as for the Gaussian models using again independent and normally distributed random effects α_{cl} with 20 clusters:

$$\begin{aligned} \eta_1 &= (1 + \alpha_{c1})\, f_1(x), \\ \eta_2 &= -0.5 + (1 + \alpha_{c2})\, f_2(x), \end{aligned}$$

with f_1 and f_2 being the sine function $\sin(x)$ in the interval $[-\pi, \pi]$. We again vary the number of observations per cluster as well as the variance of the random effects according to Table 1. So the effects $(1+\alpha_{cl})f_l(x)$ are the same as in the Gaussian models, see Figure 1 for illustration.

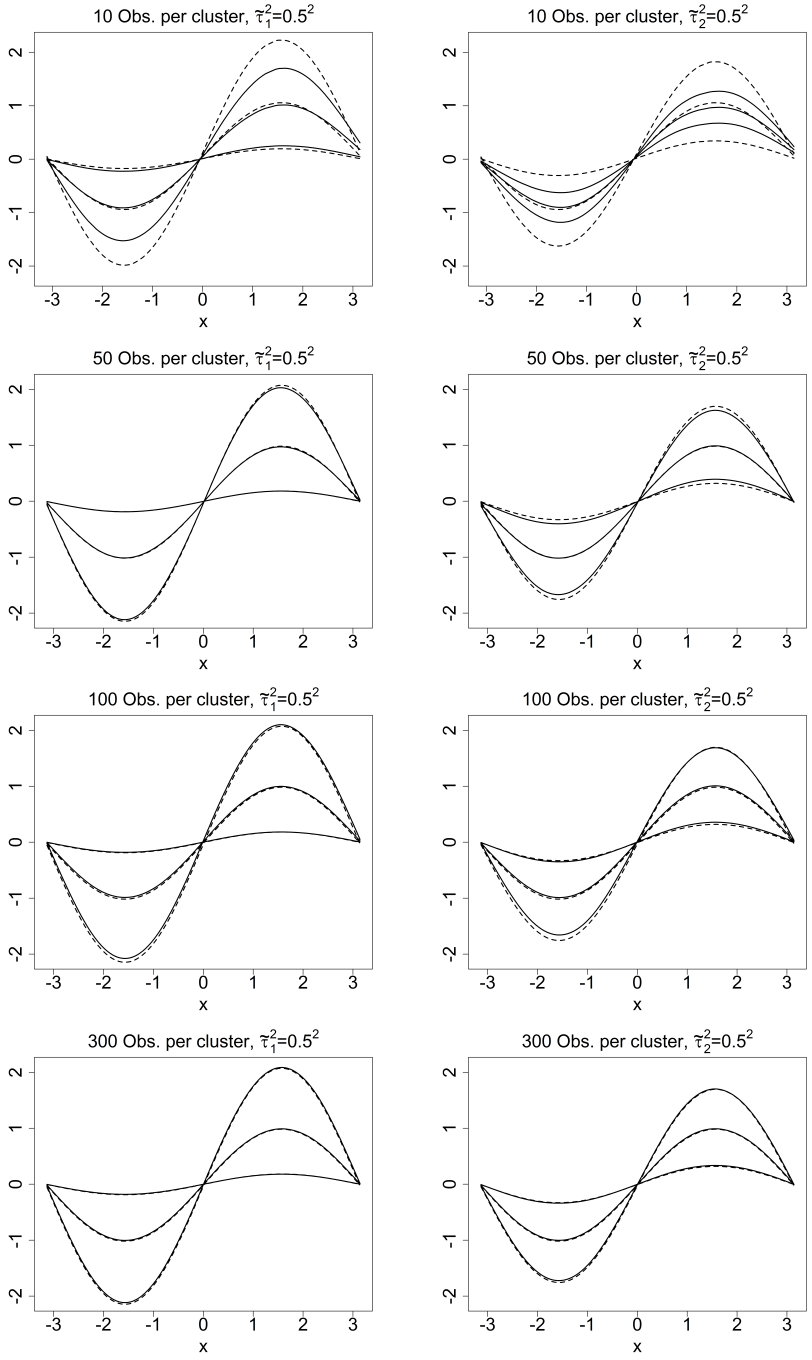

Figure 4: The average estimates of the functions f_l as well as of the smallest and largest cluster-specific effects $(1 + \alpha_{cl})\, f_l(x)$ (solid) and the respective true effects (dashed) for the Gaussian models 1 to 4. The first column shows the effects of the μ-equation, the second column those of the σ-equation.

Results

We again generate 250 replications of the six models and do the same MCMC runs as for the Gaussian models. As we can see from Figures 6 and 7, showing the sampling paths of the random effects α_{cl} of the first cluster in one of the replications of model 2 and the corresponding autocorrelation functions, the draws again are practically independent.

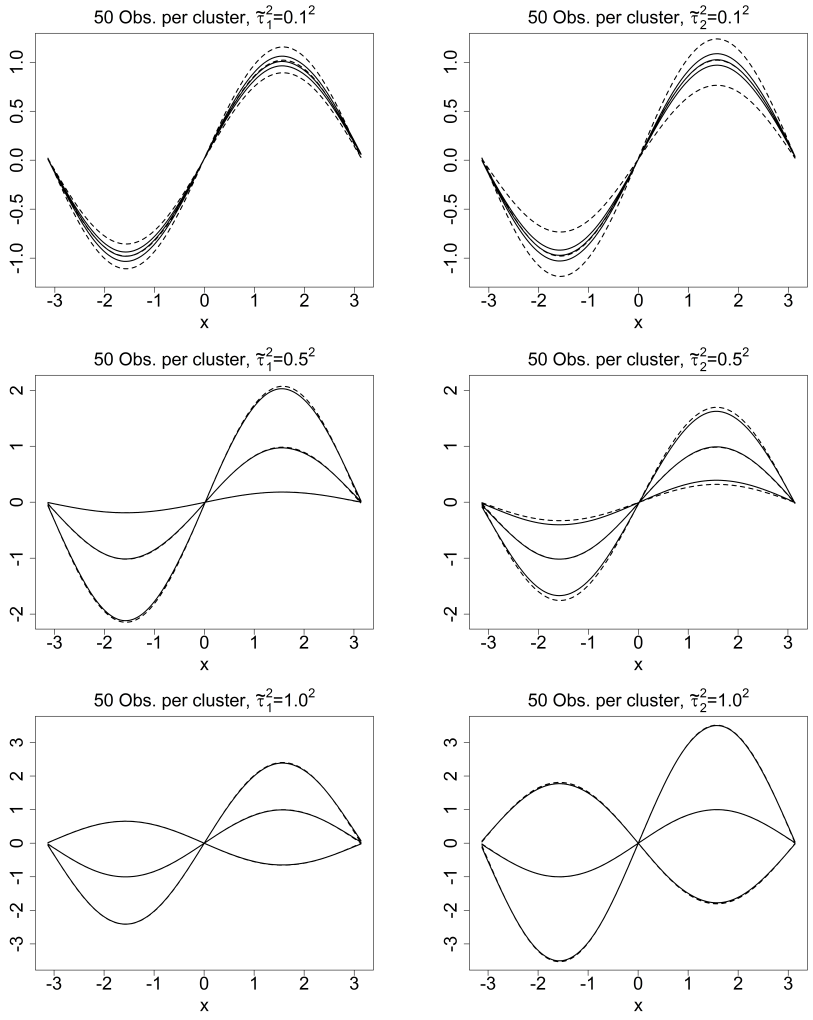

Figure 5: The average estimates of the functions f_l as well as of the smallest and largest cluster-specific effects $(1 + \alpha_{cl})\, f_l(x)$ (solid) and the respective true effects (dashed) for the Gaussian models 5, 2 and 6. The first column shows the effects of the μ-equation, the second column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

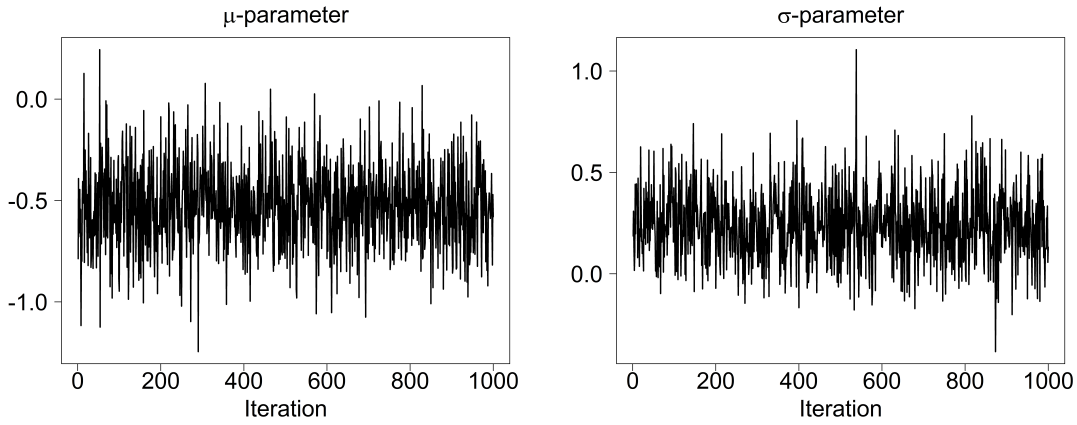

Figure 6: Sampling paths of the random effects α_{cl} of the first cluster in one of the replications of the Gamma model 2.

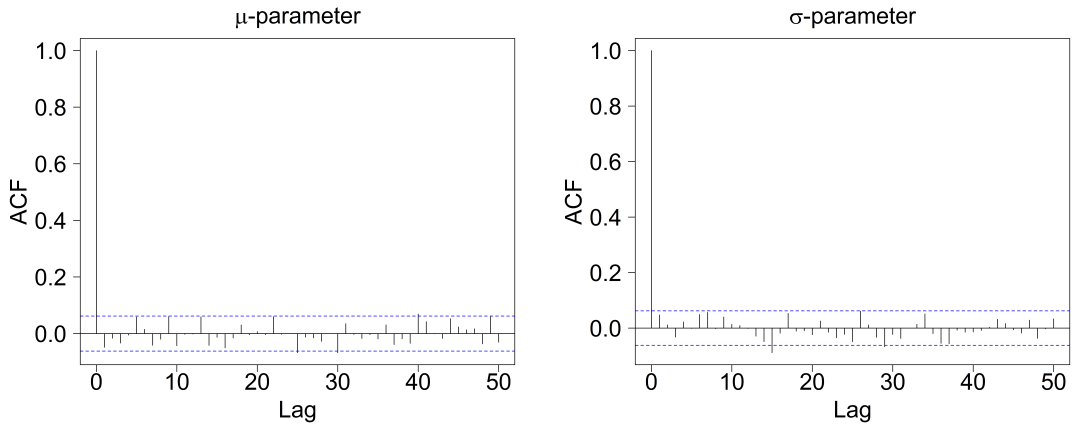

Figure 7: Autocorrelation functions of the random effects α_{cl} of the first cluster in one of the replications of the Gaussian model 2.

In contrast to the Gaussian model, it is hardly possible in the Gamma model to identify any random effect when having only 10 observations per cluster (first row of Figure 8). However, if we increase the number of observations per cluster, the results for both parameters significantly get better with almost perfect results for 100 or more observations per cluster. The impact of the variance of the random effects is similar to the Gaussian models. While the effects cannot be separated from the overall noise in the data for a small variance, the estimates gets better the larger the variance is (see Figure 9).

3.3 Binomial models

Setup

Finally, we analyze models with a binomial distributed response y with probability p:

$$y = \mathcal{B}(1, p).$$

The parameter p is linked to a predictor η via the logistic distribution function,

$$p = \frac{\exp(\eta)}{1 + \exp(\eta)}.$$

We set up the same predictor as for the μ-parameter in the Gaussian models using again independent and normally distributed random effects α_c with 20 clusters:

$$\eta = (1 + \alpha_c)\, f(x),$$

with f being the sine function $\sin(x)$ in the interval $[-\pi, \pi]$. We again vary the number of observations per cluster as well as the variance of the random effects according to Table 1. So the effects $(1+\alpha_c)f(x)$ are the same as for the μ-parameter in the Gaussian models, see Figure 1 (left column) for illustration.

Results

We again generate 250 replications of the six models and do the same MCMC runs as for the Gaussian models. As we can see from Figures 10 and 11, showing the sampling paths

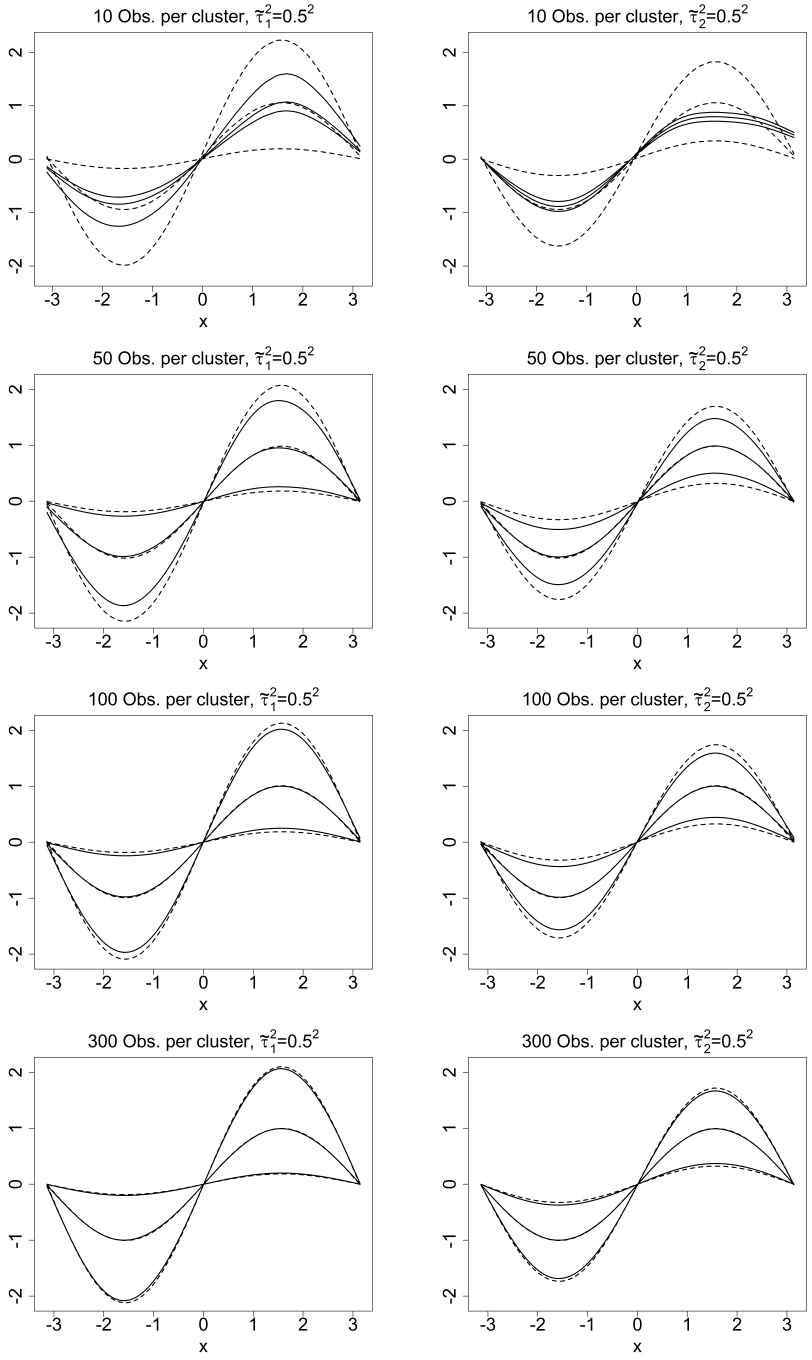

Figure 8: The average estimates of the functions f_l as well as of the smallest and largest cluster-specific effects $(1 + \alpha_{cl})\, f_l(x)$ (solid) and the respective true effects (dashed) for the Gamma models 1 to 4. The first column shows the effects of the μ-equation, the second column those of the σ-equation.

of the random effects α_c of the first cluster in one of the replications of model 2 and the corresponding autocorrelation functions, the draws are again practically independent.

Compared to the Gaussian and the Gamma models, we now need much more observations per cluster to estimate the random effects properly (see Figure 12). Even with 300 observations per cluster there is still a (small) bias. When analyzing the impact of the variance of the random effects, we again find that the estimation results get better the larger the variance is (see Figure 13).

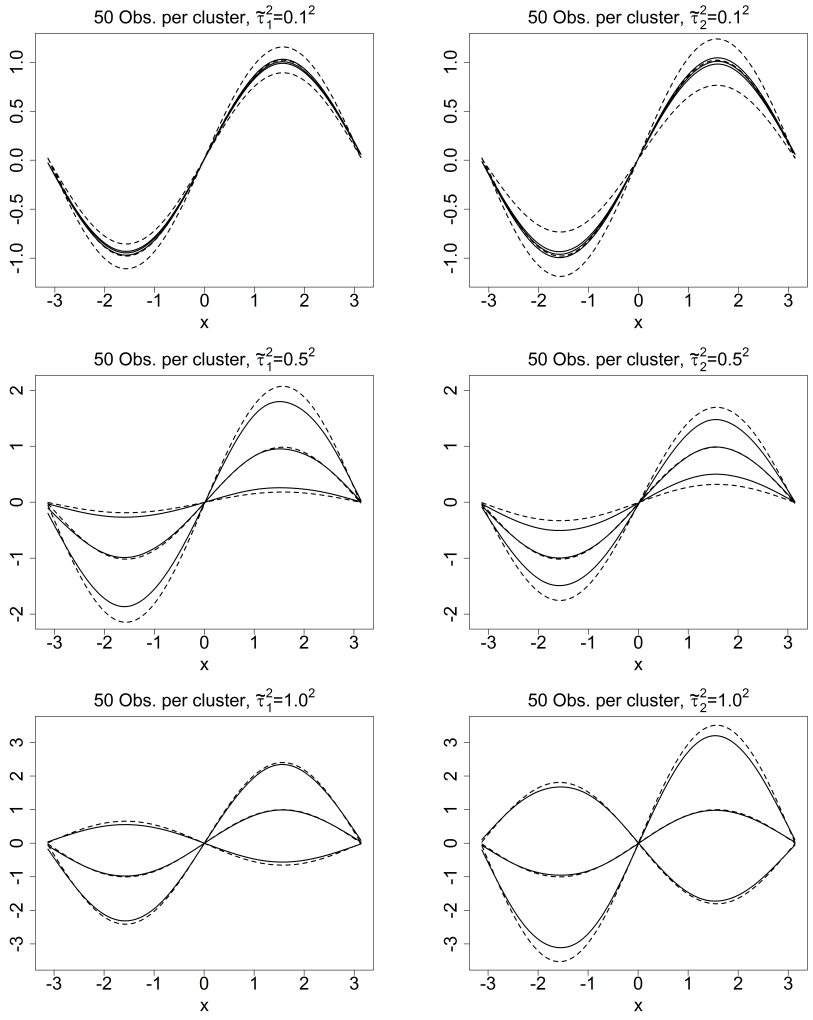

Figure 9: The average estimates of the functions f_l as well as of the smallest and largest cluster-specific effects $(1 + \alpha_{cl})\, f_l(x)$ (solid) and the respective true effects (dashed) for the Gamma models 5, 2 and 6. The first column shows the effects of the μ-equation, the second column those of the σ-equation.

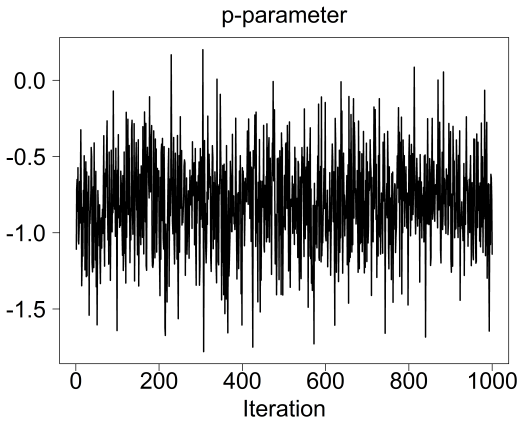

Figure 10: Sampling path of the random effect α_c of the first cluster in one of the replications of the Binomial model 2.

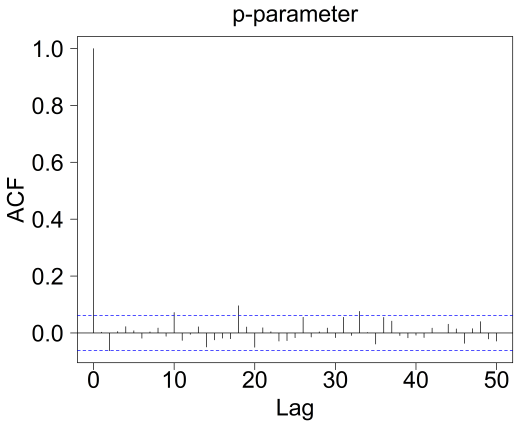

Figure 11: Autocorrelation function of the random effect α_c of the first cluster in one of the replications of the Binomial model 2.

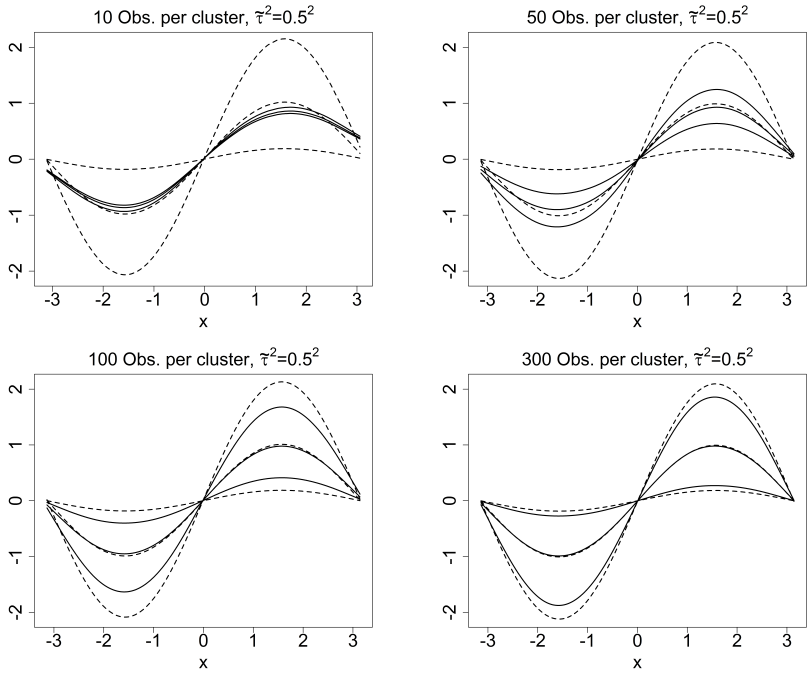

Figure 12: The average estimates of the function f as well as of the smallest and largest cluster-specific effects $(1 + \alpha_c) f(x)$ (solid) and the respective true effects (dashed) for the Binomial models 1 to 4.

3.4 Model evaluation

In order to evaluate the performance of our simultaneous estimation approach, we refer to 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 two-stage estimation procedure sketched in Section 2.3 and again calculate the corresponding MSEs.

Figure 14 shows boxplots of the MSEs of the 250 replications in the Gaussian models 1-4 with 10, 50, 100 and 300 observations per cluster and a variance of the scaling factors

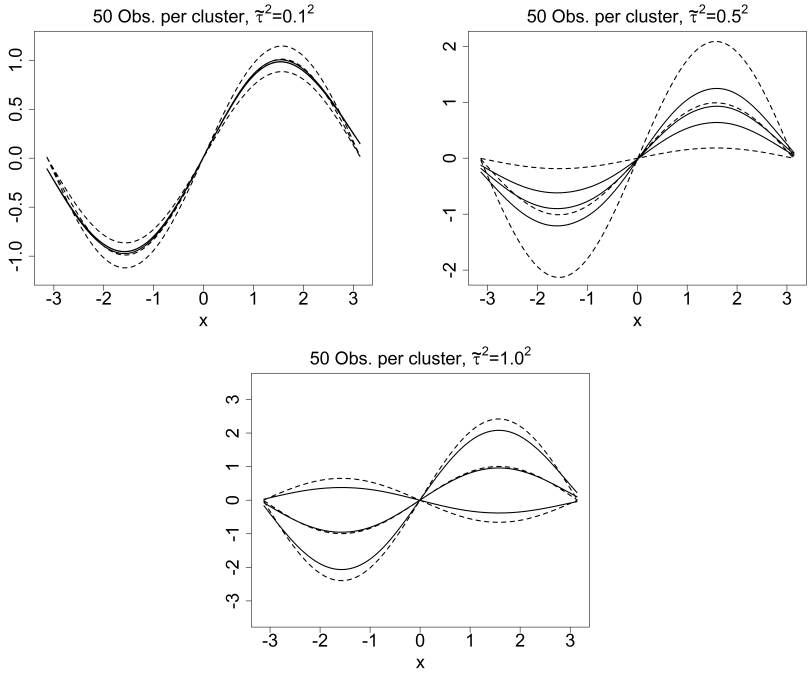

Figure 13: The average estimates of the function f as well as of the smallest and largest cluster-specific effects $(1 + \alpha_c)\, f(x)$ (solid) and the respective true effects (dashed) for the Binomial models 5, 2 and 6.

of $\tilde{\tau}_l^2 = 0.5^2$. As we can see, the simultaneous estimation approach ("1") yields lower MSEs than the two-stage procedure ("2") for all models and both parameters. Especially for the σ-parameter, the estimation results are considerably better with our simultaneous approach. The continuous decrease of the MSEs for both parameters in our simultaneous estimation approach for models with a larger number of observations per cluster corresponds to the improving estimation results that we have seen in Figure 4.

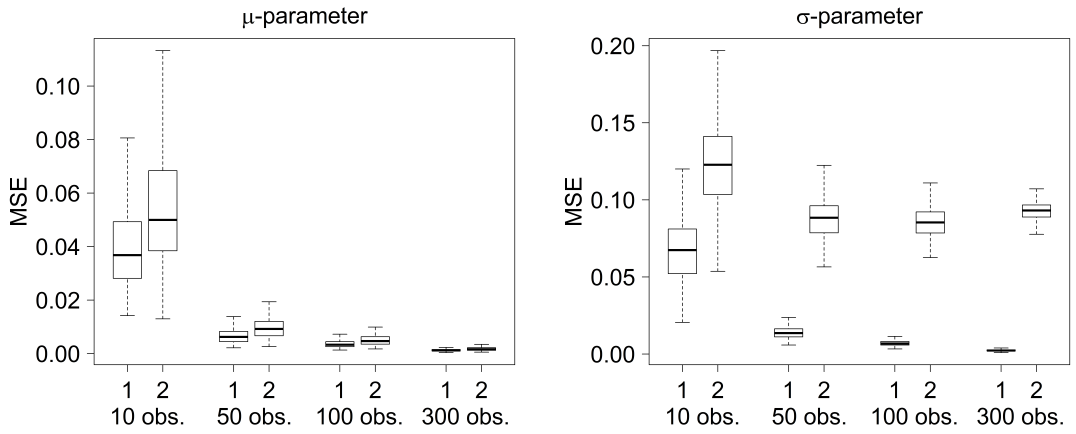

Figure 14: MSEs for the μ- and σ-parameter in the Gaussian models with $\tilde{\tau}_l^2 = 0.5^2$ from the simultaneous estimation approach ("1") and the two-stage procedure ("2").

With respect to the variance of the random scaling factors, we find that the performance of the simultaneous estimation approach compared to the two-stage procedure substantially gets better the higher the variance is, see Figure 15 that depicts boxplots of the MSEs for the Gaussian models $5, 2$ and 6 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 50 observations per cluster. Particularly for large variances (e.g. $\tilde{\tau}_l^2 = 1.0^2$) the two-stage procedure yields peculiar estimation results that cause huge MSEs.

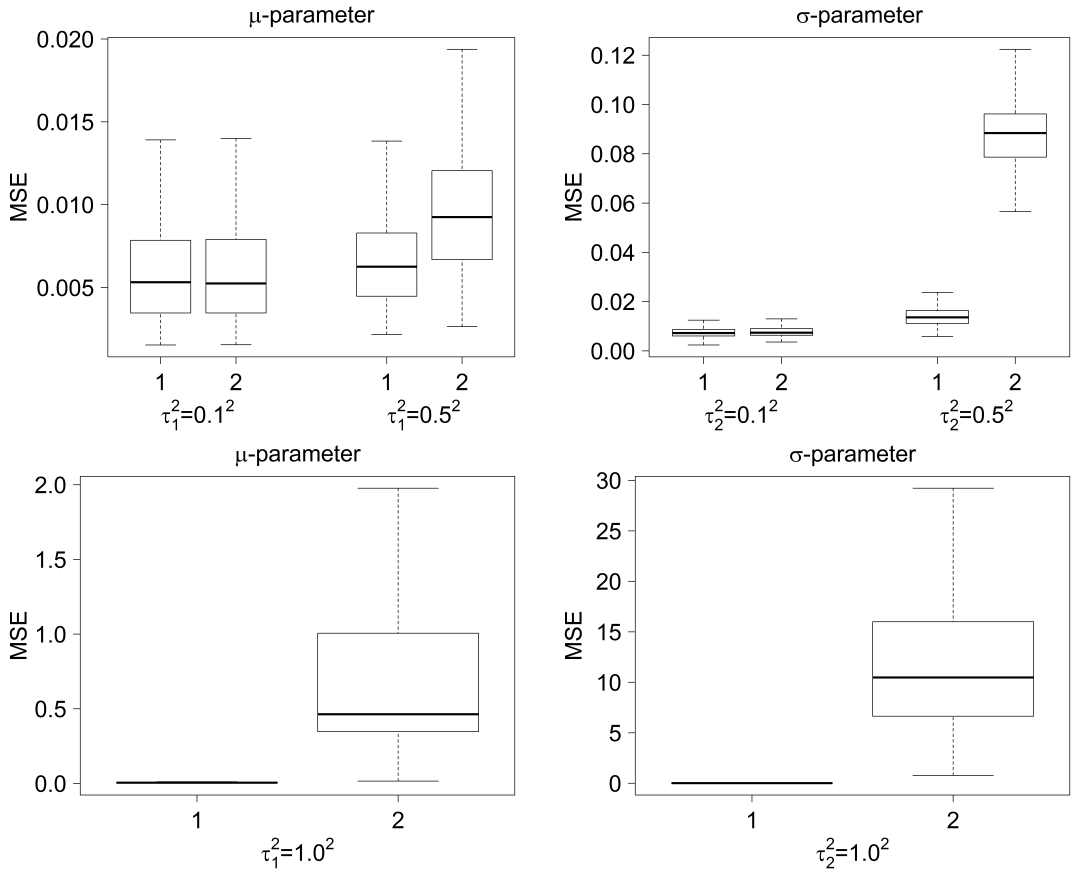

Figure 15: MSEs for the μ- and σ-parameter in the Gaussian models with 50 observations per cluster from the simultaneous estimation approach ("1") and the two-stage procedure ("2").

For the Gamma models, we are again able to reduce the MSEs for both parameters with the simultaneous estimation approach (see Figure 16 and Figure 17). However, the improvement for the σ-parameter is not as marked as in the Gaussian models. We again find that the performance of the simultaneous estimation approach compared to the two-stage procedure substantially gets better the higher the variance of the random scaling factors is, see Figure 17.

Eventually, there are only minor differences with respect to the MSE between the simultaneous estimation approach and the two-stage procedure for the p-parameter in the Binomial models, as we can see from Figure 18.

We additionally analyze if the MSE is related to the magnitude of the scaling in the respective cluster. For this purpose, we calculate the average MSE in each of the 20 clusters over all observations and all replications and plot it against the corresponding random effect (which is fixed over all replications). Figure 19 shows the resulting scatter plots for the different parameters of our three models, each with 300 observations per cluster. For all models and all parameters, we find that the MSE increases the larger the size of the scaling is.

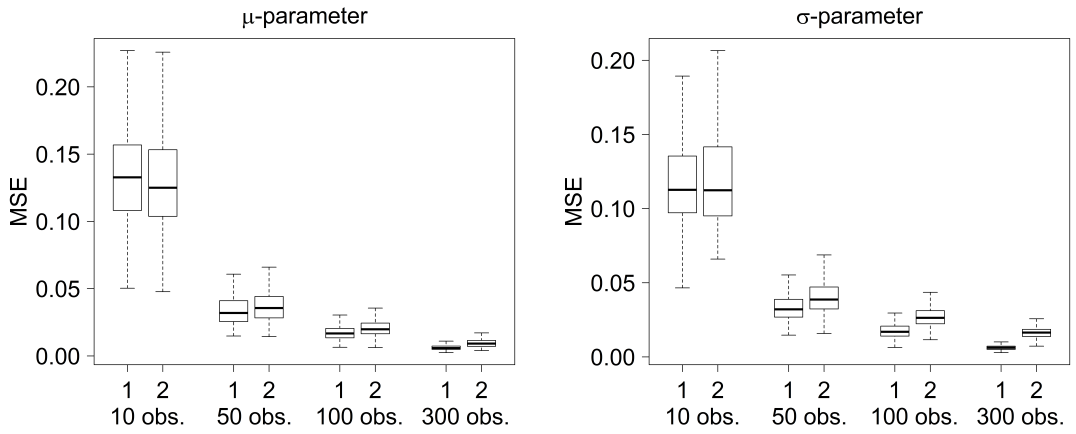

Figure 16: MSEs for the μ- and σ-parameter in the Gamma models with $\tilde{\tau}_l^2 = 0.5^2$ from the simultaneous estimation approach ("1") and the two-stage procedure ("2").

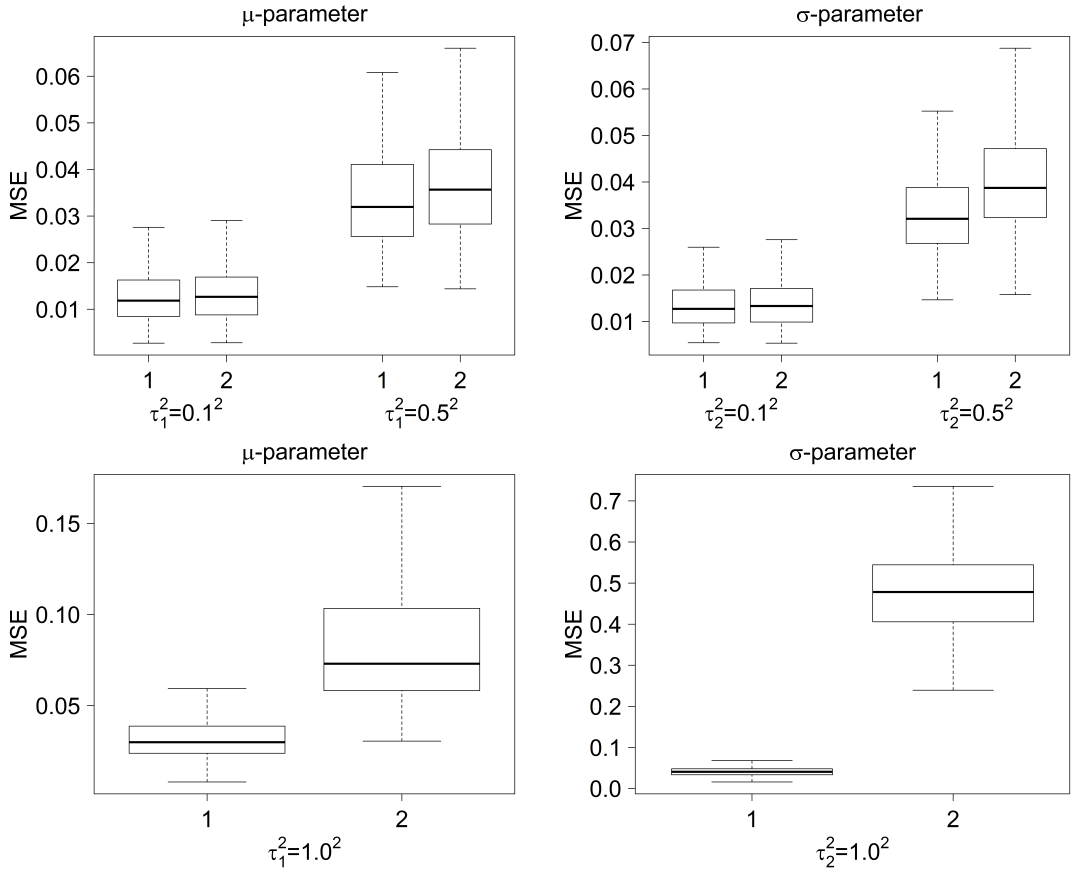

Figure 17: MSEs for the μ- and σ-parameter in the Gamma models with 50 observations per cluster from the simultaneous estimation approach ("1") and the two-stage procedure ("2").

4 Data description and model specification

We apply our methodology to a dataset of almost 100,000 single family homes all over Germany. The data was provided by F+B Research and Consulting for Habitation, Real Estate and Environment Ltd, a business consultancy in Hamburg, Germany.

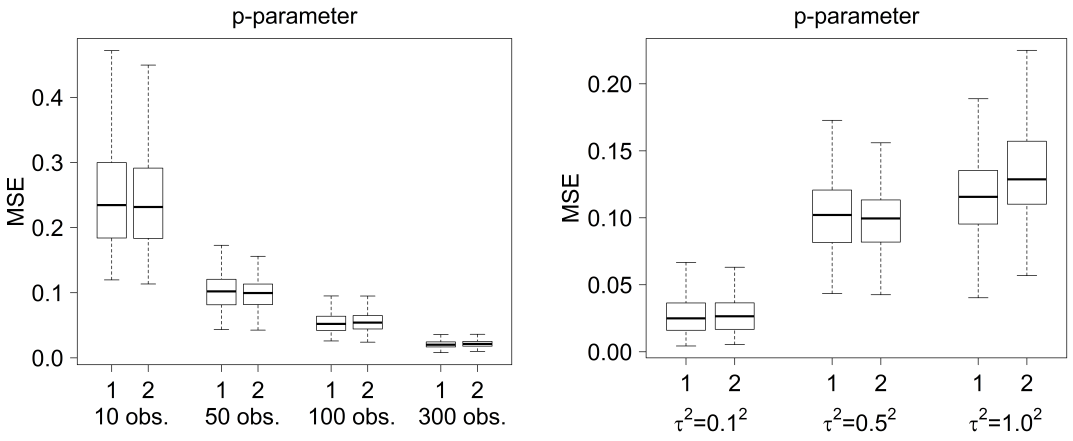

Figure 18: MSEs for the p-parameter in the Binomial models.

Since the raw data initially was supply data, prices obviously were upward biased. Thus, F+B adjusted these prices by a transaction discount estimated from a regression model in order to provide realistic purchase prices. Dividing these adjusted prices by the floor area of the houses leads to the prices per square meter (p_{qm}), which we use as the dependent variable in our analysis. The set of explanatory variables include continuous and categorical covariates that characterize the building and its location:

- Continuous covariates: The floor area of the building (*area*) is expected to have a decreasing effect on the price per square meter due to the law of diminishing marginal utility, while the plot area where the house is built on (*plot_area*) should have a positive effect. Due to depreciation over time the year of construction (*year*) in general should also have an increasing effect on house prices per square meter. Besides from these structural covariates, an expert rating (*rating*) is included that characterizes the area in which the building is situated. Since the rating ranges from 1 (excellent) to 9 (bad), we expect a negative effect on the prices.

- Categorical covariates: The equipment of the house (*equipment*) is classified by four categories. Obviously, we expect an increasing effect for better equipments. In order to control for state-specific price differences, we include dummy variables for the states in which the buildings are located in.

As the most basic model we set up a Gaussian model with parameters μ and σ, which we link to predictors η_1 and η_2 via

$$\begin{aligned} \mu &= \eta_1, \\ \sigma &= \exp(\eta_2). \end{aligned}$$

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

$$\begin{aligned} \boldsymbol{\eta_l} \;=\; & \boldsymbol{D}_{1l}\boldsymbol{f}_{1l}(area) + \boldsymbol{D}_{2l}\boldsymbol{f}_{2l}(plot\_area) \\ & + \boldsymbol{D}_{3l}\boldsymbol{f}_{3l}(year) + \boldsymbol{f}_{4l}(rating) + \boldsymbol{X}\boldsymbol{\gamma}_l. \end{aligned} \tag{6}$$

$\boldsymbol{f}_{1l}, \ldots, \boldsymbol{f}_{4l}$ are possibly nonlinear functions of the continuous covariates and will be modeled with P-splines. $\boldsymbol{D}_{jl} = \text{diag}(1 + \alpha_{jc_1l}, \ldots, 1 + \alpha_{jc_nl})$, $c_i \in \{1, \ldots, C\}$, contain random scaling factors for the C districts that allow for regional heterogeneity in the respective

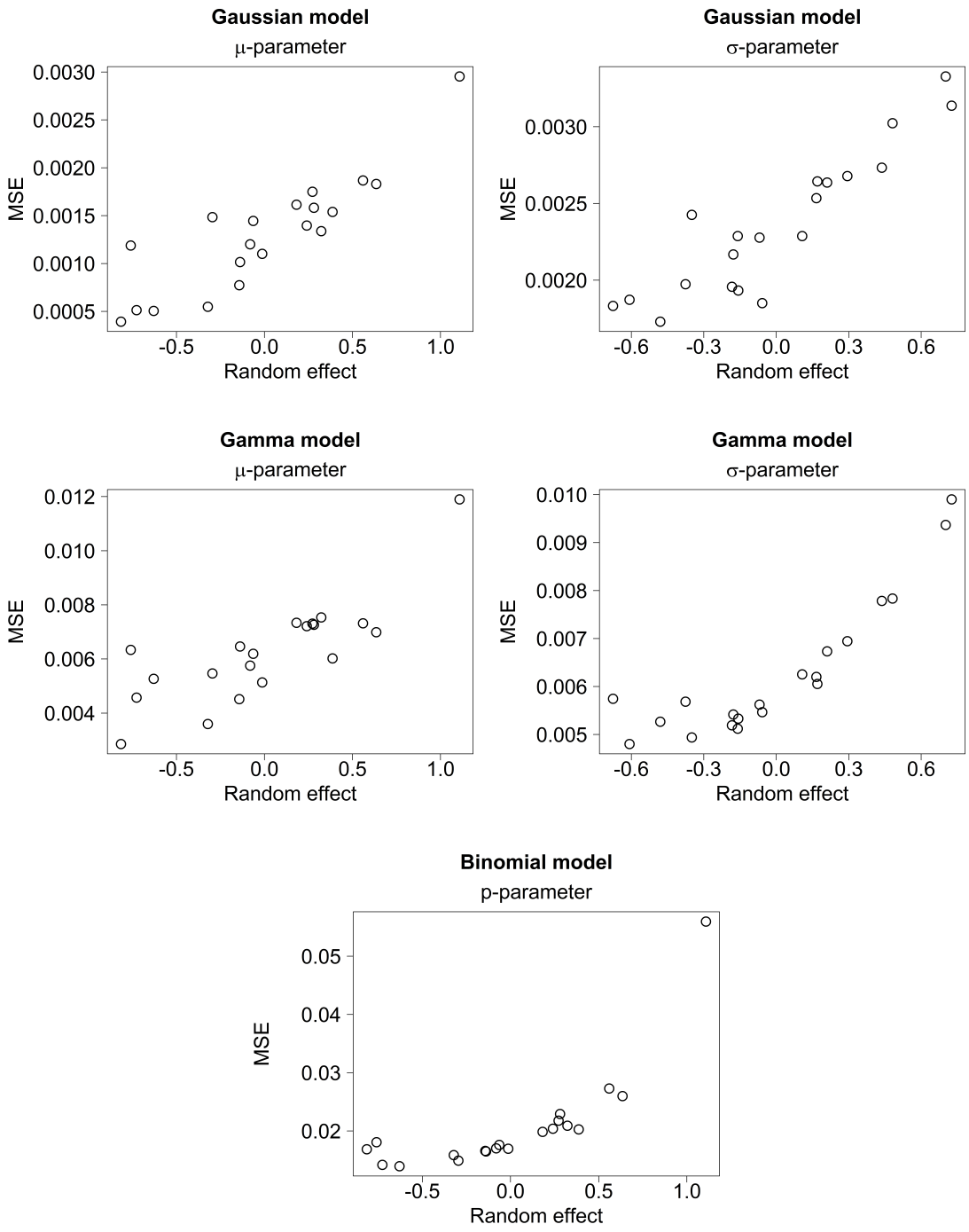

Figure 19: MSE in dependence of the size of the random effect.

price response functions. Here, we particularly expect spatial differences between Eastern and Western Germany. 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 $\boldsymbol{\gamma}_l$.

Since house prices typically are skewed we additionally set up a Loggaussian model (with location parameter μ and scale parameter σ) as well as a Gamma model (with mean parameter μ and shape parameter σ). In both cases, we link these parameters to predictors

η_1 and η_2, which again are constructed according to (6), via

$$\begin{aligned}\mu &= \eta_1, \\ \sigma &= \exp(\eta_2),\end{aligned}$$

in the Loggaussian model and via

$$\begin{aligned}\mu &= \exp(\eta_1), \\ \sigma &= \exp(\eta_2),\end{aligned}$$

in the Gamma model.

5 Results

The estimation results for the models described in the previous section 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 draws from the posterior. Comprehensive MCMC diagnostics show that they are practically independent. For illustration, Figures 20 and 21 show the sampling paths of the random effects α_{cl} of the floor area of one of the districts and the corresponding autocorrelation functions for both the μ- and the σ-parameter in the Gamma model (which turned out to be the best model, see Section 5.2).

When analyzing the results, one has to be aware that the parameters of the distributions do not exactly correspond to each other. For example, the mean of the Gaussian and the Gamma model immediately is given by the respective μ-parameter, while it is given by $\exp(\mu+\sigma^2/2)$ in the Loggaussian model. Therefore, a simple comparison of the parameters (or of the involved scaling factors) is not reasonable. Instead, we always derive the mean effects of the three models and compare them to each other. However, we additionally present the results of the individual scaling factors at least for the best model (the Gamma model).

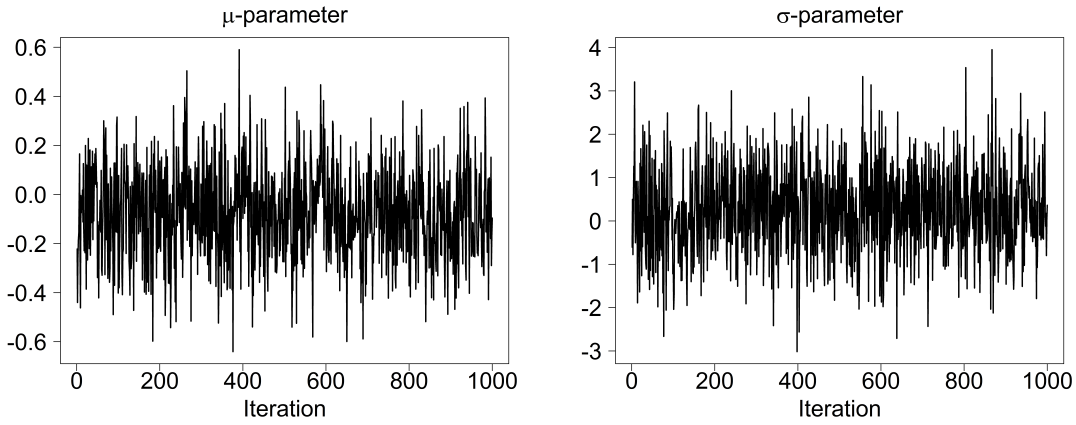

Figure 20: Sampling paths of the random effects α_{cl} of the floor area of one district in the Gamma model.

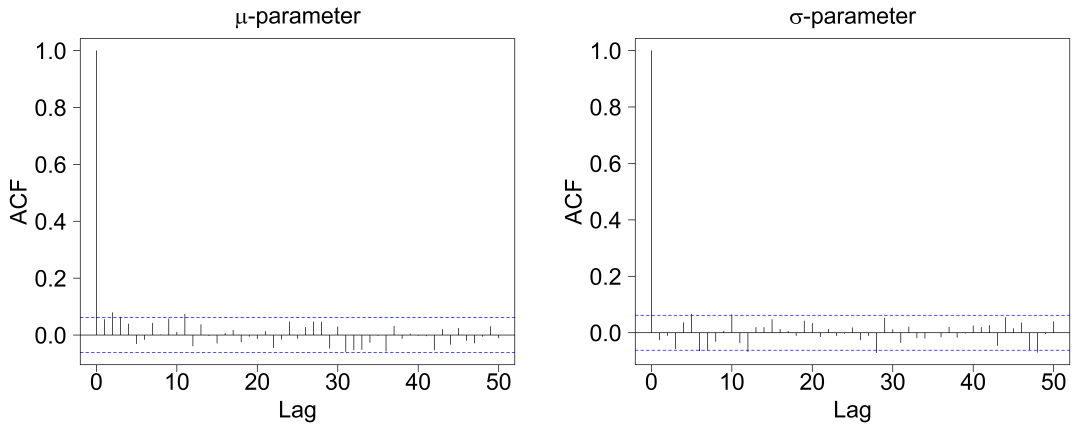

Figure 21: Autocorrelation functions of the random effects α_{cl} of the floor area of one district in the Gamma model.

5.1 Effect estimates

For the sake of illustration, we restrict the presentation of the results to the covariates being multiplied by random scaling factors, i.e. the floor area, the plot area and the year of construction.

As expected, the mean effect of the floor area on house prices per square meter, averaged over all districts, is monotonically decreasing in all our models (see panel [a] of Figure 22). Here, in order to get an impression of the magnitude of the effects and to make the results comparable, the other continuous covariates are held constant at mean level of attributes and the categorical variables are held at their mode level (which we will call the *mean level*). As we can see, the results of the Loggaussian and the Gamma model almost coincide, while the effect estimated by the Gaussian model is consistently lower.

With respect to the scaling, we find that the effect of the floor area considerably differs between the districts. Panels [b] to [d] of Figure 22 show the scaled effects of the different districts for the three models together with the respective average effect. In the Gamma model, for example, the effect of the floor area accounts for a variation between 350 Euro per square meter in the district with the smallest scaling factor and $1,200$ Euro per square meter in the district with the largest scaling factor.

For illustration, we now have a closer look to the scaling factors of the floor area in the Gamma model. According to Table 2, the scaling factors of the mean parameter range from 0.49 to 1.60, those of the shape parameter go from -2.16 to 3.04.

Parameter	Min	Mean	Max
μ	0.49	1.00	1.60
σ	-2.16	1.00	3.04

Table 2: *Random scaling factors of the floor area in the Gamma model*

With respect to the geographic distribution, the scaling factors of the mean parameter tend to be higher in Western Germany, while they are considerably lower in Eastern Germany (panel [a] of Figure 23). In order to verify the significance of these differences, we refer to the posterior probabilities based on a nominal level of 80%. If the 80% credible interval of a random effect α is strictly positive or negative we assign the corresponding

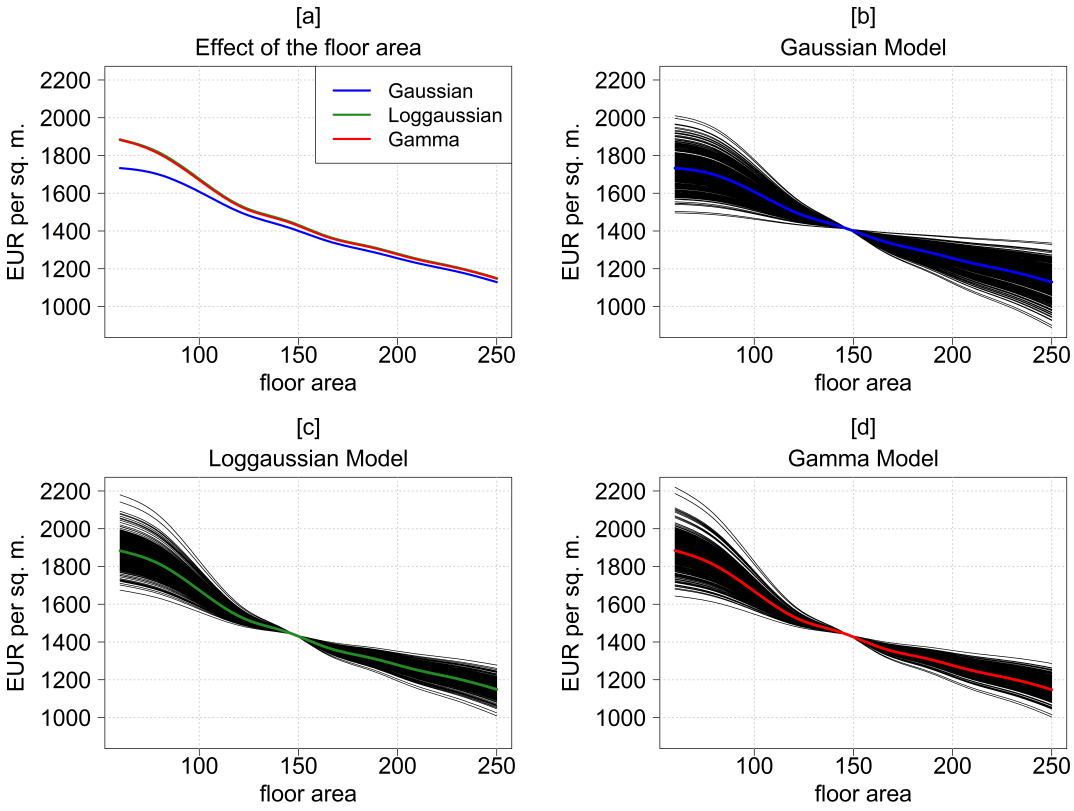

Figure 22: Mean effect of the floor area evaluated at the mean level. [a]: Average effect over all districts for the three different models. [b] – [d]: Scaled effects of the individual districts for the three models together with the respective average effect.

district a value of 1 or -1, respectively. If the credible interval contains zero, we record a value of 0 for this district. As panel [b] of Figure 23 shows, the scaled effects of the floor area significantly differ from the average effect in about one fourth of the districts.

For the shape parameter σ, the marginal effect of the floor area, averaged over all districts, is increasing up to an area of about 120 square meters and then levels off (see Figure 24). However, as can be seen from the simultaneous 95% confidence bands, the effect is not significant for a wide range of the floor area. Accordingly, the majority of the respective scaling factors are not significant either, nor do we find a clear geographic pattern (see Figure 25).

For the plot area, the average mean effect over all districts is monotonically increasing up to an area of $1,900$ square meters and then slightly reverses for all three models (panel [a] of Figure 26). However, this minor decrease for very large plots is not significant according to the simultaneous 95% confidence bands (not depicted in the figure).

There is again considerable variation in the effects for the different districs (panels [b] – [d] of Figure 26). In the Gamma model, for example, the respective scaling factors of the mean parameter range from -0.08 to 2.05, see Table 3.

Parameter	Min	Mean	Max
μ	-0.08	1.00	2.05
σ	-0.50	1.00	3.30

Table 3: *Random scaling factors of the plot area in the Gamma model*

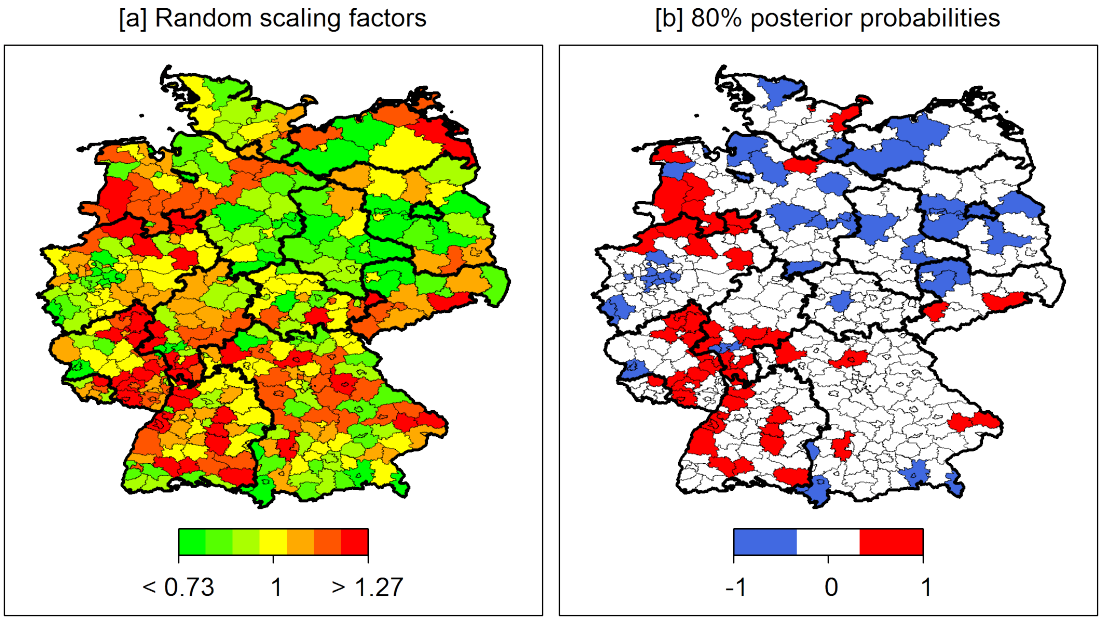

Figure 23: Panel [a]: Random scaling factors of the floor area for the mean parameter in the Gamma model. [b]: 80% posterior probabilities.

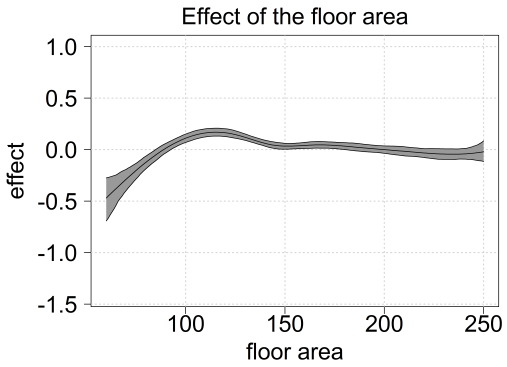

Figure 24: Average marginal effect of the floor area for the shape parameter in the Gamma model together with simultaneous 95% confidence bands.

According to these scaling factors, the magnitude of the mean effect of the plot area is much more pronounced in the southwestern states as well as in the surroundings of Berlin and Hamburg than in the remaining parts of Germany (panel [a] of Figure 27). This coincides with the regions that have a higher population density and where land therefore is a scarcer resource. The deviation from the average effect is significant in about half of the districts, see panel [b] of Figure 27.

For the shape parameter of the Gamma model, the average marginal effect of the plot area is slightly decreasing, see Figure 28. Again, there are considerable differences in the scaling of the effect over the districts with the corresponding scaling factors ranging from -0.50 to 3.30.

The largest scaling factors can be found in the most western states of Germany (panel [a] of Figure 29). However, especially in the eastern half of Germany, there are hardly any districts where the scaled effects significantly differ from the average effect, see panel [b] of Figure 29.

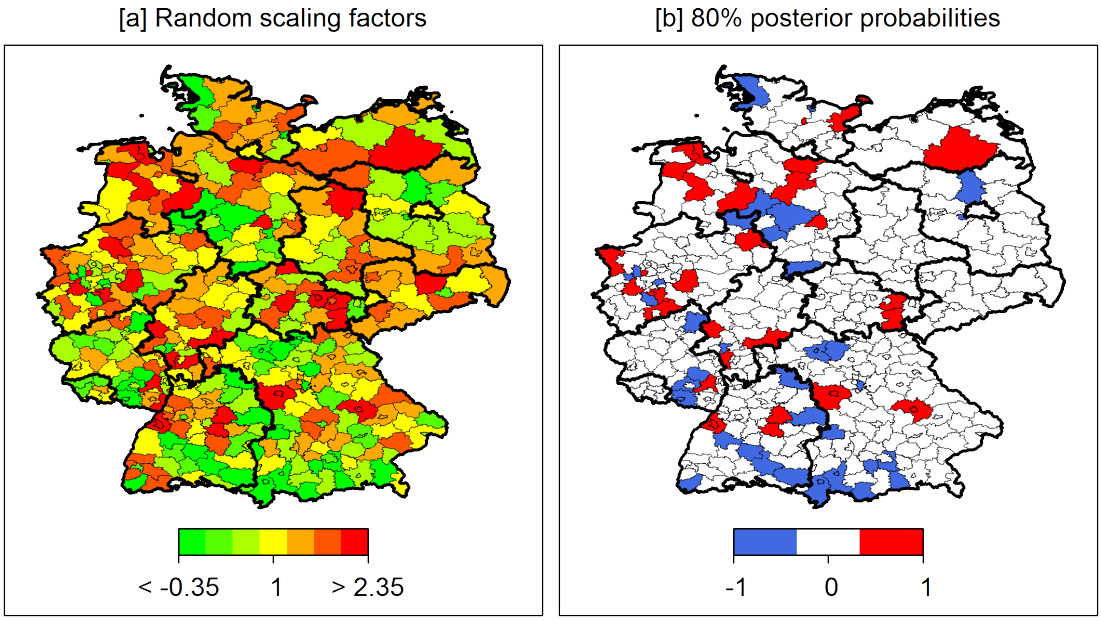

Figure 25: Panel [a]: Random scaling factors of the floor area for the shape parameter in the Gamma model. [b]: 80% posterior probabilities.

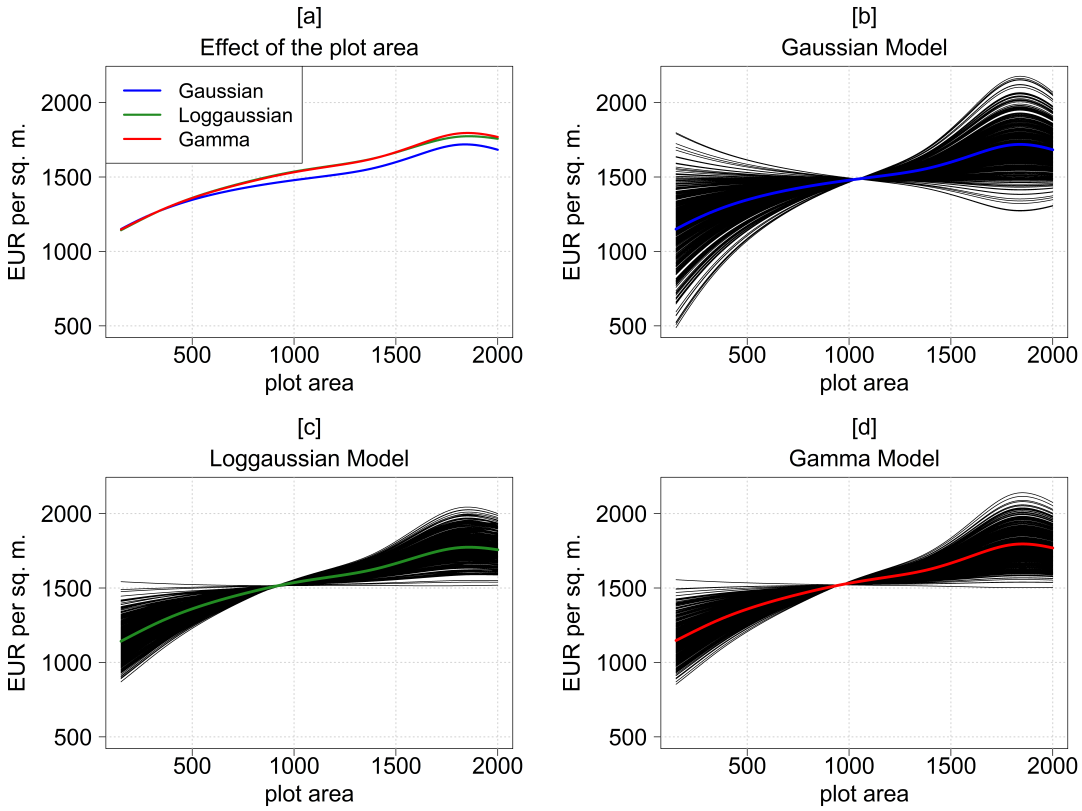

Figure 26: Mean effect of the plot area evaluated at the mean level. [a]: Average effect over all districts for the three different models. [b] – [d]: Scaled effects of the individual districts for the three models together with the respective average effect.

In general, the results for the year of construction show an increasing effect on the mean of the house prices per square meter (panel [a] of Figure 30). However, there are three

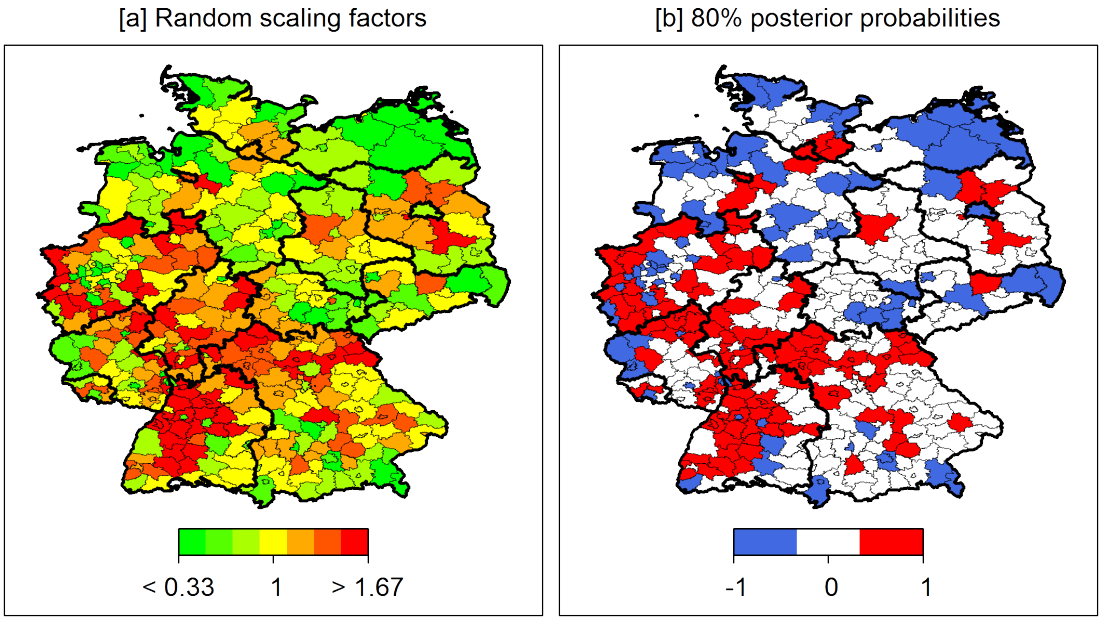

Figure 27: Panel [a]: Random scaling factors of the plot area for the mean parameter in the Gamma model. [b]: 80% posterior probabilities.

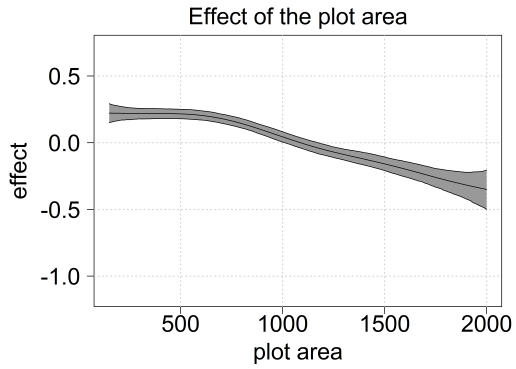

Figure 28: Average marginal effect of the plot area for the shape parameter in the Gamma model together with simultaneous 95% confidence bands.

periods where the effect is negative: during and shortly after World War One construction naturally was cheaper. The same holds for the time of World War Two. The third period showing a negative effect is the time after the latest financial crisis when the real estate market in Germany was in trouble. Indeed, a recovery started in recent years, but this is not yet covered by our data ending in 2012. Again, the scaling factors, which for the Gamma model vary between 0.14 and 1.66 (see Table 4), reveal considerable spatial heterogeneity in the magnitude of the effects (panels [b] – [d] of Figure 30).

Parameter	Min	Mean	Max
μ	0.14	1.00	1.66
σ	-0.10	1.00	2.38

Table 4: *Random scaling factors of the year of construction in the Gamma model*

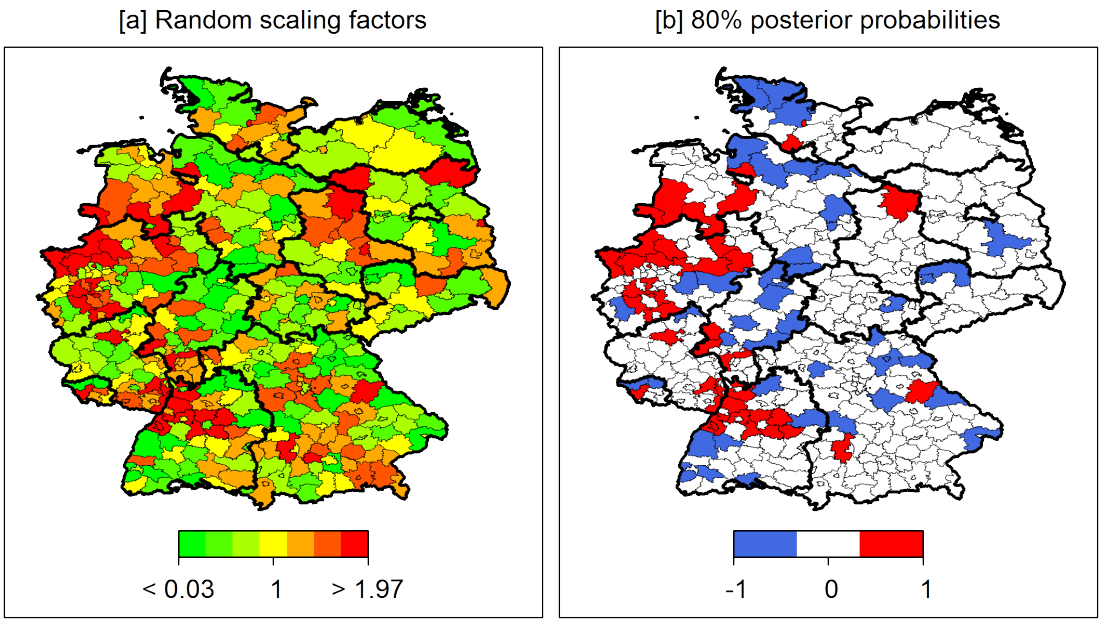

Figure 29: Panel [a]: Random scaling factors of the plot area for the shape parameter in the Gamma model. [b]: 80% posterior probabilities.

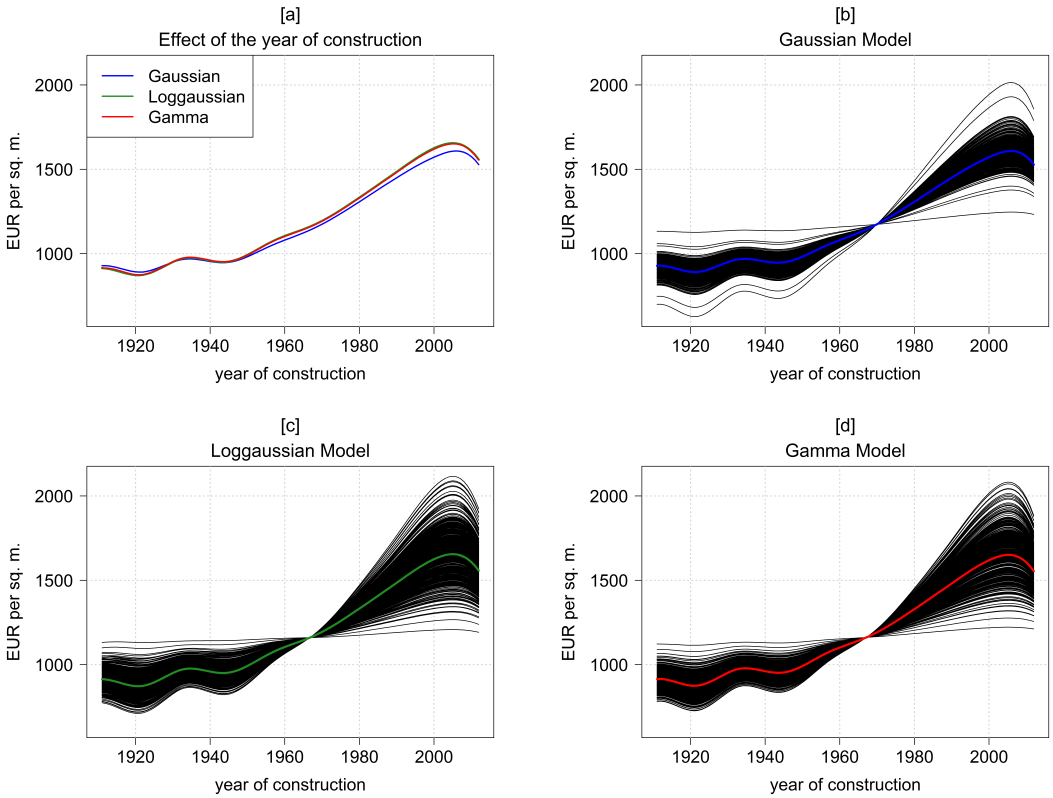

Figure 30: Mean effect of the year of construction evaluated at the mean level. [a]: Average effect over all districts for the three different models. [b] – [d]: Scaled effects of the individual districts for the three models together with the respective average effect.

With the exception of the surrounding of Berlin, the mean effect of the year of construction is more pronounced in Eastern Germany than in most parts of Western Germany, see panel

[a] of Figure 31. The main reason is that the construction industry in Eastern Germany was much more affected by the reunification of 1990, causing at first a huge boom in the eastern states followed by a severe downturn afterwards. Thus, the year of construction has more influence on house prices in Eastern than in Western Germany. Overall, the scaled effects significantly differ from the average effect in about two-thirds of the districts (panel [b] of Figure 31).

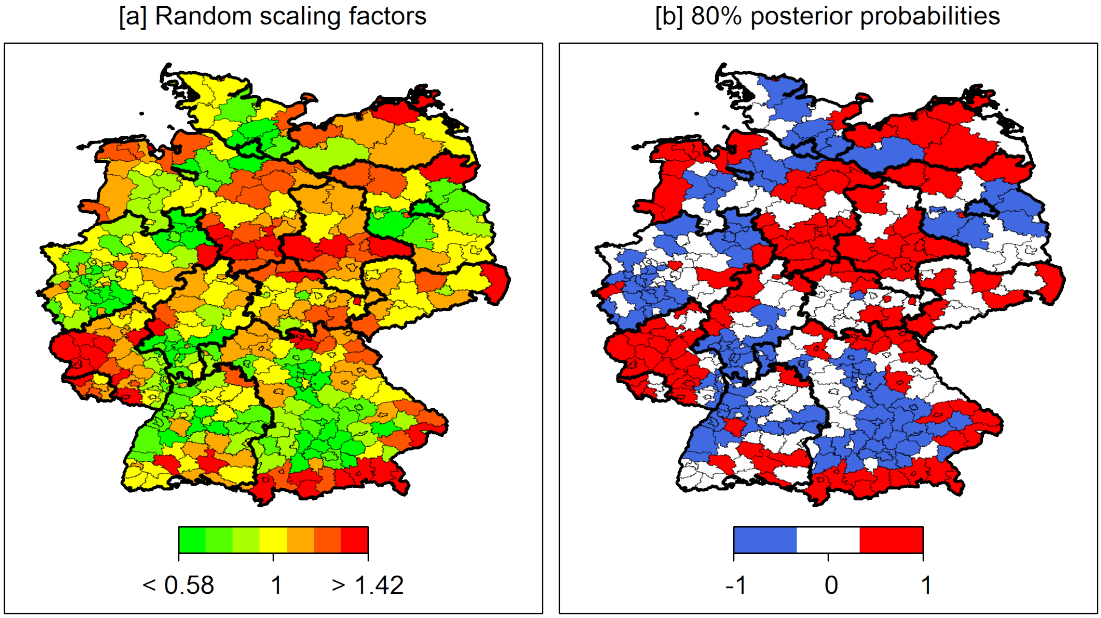

Figure 31: Panel [a]: Random scaling factors of the year of construction for the mean parameter in the Gamma model. [b]: 80% posterior probabilities.

Except for the last few years, the effect of the year of construction for the shape parameter in the Gamma model looks similar to the one for the mean parameter (Figure 32). The corresponding scaling factors range from -0.10 to 2.38, see Table 4.

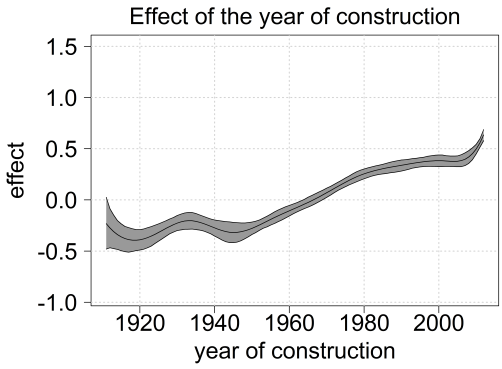

Figure 32: Average marginal effect of the year of construction for the shape parameter in the Gamma model together with simultaneous 95% confidence bands.

The effect is more pronounced in Eastern Germany than in the western parts, see panel [a] of Figure 33, with significant differences from the average effect in at least one third of the districts (panel [b] of Figure 33).

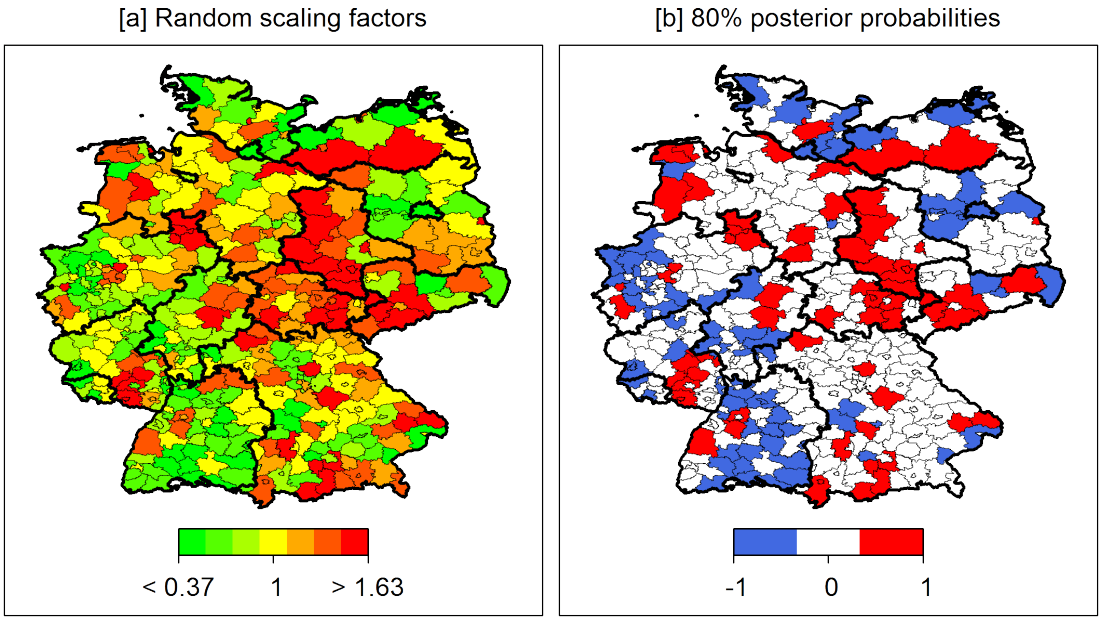

Figure 33: Panel [a]: Random scaling factors of the year of construction for the shape parameter in the Gamma model. [b]: 80% posterior probabilities.

5.2 Model comparison

First of all, we want to compare our models to standard Gaussian, Loggaussian and Gamma models without scaling factors. In such models, the predictors for the respective μ- and σ-parameters are given by

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

For illustration, Figure 34 shows the marginal mean effects of the standard Gamma model (solid lines). For comparison, the average effects of the extended Gamma model with scaling factors are displayed as well (dotted lines). We see that for all covariates the estimated effects of the two models are very similar.

In order to evaluate the performance of the models we refer to the deviance information criterion (DIC) of Spiegelhalter et al. (2002), which takes into account both the fit of the data and the model complexity. If we denote by $\boldsymbol{\theta}$ the vector of model parameters and by $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(T)}$ an MCMC sample from the posterior distribution of model parameters, then the deviance of a model with response $\boldsymbol{y}$ is given by

$$D(\boldsymbol{\theta}) = -2 \cdot \log\left(p(\boldsymbol{y}|\boldsymbol{\theta})\right).$$

Furthermore, the effective number of parameters in the model, p_D, is given by

$$p_D = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}})$$

where

$$\overline{D(\boldsymbol{\theta})} = \frac{1}{T}\sum_{t=1}^{T} D\left(\boldsymbol{\theta}^{(t)}\right) \quad \text{and} \quad \bar{\boldsymbol{\theta}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\theta}^{(t)}.$$

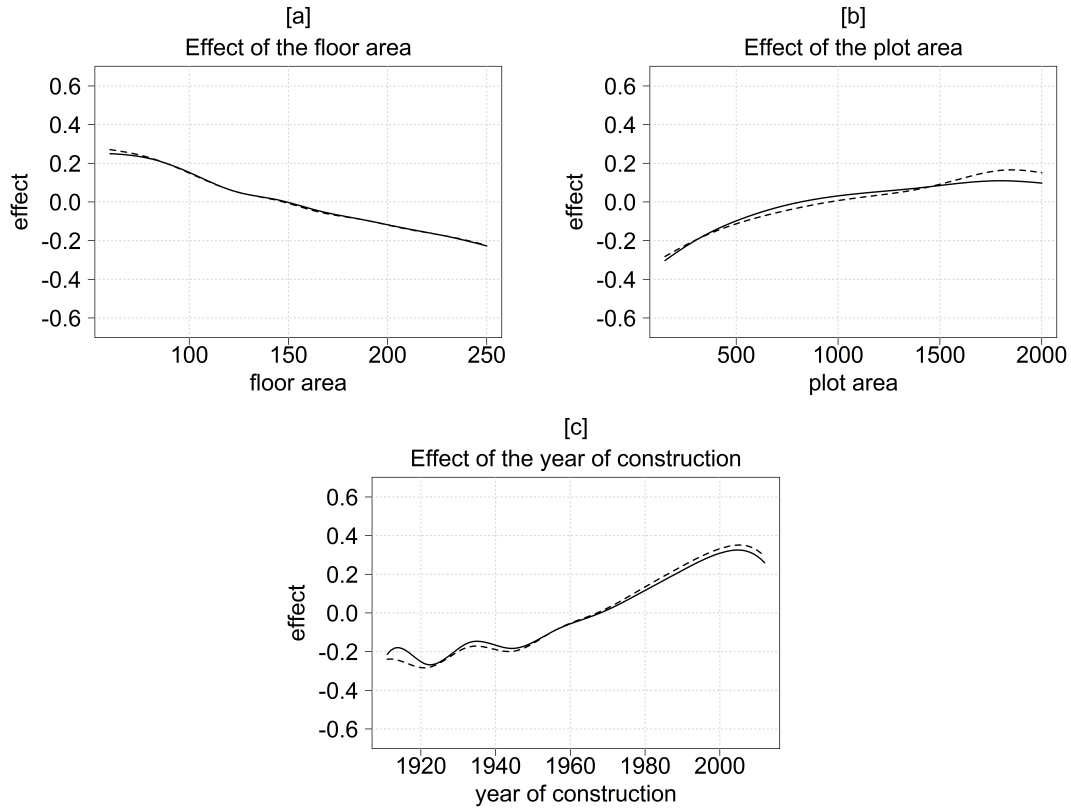

Figure 34: Mean effects of the standard Gamma model (solid) together with average effects of the extended Gamma model with scaling factors (dotted). [a]: Effect of the floor area. [b]: Effect of the plot area. [c]: Effect of the year of construction.

Then, the DIC is defined as

$$\text{DIC} = \overline{D(\boldsymbol{\theta})} + 2 \cdot p_D = 2 \cdot \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}).$$

Models with a lower DIC are superior compared to models with a higher DIC, where differences of 10 or more usually are considered to be significant. As we can see from Table 5, the DIC of the models with scaling factors are by far lower than the DIC of the standard models, showing that the inclusion of random scaling factors leads to a significant improvement of the results.

DIC	Gaussian	Loggaussian	Gamma
Standard model	68,267	62,543	62,878
Model with scaling factors	56,013	50,886	50,753

Table 5: *DIC of the standard models and the extended models with random scaling factors*

Furthermore, Klein et al. (2015) have shown that the DIC also can be used to discriminate between different types of response distributions in distributional regression. Thus, the previous results additionally suggest that within the models with scaling factors the Gamma model is the best one. In order to verify this finding, we further calculate the scores proposed by Gneiting and Raftery (2007), which are suited to compare the predictive ability of parametric models in terms of probabilistic forecasts based on the predictive distribution of the actual realizations.

In order to evaluate the scores for our three models we do a five-fold cross validation, i.e. we randomly divide the data set into five subsets $\Omega_1, \ldots, \Omega_5$ of virtually equal size and estimate the models based on four of those subsets. For the remaining subset, without loss of generality $\Omega_1 = \{y_1, \ldots, y_R\}$, we derive the predictive distributions with densities $p_1, \ldots, p_R$ based on the predictive parameters $\mu_1, \ldots, \mu_R$ and $\sigma_1, \ldots, \sigma_R$. A proper scoring rule S then leads to a score S_{Ω_1} for this subset by summing up the individual contributions

$$S_{\Omega_1} = \frac{1}{R} \sum_{r=1}^{R} S(p_r, y_r).$$

The conclusive score $\mathcal{S}$ is then given by the average score of the five subsets

$$\mathcal{S} = \frac{1}{5} \sum_{i=1}^{5} S_{\Omega_i}.$$

Following Gneiting and Raftery (2007), we consider the logarithmic score (LogS), the quadratic score (QuadS) and the spherical score (SpherS), which are defined by

$$\begin{aligned}
\text{LogS}(p_r, y_r) &= \log\left(p_r(y_r)\right), \\
\text{QuadS}(p_r, y_r) &= 2p_r(y_r) - \int p_r(\omega)^2 \mathrm{d}\omega, \\
\text{SpherS}(p_r, y_r) &= \frac{p_r(y_r)}{\left(\int p_r(\omega)^2 \mathrm{d}\omega\right)^{1/2}},
\end{aligned}$$

as well as the continuous ranked probability score (CRPS)

$$\text{CRPS}(p_r, y_r) = -\int_{-\infty}^{\infty} \left(F_r(x) - \mathbb{1}_{\{x \geqslant y_r\}}\right)^2 \mathrm{d}x,$$

with predictive cumulative distribution function $F_r(x) = \int_{-\infty}^{x} p_r(u)\mathrm{d}u$. Since all of these scores are proper, higher scores correspond to better probabilistic forecasts when comparing different models.

As we can see from Table 6, the scores of the Gaussian model are consistently lower than those of the Loggaussian and the Gamma model. The latter are almost identical, reflecting the similar results that we have seen for these two models. Strictly speaking, however, the scores slightly favor the Gamma model, confirming the results of the DIC.

Model	Logarithmic score	Quadratic score	Spherical score	CRPS
Gaussian	-0.2931	0.9594	0.9680	-0.1901
Loggaussian	-0.2622	0.9754	0.9754	-0.1886
Gamma	**-0.2621**	**0.9759**	**0.9759**	**-0.1885**

Table 6: *Comparison of average score contributions of the three models*

6 Conclusion

This paper presents a simultaneous estimation approach for Bayesian distributional regression models with random scaling factors and provides a comprehensive simulation study showing the accuracy of this approach. A comparison to an ordinary two-stage estimation procedure reveals considerable improvement particularly for models where the response distribution depends on more than one parameter as well as for models where the variance of the random scaling factors is large.

We apply our methodology to real estate data from Germany and identify district-specific random scaling factors that are significant in up to two-thirds of the districts. The results confirm expected spatial heterogeneity in the covariates' effects and provide new insights into the valuation of house prices. Furthermore, allowing for district-specific random scaling factors significantly improves the performance of the models with respect to their DIC.

The spatial structure in the scaling factors that we found for the different covariates suggest a conceivable starting point for further research we are already working on. Instead of using spatially uncorrelated random scaling factors we plan to introduce correlated factors. Another direction for further research could be an automated variable selection for random scaling factors, which would be of interest especially for more complex models.

References

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

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.

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

Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

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

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

Hastie, T. J. and R. J. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society B 55*, 757–796.

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.

Lang, S., W. Steiner, A. Weber, and P. Wechselberger (2015). Accommodating heterogeneity and nonlinearity in price effects for predicting brand sales and profits. *European JournalofOperationalResearch 246*, 232–241.

Lang, S., N. Umlauf, P. Wechselberger, K. Harttgen, and T. Kneib (2014). Multilevel structured additive regression. *Statistics and Computing 24*, 223–238.

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

Rigby, R. A. and D. M. Stasinopoulos (2005). Generalized Additive Models for Location, Scale and Shape. *Applied Statistics 54*, 507–554.

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 (2015). A comparison of semiparametric and heterogeneous store sales models for optimal category pricing. Technical report.

Wechselberger, P., S. Lang, and W. J. Steiner (2008). Additive Models with Random Scaling Factors: Applications to Modeling Price Response Functions. *Austrian Journal of Statistics 37*, 255–270.

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

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

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

2016-30 **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

2016-30

Alexander Razen, Stefan Lang

Random scaling factors in Bayesian distributional regression models with an application to real estate data

Abstract
Distributional structured additive regression provides a flexible framework for modeling each parameter of a potentially complex response distribution in dependence of covariates. Structured additive predictors allow for an additive decomposition of covariate effects with nonlinear effects and time trends, unit- or cluster-specific heterogeneity, spatial heterogeneity and complex interactions between covariates of different type. Within this framework, we present a simultaneous estimation approach for multiplicative random effects that allow for cluster-specific heterogeneity with respect to the scaling of a covariate's effect. More specifically, a possibly nonlinear function $f(z)$ of a covariate z may be scaled by a multiplicative cluster-specific random effect $(1+\alpha_c)$. Inference is fully Bayesian and is based on highly efficient Markov Chain Monte Carlo (MCMC) algorithms. We investigate the statistical properties of our approach within extensive simulation experiments for different response distributions. Furthermore, we apply the methodology to German real estate data where we identify significant district-specific scaling factors. According to the deviance information criterion, the models incorporating these factors perform significantly better than standard models without random scaling factors.

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