Bayesian generalized additive mixed models. A simulation study

Stefan Lang and Ludwig Fahrmeir
University of Munich, Ludwigstr. 33, 80539 Munich
email:lang@stat.uni-muenchen.de and fahrmeir@stat.uni-muenchen.de

Abstract

Generalized additive mixed models extend the common parametric predictor of generalized linear models by adding unknown smooth functions of different types of covariates as well as random effects. From a Bayesian viewpoint, all effects as well as smoothing parameters are random. Assigning appropriate priors, posterior inference can be based on Markov chain Monte Carlo techniques within a unified framework.

Given observations on the response and on covariates, questions like the following arise: Can the additive structure be recovered? How well are unknown functions and effects estimated? Is it possible to discriminate between different types of random effects?

The aim of this paper is to obtain some answers to such questions through a careful simulation study. Thereby, we focus on models for Gaussian and categorical responses based on smoothness priors as in Fahrmeir and Lang (2001a, b). The result of the study provides valuable insight into the facilities and limitations of the models when applying them to real data.

1 Introduction

Generalized additive mixed models (GAMMs) provide a broad and flexible framework for regression analyses in realistically complex situations with cross-sectional, longitudinal and spatial data. They extend the class of generalized linear models by adding unknown smooth functions of metrical covariates, time scales and spatial covariates as well as random effects to the common linear fixed effects part of the predictor. In fully Bayesian GAMMs, all these effects as well as smoothing parameters or other hyperparameters are considered as random, and specific models are obtained by assigning appropriate priors to them.

The two main concepts in the rapid development of Bayesian non- and semiparametric function estimation are based on (i) adaptive basis function approaches and (ii) smoothness priors. In basis function approaches, priors are designed to control the significance of basis function coefficients (e.g. Smith and Kohn, 1996; Yau, Kohn and Wood, 2000), the number and location of knots of piecewise constant functions or regression splines (Denison et al., 1998; Biller, 2000), or - in a spatial context - partitions of a spatial domain (Heikkinen and Arjas, 1998; Knorr-Held and Raßer, 1999).

Smoothness priors approaches are stochastic analogues of penalized likelihood methods. They assign appropriate priors, differing in the type and degree of smoothness

for function values or parameters, which are neighbours in the domain of a metrical covariate or a time scale, or in space. These concepts have been used for Bayesian smoothing splines (Hastie and Tibshirani, 2000) , in dynamic models (Knorr-Held, 1999) or in Markov random field models for spatial smoothing (Besag et al., 1991). Fahrmeir and Lang (2001a, b) combine these concepts for a unified treatment of GAMMs. Using the fact that smoothness priors have a common general multivariate Gaussian form, posterior inference is possible with algorithmically efficient MCMC techniques. Finally, Lang and Brezger (2001) develop Bayesian P-splines models, which may be seen as a combination of basis functions and smoothness priors approaches.

Bayesian GAMMs offer a powerful tool for regression analysis of data correlated over time or space. However, in applications some of the following questions typically arise: How well can an underlying structure, including additive effects of covariates, time and space, be recovered on the basis of the observed data? Is it possible to distinguish between spatially correlated random effects modelling a smooth spatial structure and uncorrelated random effects, which should capture unobserved heterogeneity? Assuming a true underlying regression function, can we say something about bias and variance of the estimate? What is the influence of signal-to-noise or variance-to-variance-ratios? How big is the influence of the information content of different types of responses?

To find answers to these questions for realistically complex models, analytical results are generally not available. Consequently, there is an obvious need for thorough simulation studies, which can give at least partial answers on an empirical basis. Surprisingly, there is a clear lack of such studies in the literature. Some exceptions are Smith, Wong and Kohn (1998), Smith and Kohn (1996) and Bernardinelli, Clayton and Montomoli (1995).

The aim of this paper is to explore some of the open questions mentioned above for Bayesian GAMMs developed in Fahrmeir and Lang (2001a, b). We focus here on Gaussian and (multi-)categorical probit models, with predictors including all the different types of effects described at the beginning in additive stylized form. The results provide further insight into the properties of posterior inference with these models and give guidance for their facilities and limitations in real data applications.

2 Bayesian generalized additive mixed models

Generalized linear models assume that, given covariates w, the distribution of the response variable y belongs to an exponential family with mean $\mu = E(y|w)$ linked to a linear predictor $\eta = \gamma' w$ by $\mu = h(\eta)$. Here h is a known link or response function and $\gamma = (\gamma_1, \ldots, \gamma_q)$ is the vector of unknown regression coefficients to be estimated. GAMMs extend GLMs by replacing the simple parametric linear predictor through a flexible semiparametric additive predictor

$$\eta = f_1(x_1) + \cdots + f_p(x_p) + \gamma' w + b_g.$$

The unknown functions $f_1, \ldots, f_p$ are (more or less) smooth functions of metrical covariates, time scales and spatial covariates. The parameters $b_g, g \in \{1, \ldots, G\}$, are

uncorrelated (unstructured) random effects, mainly introduced to model unobserved heterogeneity. In most cases the grouping variable is an index that identifies different units or clusters. It may also be an indicator of spatial sites to incorporate spatial heterogeneity. Then, a common approach is to decompose a spatial effect $f_{spat}(s)$, say, where $s \in \{1, \ldots, S\}$ denotes the sites on a lattice, into a smooth, structured component $f_{str}(s)$ and an unstructured random effect b_s. A rational for an additive decomposition $f_{spat}(s) = f_{str}(s) + b_s$ is that a spatial effect is usually a surrogate of many underlying unobserved influential factors. Some of them may obey a strong spatial structure others may be present only locally. Such models are common in spatial epidemiology (Besag et al., 1991; Knorr-Held and Besag, 1998) and have also been used in Fahrmeir and Lang (2001b). With similar arguments, we could also decompose functions of a time scale or a metrical covariate into a smooth component and an uncorrelated random effect. A main focus of the simulation study in Section 3 is to investigate whether such a decomposition is identifiable given the data and the priors. (Note that the decomposition is not likelihood identifiable!)

Smoothness priors for the unknown functions $f_1, \ldots, f_p$ depend upon the type of covariate. Alternatives for metrical covariates and time scales are random walk priors (Fahrmeir and Lang, 2001a), P-spline coefficient priors (Lang and Brezger, 2001) and smoothing spline priors (Hastie and Tibshirani, 2000). In any case, it is possible to express an unknown smooth function f, more exactly the corresponding vector of function evaluations $f = (f(x_1), \ldots, f(x_n))$, as the matrix product of a design matrix X and a vector of unknown parameters β, i.e. $f = X\beta$. All priors for a coefficient vector β have the same general Gaussian form

$$\beta|\tau^2 \propto \exp\left(-\frac{1}{2\tau^2}\beta' K \beta\right). \tag{1}$$

This implies that $\beta|\tau^2$ follows a partially improper Gaussian prior

$$\beta|\tau^2 \sim N(0, \tau^2 K^-),$$

where K^- is a generalized inverse of the penalty matrix K. The amount of smoothness of a function f is controlled by the variance parameter τ^2. For a fully Bayesian analysis, a hyperprior for τ^2 is introduced in a further stage of the hierarchy. This allows for simultaneous estimation of the unknown function and the amount of smoothness. A common choice is a highly dispersed but proper inverse gamma prior $\tau^2 \sim IG(a, b)$. (This type of prior is also chosen for $var(\epsilon) = \sigma^2$ in a Gaussian additive mixed model $y = \eta + \epsilon$.)

Functions $f(x)$ of a spatial covariate x, where x_i represents the location of an observation i in a spatial domain, can also be expressed in the form $f = X\beta$. In the simplest case, a component β_s of β is equal to $f(s)$, the value of $f(.)$ at site s, and a Markov random field prior as in Besag et al. (1991) is assumed for β. Again it can be brought into the general form (1).

For unstructured random effects, a usual assumption for the prior is that the b_g's are i.i.d. Gaussian,

$$b_g|v^2 \sim N(0, v^2), \quad g = 1, \ldots, G \tag{2}$$

with a highly dispersed hyperprior for v^2.

Apparently, there is only a slight difference to the additive smooth effects $f_1, \ldots, f_p$. In fact, instead of specifying for example first or second order random walk priors for a function f, the random effects prior (2) may be specified as well. The main difference between the two specifications is the amount of smoothness allowed for a function f. With a random effects specification successive parameters are allowed to vary more or less *unrestricted*, whereas random walk priors guarantee that successive parameters vary smoothly over the range of x. As already mentioned in some cases it may be even necessary to include both a structured and an unstructured effect into the predictor.

Bayesian inference is based on the posterior distribution of the model. In all practical situations the posterior distribution is numerically intractable and posterior analysis is carried out with MCMC simulation. The exact MCMC simulation techniques used are described in detail in Fahrmeir and Lang (2001b), see also Hastie and Tibshirani (2000) for the case of Gaussian responses. For Gaussian and categorical probit models, which are the focus of this paper, posterior simulations are possible with Gibbs sampling. In other cases, MH steps based on conditional prior proposals (Knorr–Held 1999) or on iteratively weighted least squares proposals (Gamerman, 1997) are a possible alternative.

All computations have been carried out with *BayesX* a software package for Bayesian inference with MCMC techniques. The program and the user manual are available via internet at `http://www.stat.uni-muenchen.de/~lang/`. The user manual of *BayesX* also contains a short survey of the MCMC techniques used in this paper.

3 Simulation study

3.1 Simulation design

In our simulation study we present results for Gaussian responses, binary probit models and multinomial probit models with three categories.

In all cases we start with a basis model that will be extended to more complicated situations. The predictor of our basis model (M1) is given by

$$\eta_{it} = const + f(x_{it}) + b_i, \quad i = 1, \ldots, 50, \quad t = 1, \ldots, 15,$$

where $f(x) = sin(x)$ and $b_i \sim N(0, v^2)$ is an unstructured random effect. The model corresponds to longitudinal data with 50 individuals observed during an observation period of 15 time points. The values of the covariate x are chosen randomly within an equidistant grid of 50 points between -3 and 3. Thus, x is a time varying covariate. In a second model (M2), x is restricted to be time constant, that is for every individual i exactly one of the 50 different covariate values of x is observed. This is technically equivalent to models where the effect of a particular covariate is split up into a structured (smooth) and an unstructured (unsmooth) effect. The question arising in such situations is whether we can really distinguish between the two different effects.

In a next step we replace the metrical covariate x in models M1 and M2 by a spatial covariate (models M3 and M4). More specifically, we assume now that x denotes

the location in the geographical map shown in Figure 1. The function f is chosen as a sinusoidal function of the centroid coordinates x_1 and x_2 of the regions, i.e. $f(x_1, x_2) = sin(x_1 x_2) + 0.148$. The constant 0.148 is added to ensure that the sum of function evaluations is zero. The true function used for simulations is shown in Figure 1 a). Since the map consists of 309 districts the number of individuals is increased from 50 to 309 in order to obtain the same situation as in models M1 and M2.

Overall, we end up with 4 different models which we summarize in Table 1. In order to investigate the effect of the variance v^2 of the random effects on estimation results we simulated the models with three different variances $v^2 = 0.25$, $v^2 = 0.5$ and $v^2 = 0.75$. For Gaussian responses the variance of the error term is always set to 1. In the case of multinomial responses with three categories a second predictor has to be designed. In all cases the second predictor is composed of the same covariates and has the same structure as the predictor of the first category but with a different functional form for f. The predictors of the second category are summarized in Table 2.

We also experimented with models where an additional nonlinear "time trend" $g(t)$ has been added to the predictors of our models. Since in all cases the nonlinear time trend was estimated almost unbiased and results for f and the random effects were close to models M1-M4 we omit to show results for these models.

As smoothness priors for the function f we usually choose a second order random walk if x is metrical, and a Markov random field prior with adjacency weights for spatial covariates. For metrical covariates we also tested other smoothness priors, e.g. P-splines or smoothing splines, but the differences in results are more or less negligible. For the hyperparameters of inverse gamma priors of variance components we set $a = 1$ and $b = 0.005$ as a standard choice.

We generated 250 replications y^1 - y^{250} for every model M1-M4 and variances $v^2 = 0.25$, $v^2 = 0.5$ and $v^2 = 0.75$. The replications are simulated always with the same random effects coefficients b_i, i.e. the random effect can be seen as a fixed but unstructured (unsmooth) function. In order to make results comparable we also used identical random effects coefficients for models M1 and M2, and M3 and M4, respectively. Usually, random effects are considered as an additional unit- or cluster specific random error similar to the overall errors. Thus, we could have also treated the random effects as additional errors and simulate new coefficients b_i, $i = 1, \ldots, n$, for every replication of the models. This strategy has been followed for example by Lin and Zhang (1999). However, some of the questions that we adress in this paper cannot be answered by treating the random effects as additional errors. Moreover, in the case of a spatial covariate the treatment of random effects as additional location specific errors is at least questionable, because then we tacitly assume that there is an indefinite number of locations.

In the next sections results are compared by computing the average estimated functions and random effects coefficients of the 250 replications, i.e.

$$\bar{f}(x) = \frac{1}{250} \sum_{j=1}^{250} \hat{f}^{y^j}(x)$$

for f and analogue terms for the random effects. We also computed the square root of the relative MSEs

$$\sqrt{MSE(j)} = \sqrt{\frac{\sum_x (\hat{f}^{y^j}(x) - f(x))^2}{\sum_x f(x)^2}}$$

for the j-th replication of the models, $j = 1, \ldots, 250$.

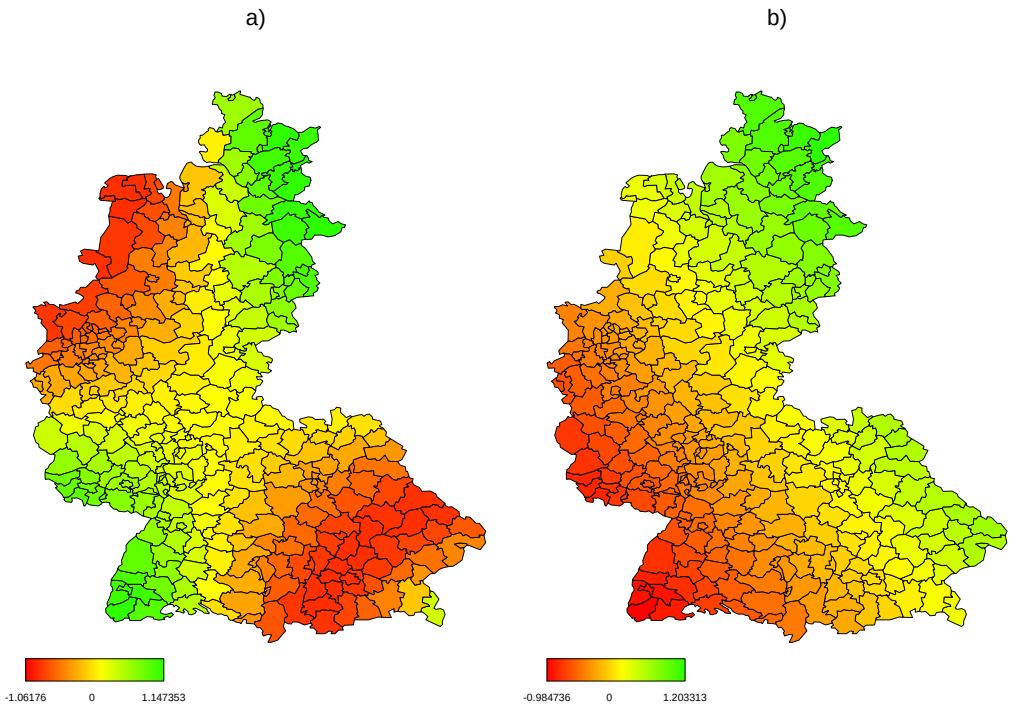

Figure 1: *Maps of the "true" function f used for simulation. Figure a) displays the map used for Gaussian responses, probit models and the first category of multicategorical probit models. Figure b) displays the map used for the second category of multicategorical probit models.*

Model	η	**covariate** x	**nonlinear functions**
M1	$f(x_{it}) + b_i$	x metrical, time varying	$f(x) = sin(x)$
M2	$f(x_i) + b_i$	x metrical, time constant	$f(x) = sin(x)$
M3	$f(x_{it}) + b_i$	x spatial, time varying	$f(x_1, x_2) = sin(x_1 x_2) + 0.148$ see also Fig. 1a)
M4	$f(x_i) + b_i$	x spatial, time constant	$f(x_1, x_2) = sin(x_1 x_2) + 0.148$ see also Fig. 1a)

Table 1: *Summary of simulated models (Gaussian responses and probit models).*

Model	nonlinear function for η_1	nonlinear function for η_2
M1,M2	$f(x) = sin(x)$	$f(x) = \frac{x^2}{4.5} - 0.694$
M3,M4	$f(x_1, x_2) = sin(x_1 x_2) + 0.148$ see also Fig. 1a)	$f(x_1, x_2) = 0.4(x_1 + x_2)$ see also Fig. 1b)

Table 2: *Summary of simulated models for multicategorical probit models.*

3.2 Results for Gaussian responses

Models M1 and M2

We first compare the results for models M1 and M2 where f is a function of the metrical covariate x. Figure 2 shows the average posterior mean estimates $\bar{f}(x)$ for f (dashed lines) for models M1 (left panel) and M2 (right panel). For comparison the true function is additionally included (solid line). The function is estimated almost unbiased for model M1 whereas for model M2 estimates are considerably biased. Also for model M2, the bias becomes larger if the variance of the random effects is increased. This is, however, not too surprising since the smooth function is considerably masked by the superimposed unstructured effect. Figure 3 displays the average deviation of the estimated random effects coefficients (dashed line) from the true coefficients (solid line). More specifically, the figure maps the true coefficients on the x-axis against the estimated and the true coefficients on the y-axis. Thus, if the coefficients are estimated without any bias the dashed and the solid line must coincide. In all cases we observe more bias if the absolute values of the true coefficients become larger. Similar to the results for the smooth function more bias is obtained for model M2 compared to model M1. There is however an important difference between the results for the smooth and the unstructured unsmooth effect. For larger variances of the random effects the bias decreases for both models M1 and M2, whereas the bias for f clearly increases. This indicates that both the smooth and the unsmooth effect is recovered better if it is less masked by the superimposed second effect. The question arises whether the lack of fit for the smooth and the unsmooth effect in model M2 also leads to a considerable reduction of the overall fit. To answer this question we compared the overall fit measured by the deviance (or the residual sum of squares) for the different models and variances v^2 of the random effects (Table 3). Surprisingly, the increase of the deviance is relatively small for model M2 compared to model M1. This implies that the main problem is more the separation of the structured and the unstructured effect rather than a decrease of the overall fit. This becomes even more obvious with Figure 4 in which boxplots of the relative MSEs for f (first graph), the random effects coefficients (second graph) and the sum of both effects for model M2 (third graph) are printed. For both, the smooth und the unsmooth effect, the relative MSEs become worse for model M2 whereas the MSEs of the sum of both effects become much smaller compared to the MSEs for individual terms.

The problem of insufficient separation of the smooth and the unsmooth effect might be caused by the specific random effects generated once and held fixed in all replications. In order to check this we rerun the simulations for model M2 treating the

random effects part in the predictor as an additional individual specific error rather than an unstructured function. Replications of the model are now obtained by simulating not only the overall errors but also the random effects. The resulting average function estimate for f is shown in Figure 5 a) for the case $v^2 = 0.5$. The panels b) - d) show the estimates corresponding to the upper tenth decile, the median and the lower tenth decile of the relative MSE measure. Similar to the simulations with random effects held fixed the bias is higher compared to model M1. Moreover, looking at the 10th percent best fit, the median fit and the 10th percent worst fit in panels b) - d) we observe that at least 50 percent of the estimates are of equal or even worse quality than the average estimate for f in Figure 2. Also, the relative MSEs are only marginally smaller compared to the simulations with random effects held fixed in every replication. That is, at least in a substantial number of cases the smooth and the unsmooth effect is separated with considerable bias.

We also estimated all models without incorporating the random effects component into the predictors in order to see how the non-consideration of the random effect changes the fit for f and the overall fit. As an example, Figure 6 shows the average posterior mean for models M1 (left panel) and M2 (right panel) in the case of $v^2 = 0.5$ for the random effects variance. Similar results are obtained for $v^2 = 0.25$ and $v^2 = 0.75$. The average deviance is 1638.64 for Model M1 and 1750.08 for M2. Neglecting the random effect leads to a dramatic increase of the deviance. Moreover, the estimate of the smooth function f is considerably masked by the random effect. This implies that a random effects component should always be incorporated although an unbiased separation of an effect into a smooth and an unsmooth part may be difficult.

Besides point estimates, an important aspect of statistical inference are credible intervals or regions for the unknown parameters. The question is: How accurate are the true values covered by the credible intervals? In a Bayesian framework based on MCMC simulation techniques pointwise credible intervals for the function evaluations of f and the random effects coefficients are estimated by computing the respective empirical quantils of the sampled parameters. Table 4 shows the average coverage (in percent) of pointwise 80 % credible intervals for function evaluations of f and the random effects coefficients, respectively. For model M1 the coverage of function evaluations of f is in all cases clearly above 80 %, for the random effects coefficients the average coverage is more or less identical to 80 % implying the suggestion that our Bayesian credible intervals are relatively conservative estimates. To assess the effect of the curvature of f on the covarage we also experimented with more or less curved functions than the sine function. We replaced the true function $f(x) = sin(x)$ by functions $f(x) = sin(ax)$ where $a = 0.25, 0.5, 2, 4$ and rerun the simulations. For $a = 0.25$ and $a = 0.5$ the resulting functions are less curved than the sine function, and for $a = 2$ and $a = 4$ the functions are more curved. We obtained average coverage rates between 84 % and 87 % implying that the curvature of the function has a negligible effect on the coverage. Not surprisingly, the average coverage for model M2 is in all cases beyond 80 %, sometimes far beyond. Clearly, the reason is that the estimation procedure is not able to distinguish correctly between the smooth and the unsmooth effect. The coverage of the sum of both effects, however, is in all cases almost identical to 80 %.

Finally, we investigated the accuracy of point estimates for the variance component v^2 of the random effects. Table 5 shows average point estimates obtained by the posterior mean and the average bias. Even for model M2 the bias can be more or less neglected.

Models M3 and M4

Compared to models M1 and M2 the covariate x is now spatial rather than metrical. Thus, x represents the location of a particular observation in the map shown in Figure 1. Model M1 corresponds to model M3, and model M2 to model M4. In general, the differences to the case of a metrical covariate are relatively small.

Figures 7 and 8, respectively, show the average posterior mean estimate and the average bias for models M3 (left panels) and M4 (right panels). To gain more insight, Figure 9 displays the average deviation of the estimated function evaluations for f (dashed line) from the true values (solid line). For model M3 the true function is recovered almost unbiased, whereas for model M4 we observe a considerable bias which increases with increasing variance v^2 of the random effects. Similar to model M2 the estimation procedure seems to have problems to distinguish between the smooth and the unsmooth effect. This is confirmed with Figures 10 and 11. Figure 10 displays the average deviation of the random effects coefficients (dashed line) from the true coefficients (solid line). For both models M3 and M4 the bias decreases with increasing variance v^2 of the random effects, i.e. the random effects coefficients are estimated more accurately if the unstructured effect is less masked by the smooth effect (and vice versa). Figure 11 shows boxplots of the relative MSEs for f (first graph), the random effects coefficients (second graph) and for the sum of both effects for model M4 (third graph). It confirms the findings for models M1 and M2. Although it seems that it is not possible to separate the two different effects in model M4, the sum of both effects is estimated satisfactorily.

In analogy to models M1 and M2 we also investigated the coverage of credible intervals. The average coverage for models M3 and M4 can be found in Table 4. The results are quite similar to the findings for models M1 and M2. Note, that the coverage for the structured spatial effect is even higher than the coverage for a metrical covariate.

Model	v^2	**average deviance**	**Model**	v^2	**average deviance**
M1	0.25	704.92	M3	0.25	4209.59
M1	0.5	699.67	M3	0.5	4180.85
M1	0.75	697.33	M3	0.75	4169.35
M2	0.25	711.52	M4	0.25	4379.39
M2	0.5	705.93	M4	0.5	4357.24
M2	0.75	703.33	M4	0.75	4350.28

Table 3: *Gaussian responses: Average deviance.*

Model	v^2	**ave. coverage** f_1	**ave. coverage** b
M1	0.25	0.87	0.80
M1	0.5	0.87	0.80
M1	0.75	0.86	0.80
M2	0.25	0.74	0.75
M2	0.5	0.61	0.67
M2	0.75	0.66	0.72
M3	0.25	0.91	0.80
M3	0.5	0.91	0.80
M3	0.75	0.91	0.80
M4	0.25	0.76	0.74
M4	0.5	0.59	0.68
M4	0.75	0.44	0.61

Table 4: *Gaussian responses: Average coverage of pointwise 80 % credible intervals for f and the random effects coefficients.*

Model	**true** v^2	$\hat{v^2}$	**ave. bias**	**Model**	**true** v^2	$\hat{v^2}$	**ave. bias**
M1	0.25	0.244	-0.015	M3	0.25	0.256	-0.001
M1	0.5	0.470	-0.021	M3	0.5	0.496	-0.002
M1	0.75	0.731	-0.022	M3	0.75	0.749	-0.001
M2	0.25	0.210	-0.049	M4	0.25	0.265	0.007
M2	0.5	0.418	-0.074	M4	0.5	0.591	0.092
M2	0.75	0.700	-0.056	M4	0.75	0.845	0.094

Table 5: *Gaussian responses: Average estimate and bias of the variance component.*

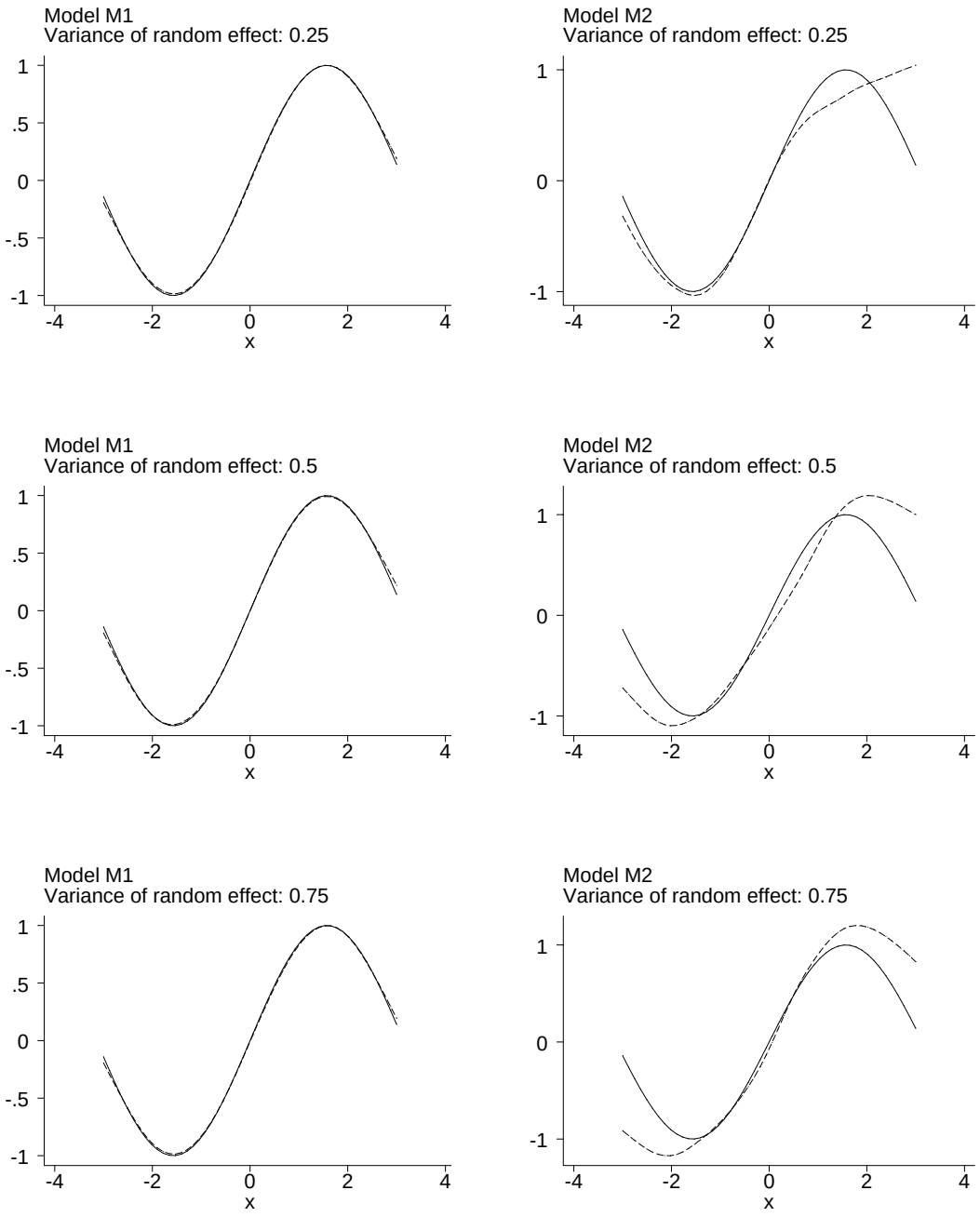

Figure 2: *Gaussian responses: True function (solid line) and average posterior mean estimate for f.*

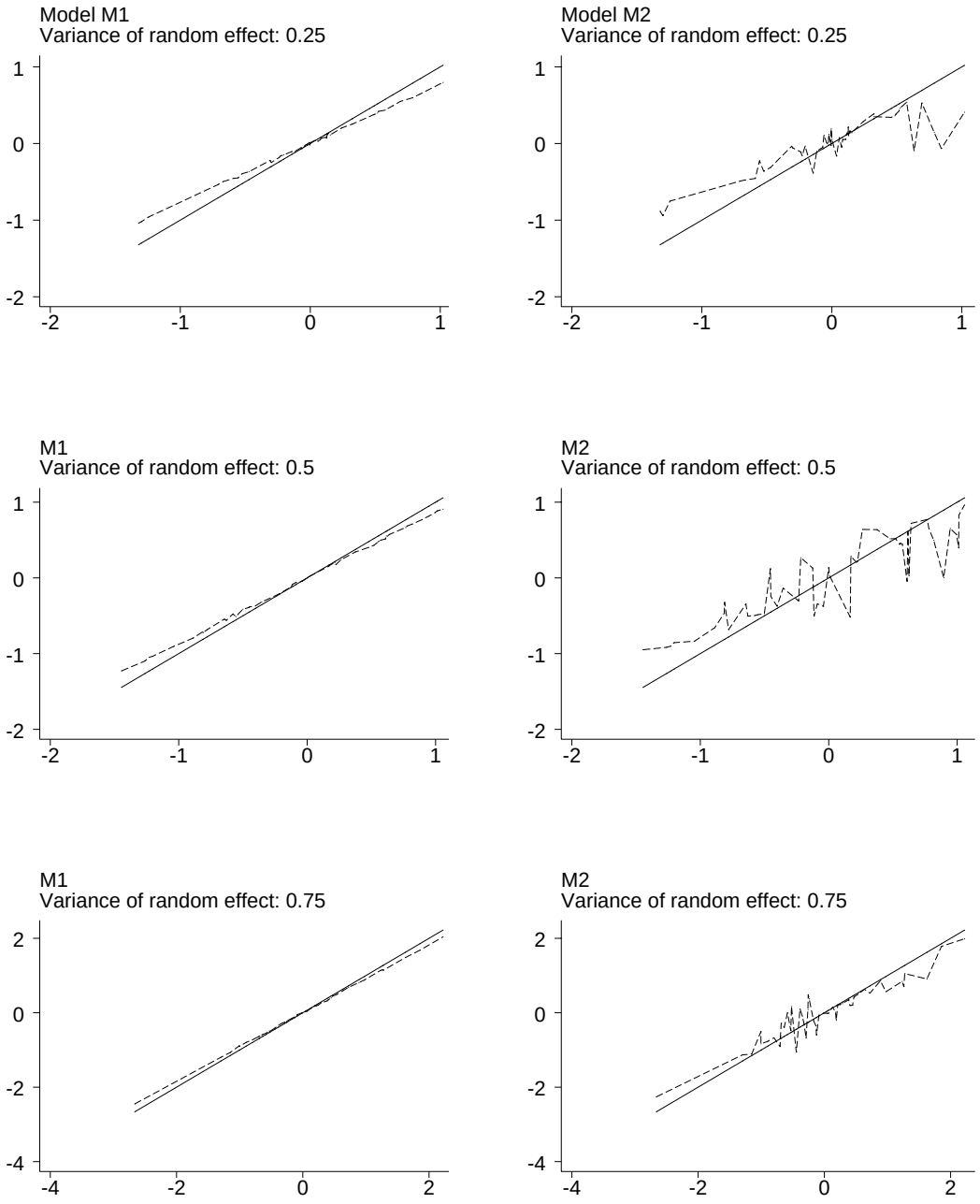

Figure 3: *Gaussian responses: Average deviation of the estimated random effects coefficients (dashed line) from the true coefficients (solid line). The graphs map the true coefficients on the x-axis against the estimated and the true coefficients on the y-axis.*

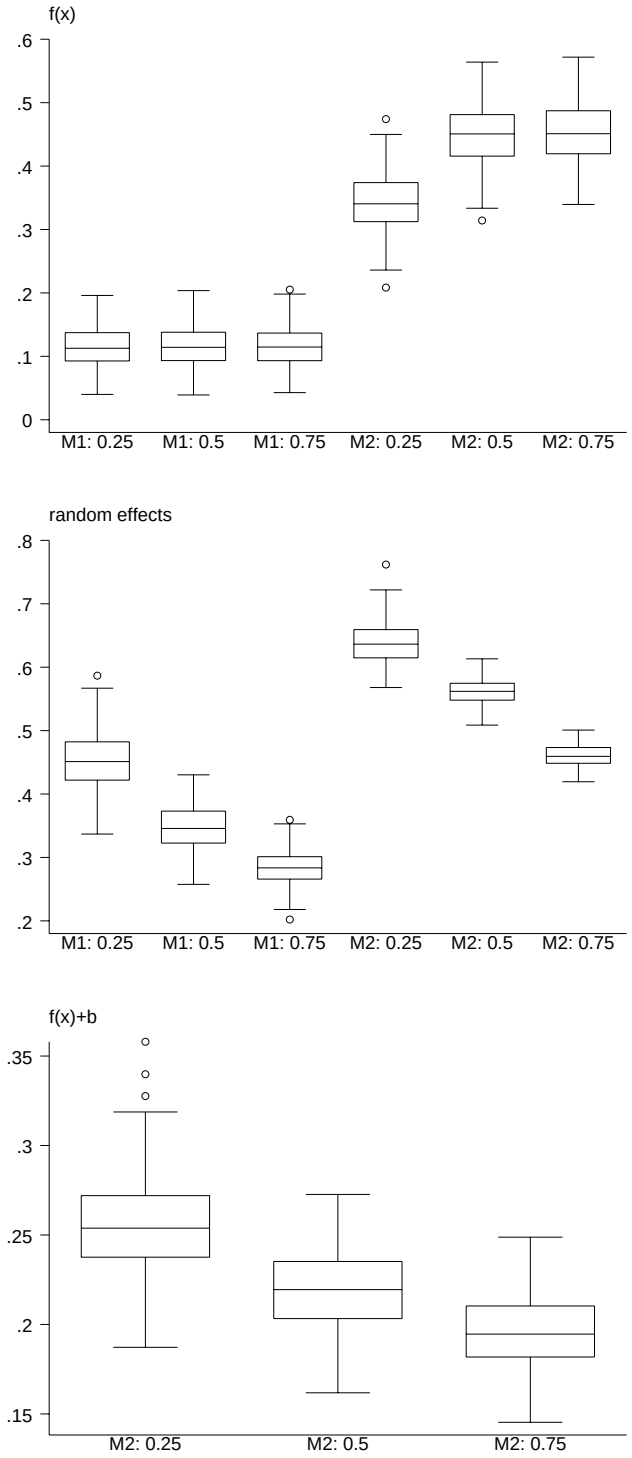

Figure 4: *Gaussian responses (models M1 and M2): Boxplots of relative MSEs for f, random effects and the sum of both effects (model M2 only).*

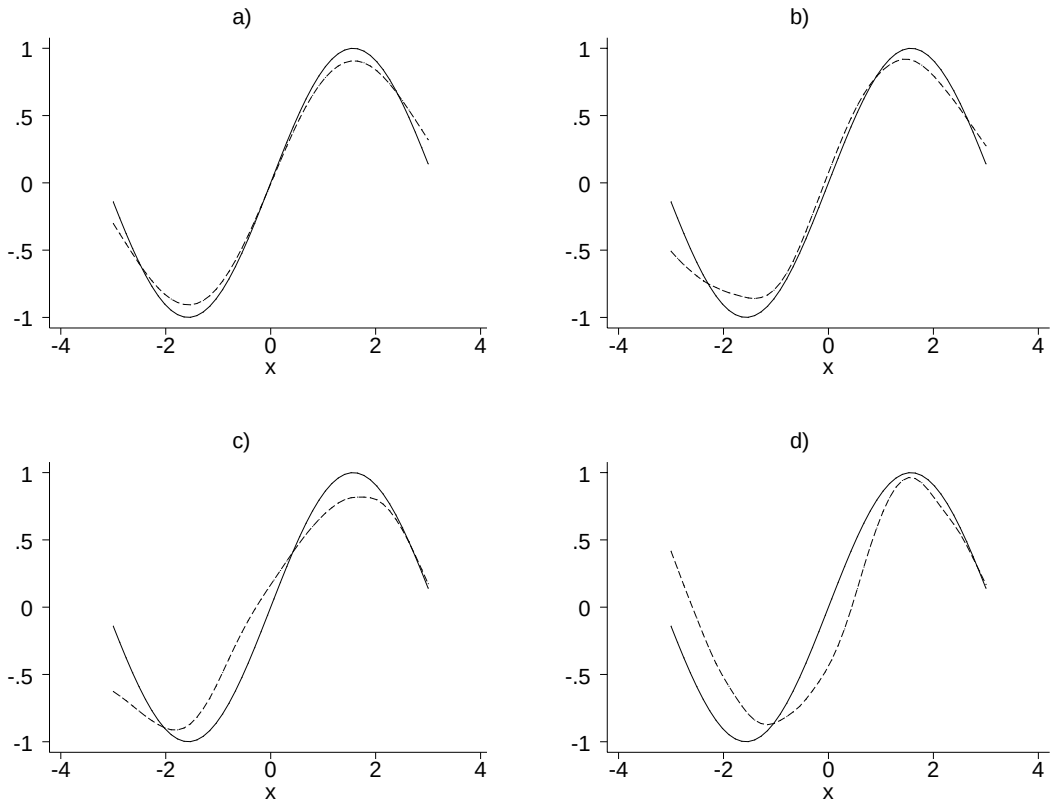

Figure 5: *Gaussian responses: Estimated function f for model M2 and $v^2 = 0.5$ if the random effects are resampled in every replication. Panel a) shows the estimate for f averaged over the 250 replications. Panels b) - d) show the estimates corresponding to the upper tenth decile, the median decile and the lower tenth decile of the relative MSE measure.*

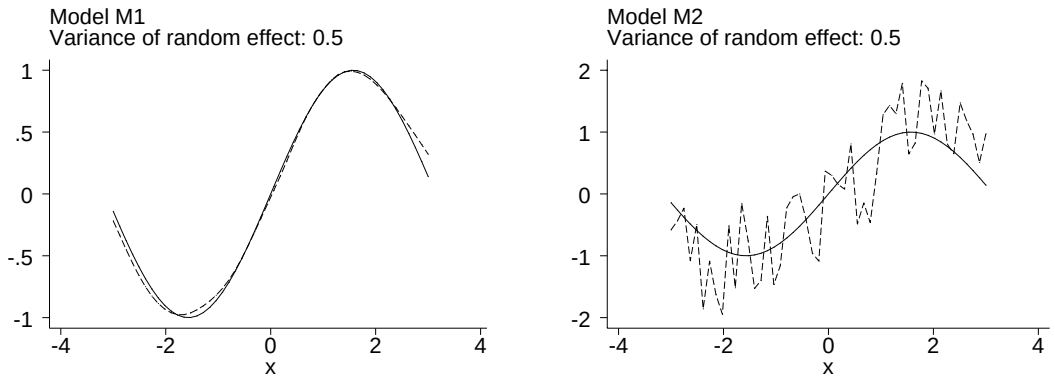

Figure 6: *Gaussian responses: True function (solid line) and average posterior mean estimate for f if the random effects are neglected.*

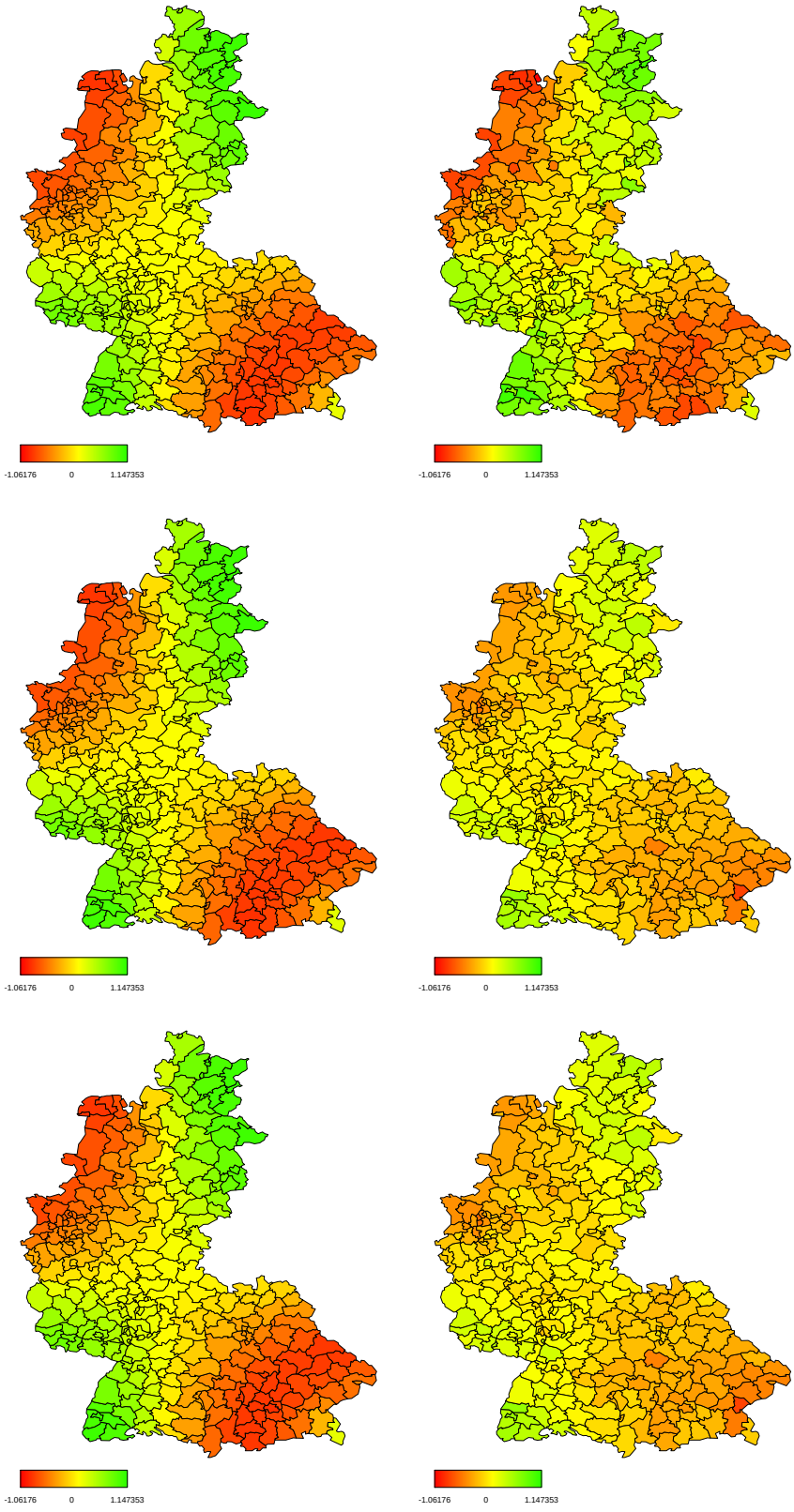

Figure 7: *Gaussian responses: Average posterior mean estimate for f. The true function can be found in Figure 1 a). The left panel refers to model M3 and the right panel to model M4. All maps including the true map are coloured in the same scale in order to guarantee comparability.*

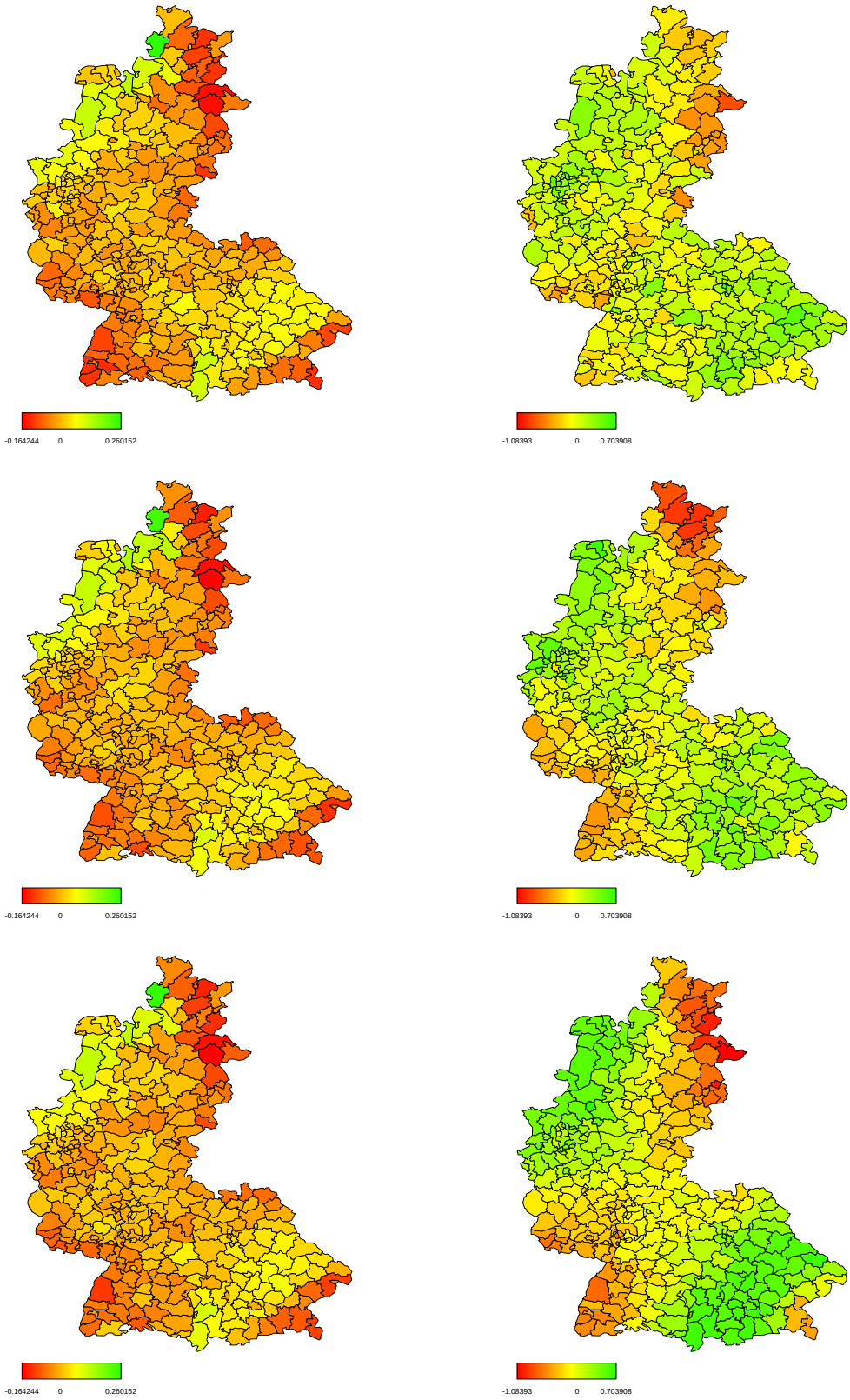

Figure 8: *Gaussian responses: Average bias for f. The left panel refers to model M3 and the right panel to model M4. The left and the right panel are coloured in different scales because of the large differences in the amount of the bias.*

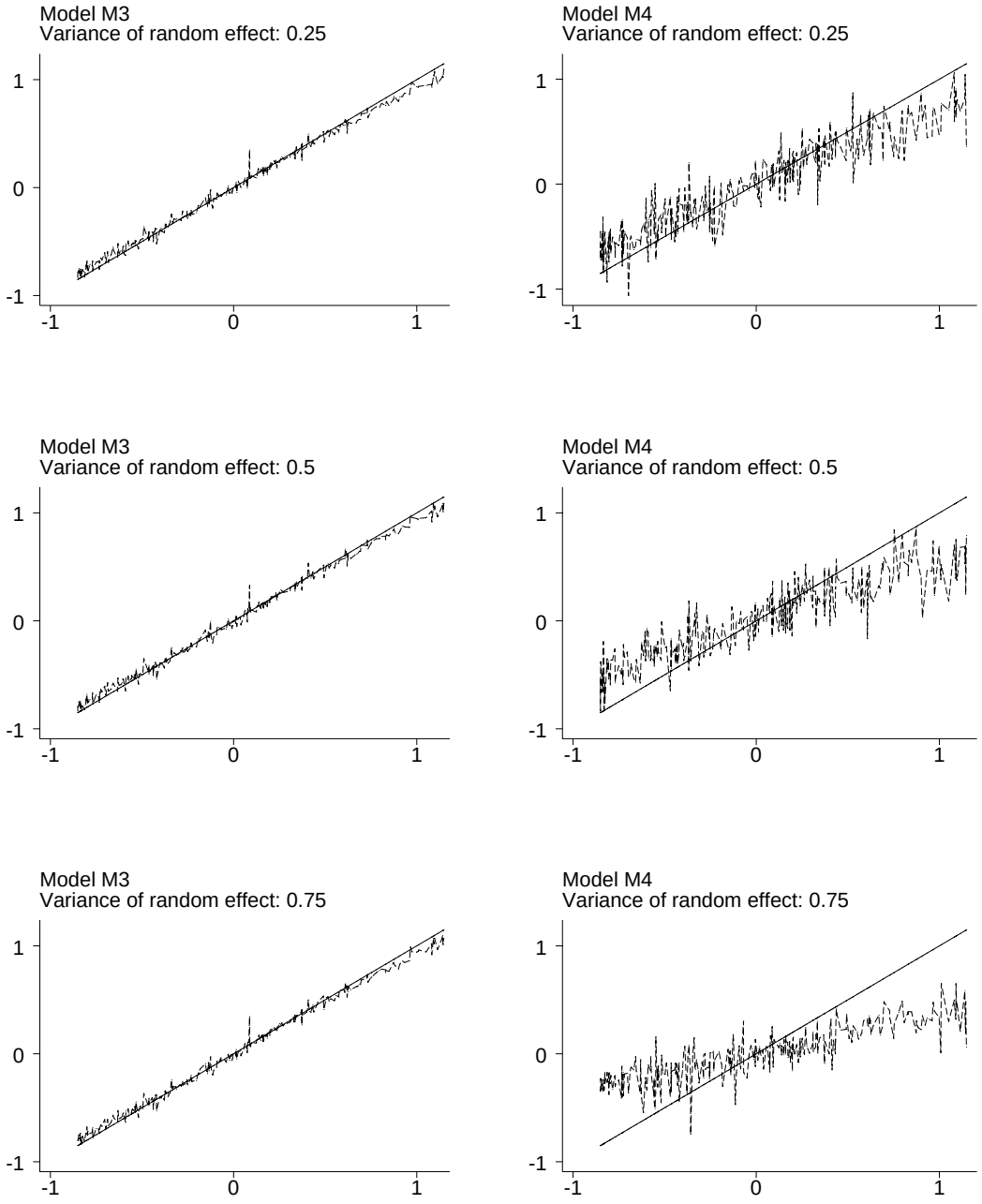

Figure 9: *Gaussian responses: Average deviation of the estimated spatial function (dashed line) from the true function (solid line). The graphs map the true values on the x-axis against the estimated and the true values on the y-axis.*

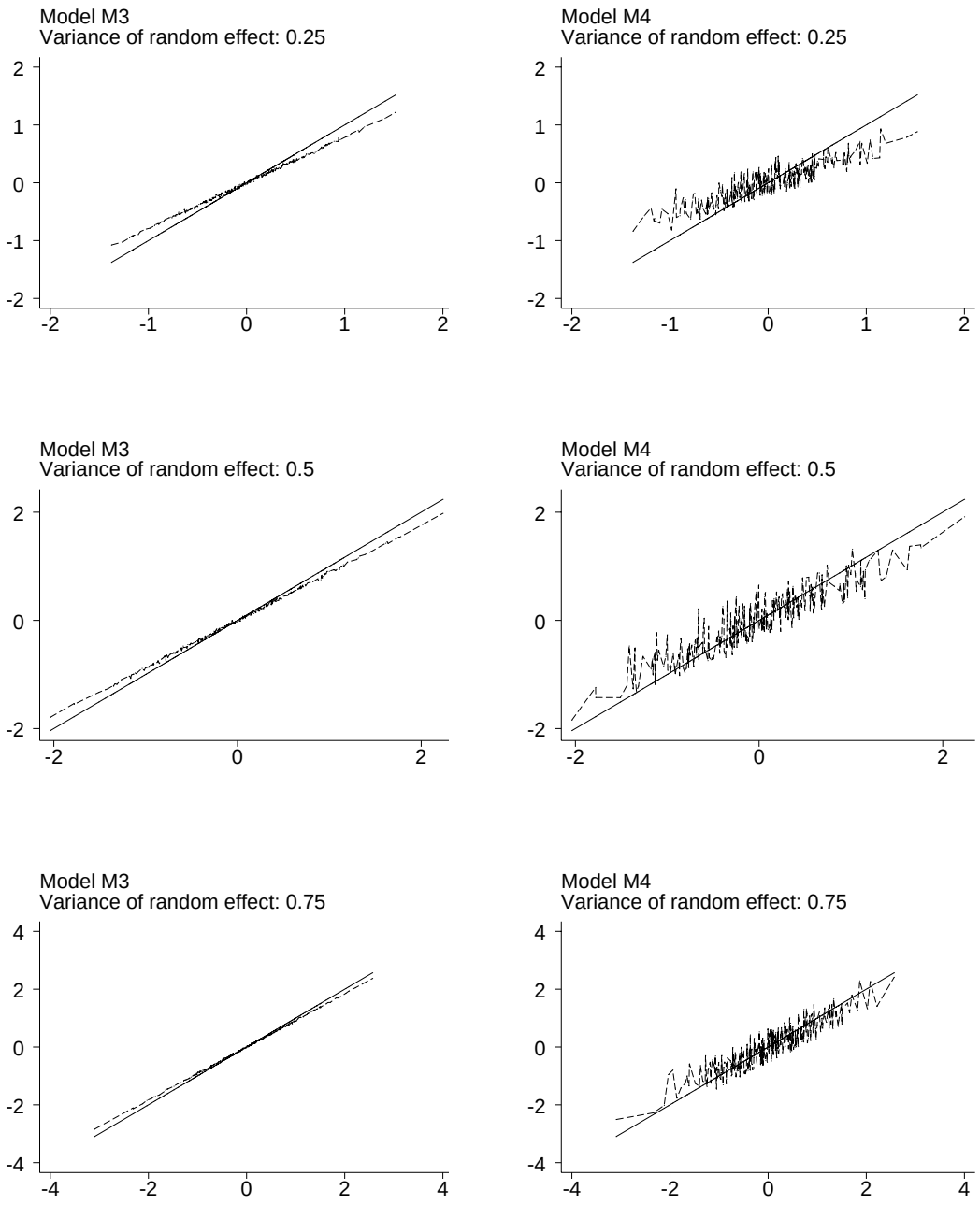

Figure 10: *Gaussian responses: Average deviation of the estimated random effects coefficients (dashed line) from the true coefficients (solid line). The graphs map the true coefficients on the x-axis against the estimated and the true coefficients on the y-axis.*

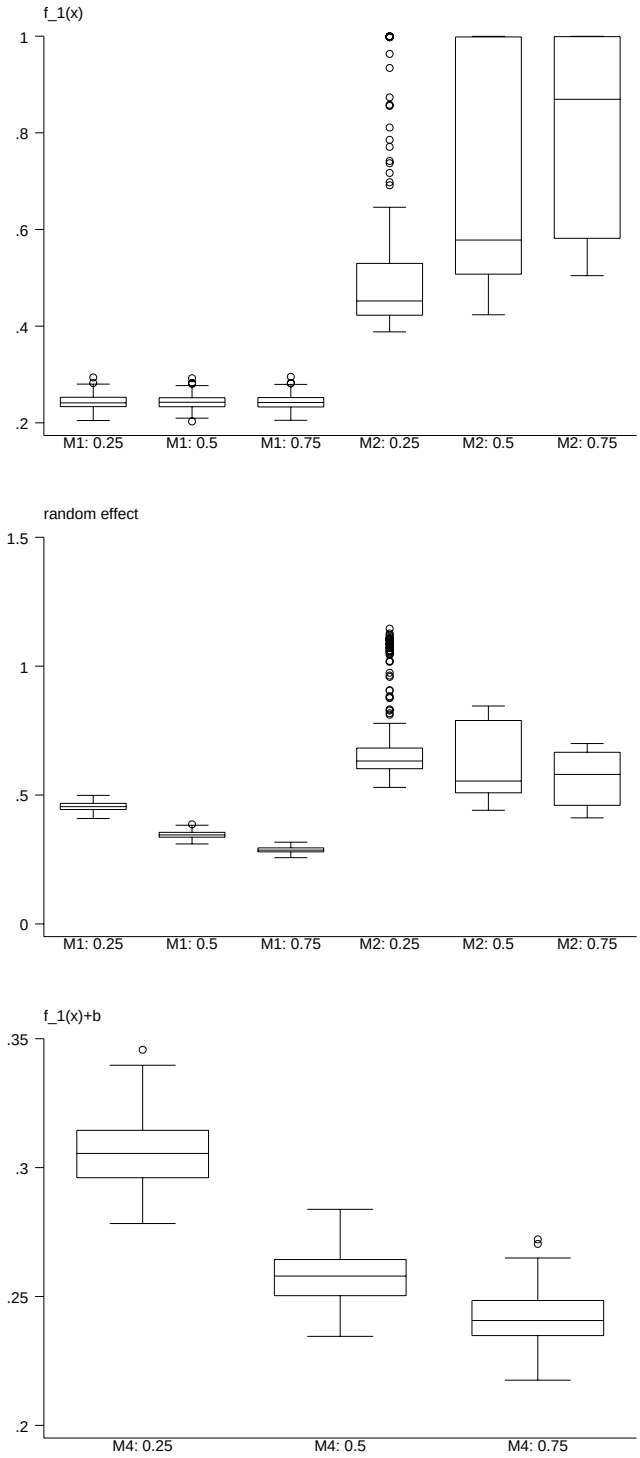

Figure 11: *Gaussian responses (models M3 and M4): Boxplots of relative MSEs for f, random effects coefficients and the sum of both effects (for model M4 only).*

3.3 Results for categorical responses

The main findings for categorical responses, binary as well as multicategorical, are virtually identical to the case of Gaussian responses. For that reason, the main focus of the presentation is restricted to the differences between metrical and categorical responses. Naturally, we would expect that both the bias and the MSE become larger for categorical data compared to metrical responses, because categorical data contain less information. Surprisingly, the loss of fit (in terms of bias and MSE) is unexpectedly small. As an example, Figure 12 compares the average estimates of the smooth function f for models M1 and M2 and $v^2 = 0.5$. The solid lines show the average posterior mean estimates for Gaussian responses and the dashed lines for binary probit models. We observe that the differences are more or less negligible. Figures 13 and 14 show the boxplots of the relative MSEs for Gaussian responses (1.–3. boxplot) and binary probit models (4.–6. boxplot). The first two graphs in each figure show the MSEs for the smooth function f. The third graphs display the MSEs for the sum of the smooth and the unsmooth effect (for models M2 and M4 only). There is an obvious but relatively small increase of the MSEs for probit models.

For binary probit models Figures 15 and 16 show the average deviations of the random effects coefficients from the true coefficients for models M1 and M2 and for models M3 and M4, respectively. They correspond to Figures 3 and 9 for Gaussian responses. We obtain virtually the same results as for Gaussian responses, but with one exception which is best seen in the left panels of Figures 15 and 16. For Gaussian responses the bias and the MSE decreases with increasing variance v^2 of the random effects. For probit models we obtain comparable results except for values of the random effects coefficients above 1.8 or below -1.8. The main reason is that the probabilities for $y = 1$ ($y = 0$) are already close to one (zero) for values of the predictor above 1.8 (below -1.8) and therefore values of the random effects coefficients above 1.8 or below -1.8 are extremly hard to identify unless we have a vast amount of data. Consequently, the bias of the posterior mean estimates for the variance component v^2 is slightly higher for $v^2 = 0.5$ and $v^2 = 0.75$ compared to Gaussian responses (Table 6).

An investigation of the coverage of 80 % credible intervals for binary responses gives similar results as for Gaussian responses. For $v^2 = 0.25$ and $v^2 = 0.75$ the average coverage of the random effects coefficients is slightly beyond 80 % in model M1. Note that the coverage of the credible interavals in models M2 and M4 is in all cases far beyond the nominal level. The coverage of the sum of both effects is, however, near or above the nominal level.

For multicategorical probit models the predictor of the first category is identical to the predictor used for binary probit models. Therefore we can assess a possible loss of fit that is obtained by switching from the binary case to the multicategorical case. Once again we observe that the loss of fit is relatively small. As an example, Figure 17 shows the relative MSEs for the smooth function f for binary probit models (1.–3. boxplots) and multicategorical probit models (4.–6-boxplots). The first graph shows the MSEs for model M1, the second graph for model M2 and the third graph for the sum of the smooth and the unsmooth effect (for model M2 only). As can be observed, the additional loss of fit for multicategorical probit models is small.

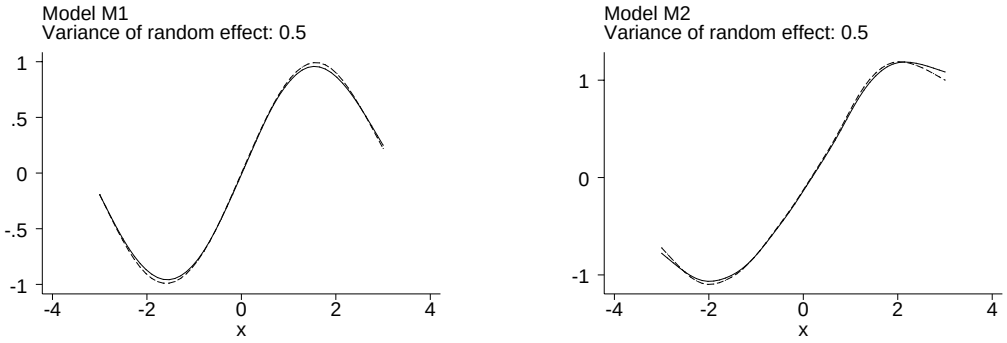

Figure 12: *Models M1 and M2 ($v^2 = 0.5$): Comparison of estimates for f for Gaussian responses (solid lines) and binary responses (dashed lines).*

Model	**true** v^2	$\hat{v^2}$	**ave. bias**	**Model**	**true** v^2	$\hat{v^2}$	**ave. bias**
M1	0.25	0.216	-0.043	M3	0.25	0.260	0.002
M1	0.5	0.497	0.005	M3	0.5	0.508	0.001
M1	0.75	0.661	-0.092	M3	0.75	0.742	-0.009
M2	0.25	0.149	-0.110	M4	0.25	0.292	0.034
M2	0.5	0.432	-0.321	M4	0.5	0.622	0.124
M2	0.75	0.571	-0.182	M4	0.75	0.847	0.096

Table 6: *Binary responses: Average estimate and bias of the variance component.*

Model	v^2	**ave. coverage** f_1	**ave. coverage** b
M1	0.25	0.84	0.76
M1	0.5	0.84	0.80
M1	0.75	0.84	0.78
M2	0.25	0.71	0.67
M2	0.5	0.63	0.71
M2	0.75	0.68	0.72
M3	0.25	0.93	0.80
M3	0.5	0.93	0.80
M3	0.75	0.93	0.80
M4	0.25	0.65	0.65
M4	0.5	0.50	0.67
M4	0.75	0.39	0.65

Table 7: *Binary responses: Average coverage of pointwise 80 % credible intervals for f and random effects coefficients.*

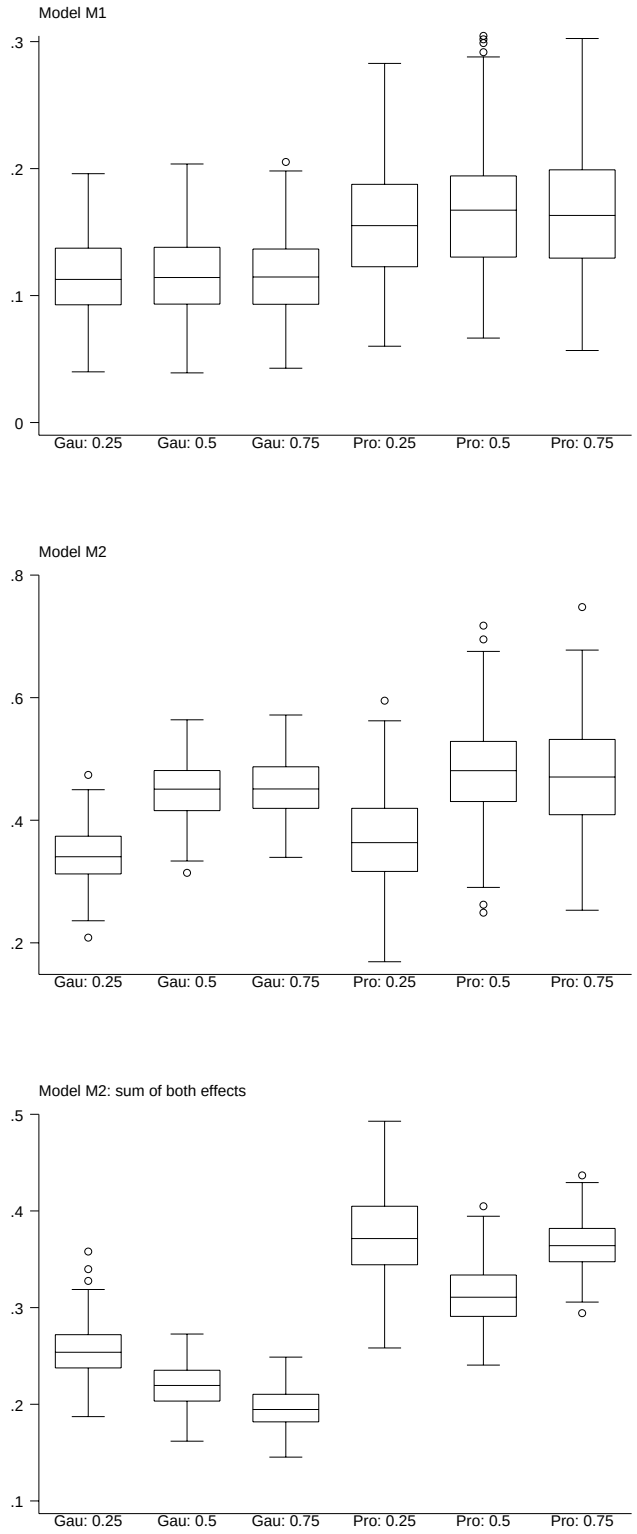

Figure 13: *Models M1 and M2: Boxplots of relative MSEs for f. The first to third boxplots refer to Gaussian responses and the fourth to sixth to binary responses. The third graph compares the MSEs for the sum of both effects (model M2 only).*

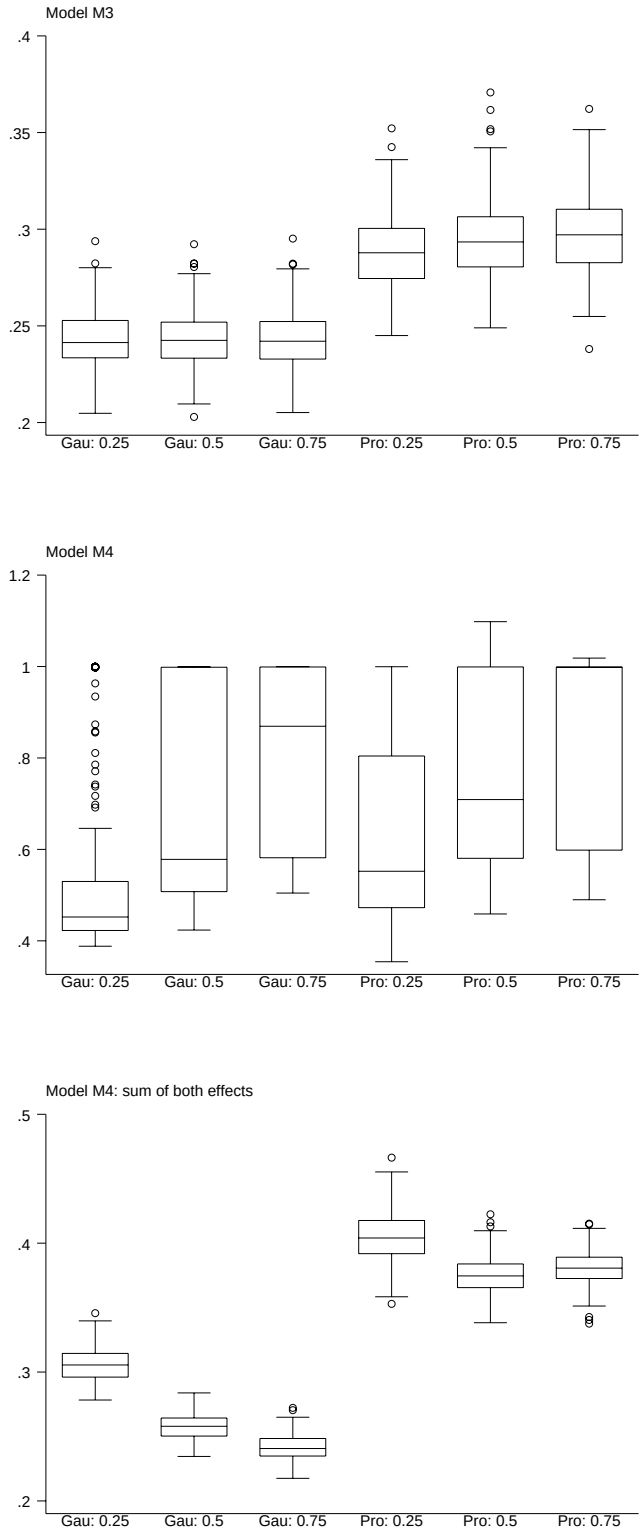

Figure 14: *Models M3 and M4: Boxplots of relative MSEs for f. The first to third boxplots refer to Gaussian responses and the fourth to sixth to binary responses. The third graph compares the MSEs for the sum of both effects (model M4 only).*

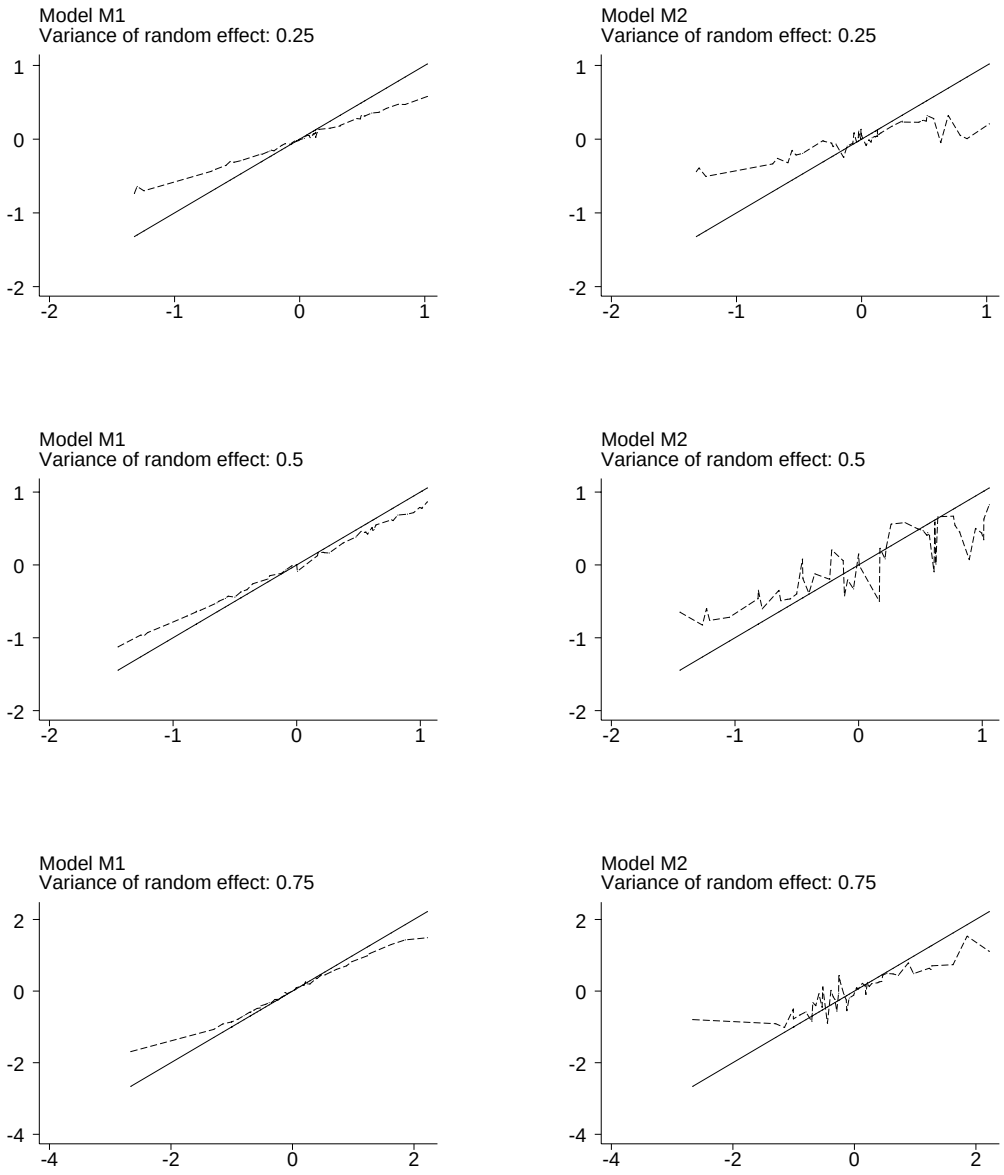

Figure 15: *Binary responses (models M1 and M2): Average deviation of the estimated random effects coefficients (dashed line) from the true coefficients (solid line). The graphs map the true coefficients on the x-axis against the estimated and the true coefficients on the y-axis.*

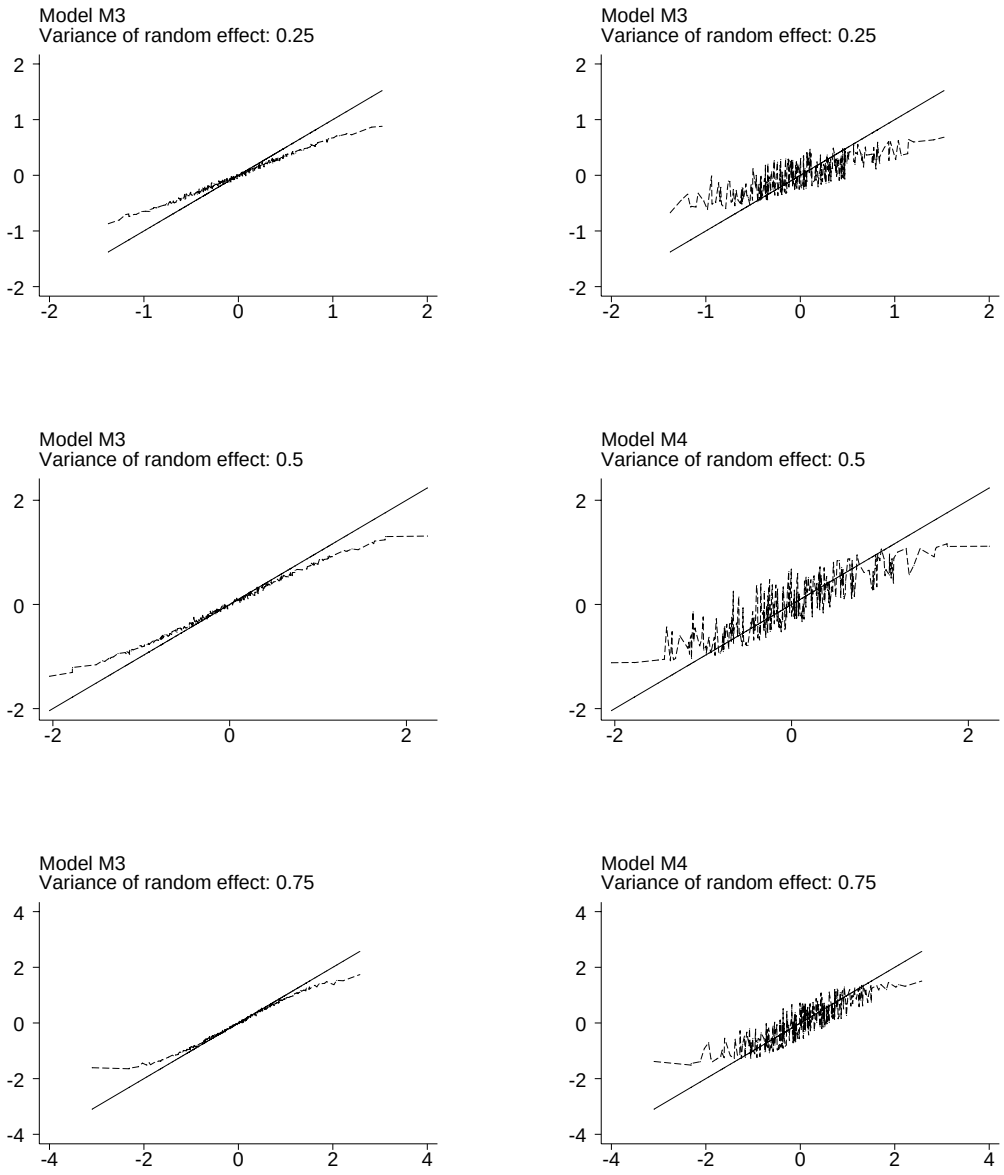

Figure 16: *Binary responses (models M3 and M4): Average deviation of the estimated random effects coefficients (dashed line) from the true coefficients (solid line). The graphs map the true coefficients on the x-axis against the estimated and the true coefficients on the y-axis.*

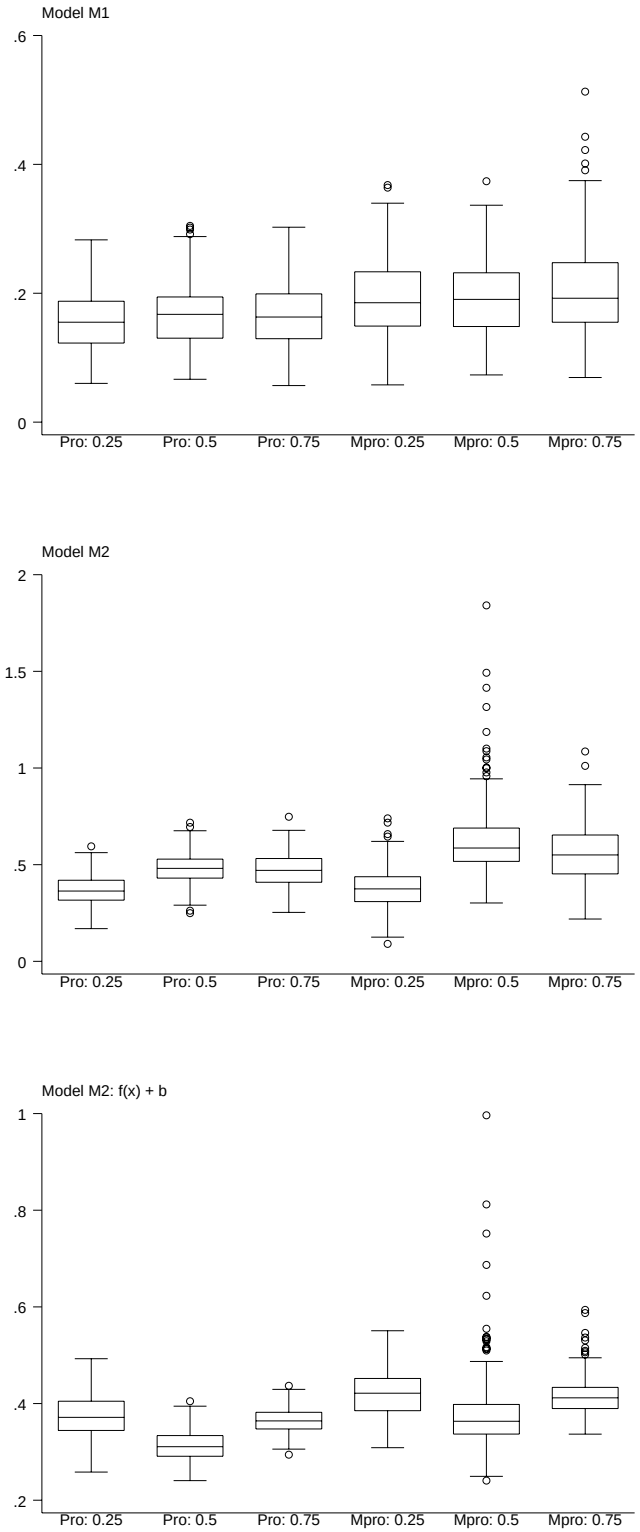

Figure 17: *Models M1 and M2: Boxplots of relative MSEs for f. The first to third boxplots refer to binary probit models and the fourth to sixth to multicategorical probit models. The comparison is based on the first category in multicategorical probit models for which the true predictor is identical to probit models. The third graph compares the MSEs for the sum of both effects (model M2 only).*

4 Conclusion

In this paper Bayesian generalized additive mixed models are studied through an extensive simulation study. The main results of the paper can be summarized as follows:

- Smooth functions and coefficients for individual specific random effects are well recovered provided that the covariate values do not only vary across individuals but also within individuals (models M1 and M3).
- If a covariate x varies only across individuals or clusters we are confronted with the following: The effect of x is split up into a smooth part $f(x)$ and an unsmooth part b_x (models M2 and M4). In this case the estimates for both the smooth and the unsmooth part are sometimes considerably biased. The sum of both effects, however, is estimated almost unbiased. Thus, the estimation procedure may have problems to separate correctly the two effects.
- The Bayesian credible intervals are rather conservative in the sense that the average coverage is usually above the nominal level.
- The differences in terms of bias and MSE between Gaussian responses, binary and multicategorical probit models are relatively small.

In this paper estimation is fully Bayesian with MCMC simulation techniques where the smooth functions, random effects coefficients and the smoothing parameters are estimated simultaneously. A comparison with an empirical Bayes approach where the smoothing parameters are estimated vias cross validation or as described in Lin and Zhang (1999) would be interesting but must be defered to future research.

References

Bernardinelli, L., D. Clayton, and C. Montomoli (1995). Bayesian Estimates of Disease Maps: How Important are Priors. *Statistics in Medicine 14*, 2411–2431.

Besag, J., Y. York, and A. Mollie (1991). Bayesian Image Restoration with two Applications in Spatial Statistics (with discussion). *Ann. Inst. Statist. Math. 43*, 1–59.

Biller, C. (2000). Adaptive Bayesian Regression Splines in Semiparametric Generalized Linear Models. *J. Comp. Stat. and Graph. Stat. 12*, 122–140.

Denison, D., B. Mallick, and A. Smith (1998). Automatic Bayesian Curve Fitting. *J. R. Statist. Soc. 60*, 333–350.

Fahrmeir, L. and S. Lang (2001a). Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors. *Appl. Statist. (JRSS C) (to appear)*.

Fahrmeir, L. and S. Lang (2001b). Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data. *Ann. Inst. Statist. Math. 53*, 11–30.

Gamerman, D. (1997). Efficient Sampling from the Posterior Distribution in Generalized Linear Models. *Statistics and Computing 7*, 57–68.

Hastie, T. and R. Tibshirani (2000). Bayesian Backfitting. *Statist. Sci. (to appear)*.

Heikkinen, J. and E. Arjas (1998). Nonparametric Bayesian Estimation of a Spatial Poisson Intensity. *Scand. J. Statist. 25*, 435–450.

Knorr-Held, L. (1999). Conditional Prior Proposals in Dynamic Models. *Scand. J. Statist. 26*, 129–144.

Knorr-Held, L. and J. Besag (1998). Modelling Risk from a Desease in Time and Space. *Statistics and Medicine 17*, 2045–2060.

Knorr-Held, L. and G. Raßer (2000). Bayesian Detection of Clusters and Discontinuities in Disease Maps. *Biometrics 56*, 13–21.

Lang, S. and A. Brezger (2001). Bayesian P-splines. SFB 386 Discussion Paper 236, University of Munich.

Lin, X. and D. Zhang (1999). Inference in Generalized Additive Mixed Models by using Smoothing Splines. *J. R. Statist. Soc. B 61*, 381–400.

Smith, M. and R. Kohn (1996). Nonparametric Regression using Bayesian Variable Selection. *Journal of Econometrics 75*, 317–343.

Smith, M., C. Wong, and R. Kohn (1998). Additive Nonparametric Regression with Autocorrelated Errors. *J. R. Statist. Soc. B 60*, 311–331.

Yau, P., R. Kohn, and S. Wood (2000). Bayesian Variable Selection and Model Averaging in High Dimensional Multinomial Nonparametric Regression. Technical report, University of NSW.