Kneib, Fahrmeir:

Supplement to "Structured additive regression for categorical space-time data: A mixed model approach"

Sonderforschungsbereich 386, Paper 431 (2005)

Supplement to
"Structured additive regression for categorical space-time data: A mixed model approach"

Thomas Kneib and Ludwig Fahrmeir

Department of Statistics, University of Munich.

1 Introduction

This technical report acts as a supplement to the paper "Structured additive regression for categorical space-time data: A mixed model approach" (Kneib and Fahrmeir, Biometrics, 2005, to appear). Details on several specific models for categorical responses are given as well as a description on how to construct design matrices in structured additive regression models. Furthermore some technical information on inferential issues and additional results from the simulation studies are provided. To ease orientation, sections in the supplement are named in analogy to the sections in the original paper. Also, formulas are presented with the same numbers.

2 Categorical response models

A general regression model for categorical responses $Y \in \{1, \dots, k\}$ can be defined in the context of multivariate generalized linear models via

$$\pi^{(r)} = P(Y = r) = h^{(r)}(\eta^{(1)}, \dots, \eta^{(q)}), \quad r = 1, \dots, q,$$

where $q = k - 1$ is the reference category, $\eta^{(r)} = v_r'\gamma$ is a predictor with appropriately defined design vector v_r, and γ is a vector of regression parameters. Defining $\pi = (\pi^{(1)}, \dots, \pi^{(q)})$, $\eta = (\eta^{(1)}, \dots, \eta^{(q)})$, $h = (h^{(1)}, \dots, h^{(q)})$ and the design matrix $V = (v_1, \dots, v_q)'$, the general model is

$$\pi = h(\eta), \quad \eta = V\gamma \tag{3}$$

with appropriately chosen multivariate response function $h : \mathbb{R}^q \to [0, 1]^q$. In the following we describe three specific models for categorial responses with unordered or ordered categories.

2.1 Models for nominal responses

The most common way to model categorical responses with unordered categories is the multinomial logit model, where, in analogy to the binary logit model, the response function is given by

$$P(Y=r)=\pi^{(r)}=h^{(r)}\left(\eta^{(1)},\dots,\eta^{(q)}\right)=\frac{\exp(\eta^{(r)})}{1+\sum_{s=1}^{q}\exp(\eta^{(s)})}$$

with linear predictor

$$\eta^{(r)}=u'\alpha^{(r)},$$

where u is a suitable vector of covariates and $\alpha^{(r)}$ is a category-specific vector of regression coefficients. Instead of defining a response function $h^{(r)}$, we can equivalently define the link function $g^{(r)}$

$$\eta^{(r)}=g^{(r)}(\pi^{(1)},\dots,\pi^{(q)})=\log\left(\frac{\pi^{(r)}}{1-\sum_{s=1}^{q}\pi^{(s)}}\right),$$

which is the inverse response function.

For a multinomial logit model, the general multinomial model (3) is obtained by defining the overall vector of regression coefficients

$$\gamma=(\alpha^{(1)\prime},\dots,\alpha^{(q)\prime})'$$

and the design matrix

$$V=\begin{pmatrix}v_1'\\ \vdots\\ v_q'\end{pmatrix}=\begin{pmatrix}u' & & 0\\ & \ddots & \\ 0 & & u'\end{pmatrix}.$$

In this classical multinomial logit model all covariates are assumed to be independent of the category while effects are category-specific. Extensions of the classic model allow for the inclusion of category-specific covariates w_q leading to the predictor

$$\eta^{(r)}=u'\alpha^{(r)}+w_r'\zeta.$$

Note that in this case, the regression coefficients ζ of category-specific effects are global, i.e. they are the same for all categories. Extensions with category-specific covariates can also be easily cast into the general form by modifying the design matrix to

$$V=\begin{pmatrix}u' & & 0 & w_1'-w_k'\\ & \ddots & & \\ 0 & & u' & w_q'-w_k'\end{pmatrix},$$

where k is the reference category, and extending the vector of regression coefficients to

$$\gamma=(\alpha^{(1)\prime},\dots,\alpha^{(q)\prime},\zeta')'.$$

Such extensions can also be defined for structured additive regression models and will be added to our implementation in a future version of BayesX.

2.2 Cumulative (threshold) models for ordinal responses

If the categorical responses can be ordered, specific models for this situation have to be employed. Such models for ordinal responses are commonly defined via a cumulative distribution function F and

$$P(Y \leq r) = F(\eta^{(r)}).$$

Therefore the response function is given by

$$\begin{aligned} P(Y=r) &= \pi^{(r)} = h^{(r)}\left(\eta^{(1)},\ldots,\eta^{(q)}\right) = F(\eta^{(r)}) - F(\eta^{(r-1)}), \quad r>1 \\ P(Y=1) &= \pi^{(1)} = h^{(1)}\left(\eta^{(1)},\ldots,\eta^{(q)}\right) = F(\eta^{(1)}) \end{aligned}$$

with linear predictor

$$\eta^{(r)} = \theta^{(r)} - u'\alpha.$$

Again, u and α denote covariates and regression coefficients, respectively and $\theta^{(1)} < \ldots < \theta^{(q)}$ are ordered thresholds. In contrast to the multinomial logit model, both the covariates and the regression coefficients other than the thresholds are assumed to be fixed for all categories.

Here, the overall vector of regression coefficients in (3) is given by

$$\gamma = (\theta^{(1)},\ldots,\theta^{(q)},\alpha')'$$

and the corresponding design matrix is

$$V = \begin{pmatrix} v_1' \\ \vdots \\ v_q' \end{pmatrix} = \begin{pmatrix} 1 & & & -u' \\ & \ddots & & \vdots \\ & & 1 & -u' \end{pmatrix}$$

Extensions of the basic predictor $\theta^{(r)} - u'\alpha$ allow for thresholds depending on covariates w or in other words, allow for category-specific effects, i.e.

$$\eta^{(r)} = \theta^{(r)} - w\zeta^{(r)} - u'\alpha,$$

where $\theta^{(r)} - w\zeta^{(r)}$ can be interpreted as a covariate-dependent threshold. Though being easily defined, these extensions lead to models of considerably increased complexity, since the order restrictions $\theta^{(1)} < \ldots < \theta^{(q)}$ have now to be fulfilled for the covariate-dependent thresholds, i.e.

$$\theta^{(1)} - w\zeta^{(1)} < \ldots < \theta^{(q)} - w\zeta^{(q)}$$

for all possible values of the covariates w.

Covariate-dependent thresholds can be cast in the general model (3) by defining

$$V = \begin{pmatrix} 1 & -w' & & & & -u' \\ & & \ddots & & & \vdots \\ & & & 1 & -w' & -u' \end{pmatrix}$$

and

$$\gamma = (\theta^{(1)}, \zeta^{(1)}, \ldots, \theta^{(q)}, \zeta^{(q)}, \alpha')'.$$

2.3 Sequential models for ordinal responses

A second possibility to model ordered responses are sequential models. In contrast to cumulative models, sequential models assume that the categories r can only be achieved successively. This leads to a model for the conditional probabilities:

$$P(Y = r | Y \geq r) = F(\theta^{(r)} - u'\alpha), \quad r = 1, \ldots, q$$

with cumulative distribution function F, covariates u, regression coefficients α and thresholds $\theta^{(1)}, \ldots, \theta^{(q)}$. Note that, in contrast to cumulative models, no ordering restriction is needed for the thresholds in sequential models. The response function is obtained as

$$P(Y = r) = \pi^{(r)} = h^{(r)}\left(\eta^{(1)}, \ldots, \eta^{(q)}\right) = F(\eta^{(r)}) \prod_{s=1}^{r-1} (1 - F(\eta^{(s)}))$$

with linear predictor

$$\eta^{(r)} = \theta^{(r)} - u'\alpha.$$

In this case, the overall vector of regression coefficients and the design matrix are equal to those obtained for cumulative responses:

$$\gamma = (\theta^{(1)}, \ldots, \theta^{(q)}, \alpha')'$$

and

$$V = \begin{pmatrix} v_1' \\ \vdots \\ v_q' \end{pmatrix} = \begin{pmatrix} 1 & & & -u' \\ & \ddots & & \vdots \\ & & 1 & -u' \end{pmatrix}.$$

Extensions with covariate-dependent thresholds can be defined in complete analogy to the cumulative case.

2.4 Categorical STAR models

Categorical structured additive regression models extend the models presented in subsections 2.1 to 2.3 through the inclusion of nonparametric effects, spatial effects and further extensions. For example, a space-time main effects model for nominal responses can be defined by

$$\eta_{it}^{(r)} = u_{it}'\alpha^{(r)} + f_1^{(r)}(x_{it1}) + \ldots + f_l^{(r)}(x_{itl}) + f_{time}^{(r)}(t) + f_{spat}^{(r)}(s_i). \tag{4}$$

Here, $f_{time}^{(r)}$ and $f_{spat}^{(r)}$ represent possibly nonlinear effects of time and space, $f_1^{(r)}, \ldots, f_l^{(r)}$ are smooth functions of the continuous covariates $x_1, \ldots, x_l$, and $u'\alpha^{(r)}$ corresponds to the usual parametric linear part of the predictor. It turns out that all unknown functions as well as extensions can be expressed as the product of appropriately defined design vectors and regression coefficients. Thus, we can always rewrite predictor (4) and extended forms as

$$\eta_{it}^{(r)} = u_{it}'\alpha^{(r)} + \sum_{j=1}^{p} z_{itj}'\beta_j^{(r)}. \tag{5}$$

In complete analogy, we can extend the linear predictor in the general multivariate model and in any of its subclasses such as cumulative or sequential models to a structured additive predictor. The general form (3) extends to

$$\eta_{it} = V_{it}\gamma + \sum_{j=1}^{p} Z_{itj}\delta_j \tag{6}$$

The matrices in (6) are constructed in a similar way as in the purely parametric models in subsections 2.1 to 2.3. The vector γ comprises fixed effects for nominal models and fixed effects and thresholds for cumulative and sequential models. Therefore V_{it} can be defined in analogy to V above. For models with nominal response we have

$$V_{it} = \begin{pmatrix} u_{it}' & & 0 \\ & \ddots & \\ 0 & & u_{it}' \end{pmatrix} \quad \text{and} \quad Z_{itj} = \begin{pmatrix} z_{itj}' & & 0 \\ & \ddots & \\ 0 & & z_{itj}' \end{pmatrix},$$

with regression coefficients

$$\gamma = (\alpha^{(1)\prime}, \ldots, \alpha^{(q)\prime})'$$

and

$$\delta_j = (\beta_j^{(1)\prime}, \ldots, \beta_j^{(q)\prime})'.$$

In cumulative and sequential models the design matrices are defined by

$$V_{it} = \begin{pmatrix} 1 & & & -u_{it}' \\ & \ddots & & \vdots \\ & & 1 & -u_{it}' \end{pmatrix} \quad \text{and} \quad Z_{itj} = \begin{pmatrix} -z_{itj}' \\ \vdots \\ -z_{itj}' \end{pmatrix},$$

and the regression coefficients are given by

$$\gamma = (\theta^{(1)}, \ldots, \theta^{(q)}, \alpha')'$$

and $\delta_j = \beta_j$.

Note that the presented structured additive regression models do not include extensions with category-specific covariates for the multinomial logit model or covariate-dependent thresholds for ordinal response models. However, such extensions could easily be included in a similar way as in the basic parametric models presented in sections 2.1 to 2.3.

3 Inference

3.1 Mixed model representation

To estimate structured additive regression models based on mixed model methodology, the original model is reparameterised based on the decomposition

$$\beta_j = Z_j^{unp}\beta_j^{unp} + Z_j^{pen}\beta_j^{pen}, \tag{13}$$

where the index r is omitted for notational simplicity. Choosing special matrices Z_j^{unp} and Z_j^{pen} in this decomposition leads to a variance components model. In general, these matrices (which have to fulfil requirements (i) to (iv) formulated in Kneib and Fahrmeir (2005)) can be obtained as follows: Z_j^{unp} contains a $d_j - k_j$ dimensional basis of the null space of K_j. Therefore requirement (iii) is automatically fulfilled. Z_j^{pen} can be obtained by $Z_j^{pen} = L_j(L_j'L_j)^{-1}$ where the full column rank $d_j \times k_j$ matrix L_j is determined by the decomposition of the penalty matrix K_j into $K_j = L_jL_j'$. This ensures requirements (i) and (iv). If we choose L_j such that $L_j'Z_j^{unp} = 0$ and $Z_j^{unp}L_j' = 0$ hold, we finally obtain requirement (ii). The decomposition $K_j = L_jL_j'$ of the penalty matrix can be based on the spectral decomposition $K_j = \Gamma_j\Omega_j\Gamma_j'$. The $(k_j \times k_j)$ diagonal matrix Ω_j contains the positive eigenvalues ω_{jm}, $m = 1, \ldots, k_j$, of K_j in descending order, i.e. $\Omega_j = \mathrm{diag}(\omega_{j1}, \ldots, \omega_{j,k_j})$. Γ_j is a $(d_j \times k_j)$ orthogonal matrix of the corresponding eigenvectors. From the spectral decomposition we can choose $L_j = \Gamma_j\Omega_j^{1/2}$. Note, that the factor L_j is not unique and in many cases numerical superior factorizations exist.

Decomposition (13) leads to the following predictor for categorical STAR models:

$$\eta = V\gamma + \sum_{j=1}^{p} \tilde{Z}_j^{unp}\delta_j^{unp} + \sum_{j=1}^{p} \tilde{Z}_j^{pen}\delta_j^{pen}. \tag{15}$$

To obtain the design matrices in this predictor, we proceed in a similar way as in subsection 2.4. For nominal responses we have

$$V_{it} = \begin{pmatrix} u_{it}' & & 0 \\ & \ddots & \\ 0 & & u_{it}' \end{pmatrix}, \quad \tilde{Z}_{itj}^{unp} = \begin{pmatrix} z_{itj}^{unp\prime} & & 0 \\ & \ddots & \\ 0 & & z_{itj}^{unp\prime} \end{pmatrix} \quad \text{and} \quad Z_{itj}^{pen} = \begin{pmatrix} z_{itj}^{pen\prime} & & 0 \\ & \ddots & \\ 0 & & z_{itj}^{pen\prime} \end{pmatrix}$$

with regression coefficients

$$\gamma = (\alpha^{(1)\prime}, \ldots, \alpha^{(q)\prime})', \quad \delta_j^{unp} = (\beta_j^{unp(1)\prime}, \ldots, \beta_j^{unp(q)\prime})' \quad \text{and} \quad \delta_j^{pen} = (\beta_j^{pen(1)\prime}, \ldots, \beta_j^{pen(q)\prime})'.$$

Similarly, for cumulative and sequential response models we have

$$V_{it} = \begin{pmatrix} 1 & & -u_{it}' \\ & \ddots & \vdots \\ & 1 & -u_{it}' \end{pmatrix}, \quad Z_{itj}^{unp} = \begin{pmatrix} -z_{itj}^{unp\prime} \\ \vdots \\ -z_{itj}^{unp\prime} \end{pmatrix}, \quad \text{and} \quad Z_{itj}^{pen} = \begin{pmatrix} -z_{itj}^{pen\prime} \\ \vdots \\ -z_{itj}^{pen\prime} \end{pmatrix},$$

and

$$\gamma = (\theta^{(1)}, \ldots, \theta^{(q)}, \alpha')', \quad \delta_j^{unp} = \beta_j^{unp} \quad \text{and} \quad \delta_j^{pen} = \beta_j^{pen}.$$

From this expressions equation (15) is yielded by defining the stacked vectors and matrices $\eta = (\eta_{it})$, $V = (V_{it})$, $\tilde{Z}_j^{unp} = (\tilde{Z}_{itj}^{unp})$ and $\tilde{Z}_j^{pen} = (\tilde{Z}_{itj}^{pen})$.
The covariance matrix Λ of the vector of penalized regression coefficients δ^{pen} is given by

$$\Lambda = \text{blockdiag}((\tau_1^{(1)})^2 I, \ldots, (\tau_1^{(q)})^2 I, \ldots, (\tau_p^{(1)})^2 I, \ldots, (\tau_p^{(q)})^2 I)$$

for nominal responses and

$$\Lambda = \text{blockdiag}(\tau_1^2 I, \ldots, \tau_p^2 I)$$

for cumulative and sequential models.

3.2 Empirical Bayes inference for categorical mixed models

The matrix of working weights $W = DS^{-1}D$ has a block diagonal structure defined by the block diagonal matrices $D = \text{blockdiag}(D_{11} \ \ldots \ D_{nT})$ and $S = \text{blockdiag}(S_{11} \ \ldots \ S_{nT})$, the $q \times q$ matrices

$$D_{it} = \frac{\partial h(\eta_{it})}{\partial \eta} = \begin{pmatrix} \frac{\partial h^{(1)}(\eta_{it})}{\partial \eta^{(1)}} & \cdots & \frac{\partial h^{(q)}(\eta_{it})}{\partial \eta^{(1)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial h^{(1)}(\eta_{it})}{\partial \eta^{(q)}} & \cdots & \frac{\partial h^{(q)}(\eta_{it})}{\partial \eta^{(q)}} \end{pmatrix}$$

and

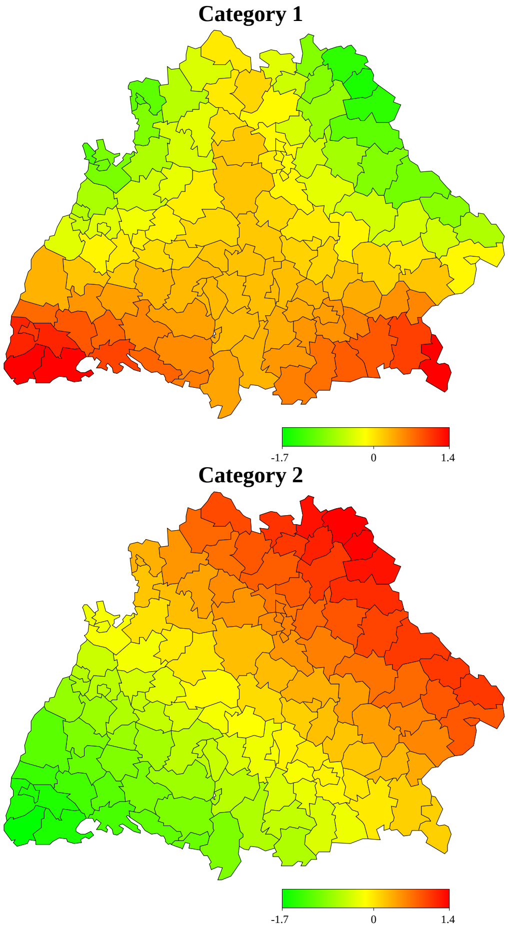
$$S_{it} = \text{cov}(y_{it}) = \begin{pmatrix} \pi_{it}^{(1)}(1-\pi_{it}^{(1)}) & -\pi_{it}^{(1)}\pi_{it}^{(2)} & \cdots & -\pi_{it}^{(1)}\pi_{it}^{(q)} \\ -\pi_{it}^{(1)}\pi_{it}^{(2)} & \ddots & & \vdots \\ \vdots & & \ddots & -\pi_{it}^{(q-1)}\pi_{it}^{(q)} \\ -\pi_{it}^{(1)}\pi_{it}^{(q)} & \cdots & -\pi_{it}^{(q-1)}\pi_{it}^{(q)} & \pi_{it}^{(q)}(1-\pi_{it}^{(q)}) \end{pmatrix}$$

The working observations $\tilde{y}$ are defined by

$$\tilde{y} = \hat{\eta} + (D^{-1})'(y - \pi).$$

4 Simulation studies

To investigate performance, we conducted several simulation studies based on a multinomial logit model and a cumulative probit model with three categories and predictors defined to be the sum of a nonparametric effect and a spatial effect (see Figures 1 and 2 for a detailed description of the simulation design). Here, we will describe the results of the simulation studies in more detail than in the original paper.



- Predictor:

$$\eta_i^{(r)} = f_1^{(r)}(x_i) + f_2^{(r)}(s_i)$$

- Category 1:

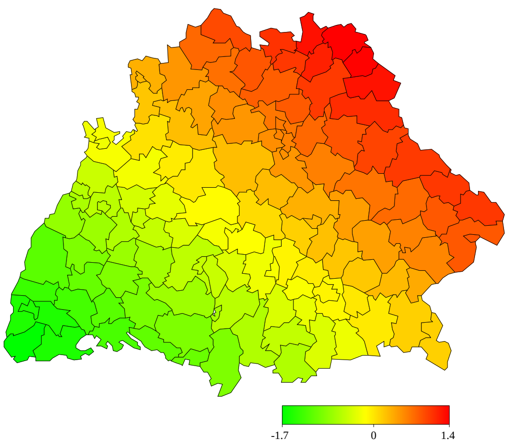
$$\begin{aligned} f_1^{(1)}(x) &= \sin[\pi(2x-1)] \\ f_2^{(1)}(s) &= -0.75|s_x|(0.5+s_y) \end{aligned}$$

- Category 2:

$$\begin{aligned} f_1^{(2)}(x) &= \sin[2\pi(2x-1)] \\ f_2^{(2)}(s) &= 0.5(s_x+s_y) \end{aligned}$$

- x is chosen from an equidistant grid of 100 values between -1 and 1.
- (s_x, s_y) are the centroids of the 124 districts s of the two southern states of Germany (see Figures).

Figure 1: Simulation design for the multinomial logit model.



- Predictor:

$$\eta_i^{(r)} = \theta^{(r)} - f_1(x_i) - f_2(s_i)$$

- Functions:

$$\begin{aligned} f_1(x) &= \sin[\pi(2x-1)] \\ f_2(s) &= 0.5(s_x+s_y) \end{aligned}$$

- x is chosen from an equidistant grid of 100 values between -1 and 1.
- (s_x, s_y) are the centroids of the 124 districts s of the two southern states of Germany (see Figure).

Figure 2: Simulation design for the cumulative probit model.

4.1 Comparison of different modelling approaches

The first aim was to compare different parameterisations of the spatial effect and different approaches to the estimation of categorical STAR models. Therefore 250 simulation runs with $n = 500$ observations were generated from the multinomial logit model described in Figure 1. We used cubic P-splines with second order random walk penalty and 20 knots to estimate effects of the continuous covariate. The spatial effect was estimated either by a Markov random field, a (full) Gaussian random field or a two-dimensional P-spline (based on 10×10 inner knots). For the competing fully Bayesian approach by Fahrmeir and Lang (2001b) and Brezger and Lang (2005), where inverse Gamma priors $IG(a, b)$ with $a = b = 0.001$ are assigned to the variances, the GRF approach was computationally to demanding due to the inversion of a full precision matrix for the spatial effect in each iteration. Therefore we excluded the fully Bayesian GRF approach from the comparison. As a further competitor we utilized the R-implementation of the procedure polyclass described in Kooperberg, Bose and Stone (1997). Here, nonparametric effects and interaction surfaces are modelled by linear splines and their tensor-products. Smoothness of the estimated curves is not achieved by penalization but via stepwise inclusion and deletion of model terms corresponding to basis functions based on AIC.

The results of the simulation study can be summarized as follows:

- Generally REML estimates have somewhat smaller median MSE than their fully Bayesian counterparts, with larger differences for spatial effects (see Figures 3a to 3d).
- Estimates for the effects of the continuous covariate are rather insensitive with respect to the model choice for the spatial effect (Figures 3a and 3b).
- Two-dimensional P-splines lead to the best fit for the spatial effect although data is provided with discrete spatial information (Figures 3c and 3d).
- Polyclass is outperformed by both the empirical and the fully Bayesian approach and therefore results are deferred to Figure 3e together with REML estimates based on two-dimensional P-splines. Presumably, the poor performance of polyclass is mainly caused by the special choice of linear splines, resulting in rather peaked estimates. Smoother basis functions, e.g. truncated cubic polynomials might improve the fit substantially but are not available in the implementation.
- Empirical and fully Bayesian estimates lead to comparable bias for the nonparametric effects. Results for function $f_1(x)$ obtained with polyclass are less biased but show some peaks caused by the modelling with linear splines. Therefore we can conclude that the poor performance of polyclass in terms of MSE is mainly introduced

by additional variance compared to empirical and fully Bayesian estimates (Figure 4).

- For spatial effects both empirical and fully Bayesian estimates tend to oversmooth the data, i.e. estimates are too small for high values of the spatial functions and vice versa. In contrast, polyclass leads to estimates which are too wiggly and therefore overestimate spatial effects (Figures 5 and 6).
- For some simulation runs with spatial effects modelled by MRFs, no convergence of the REML algorithm could be achieved. This was also the case if the spatial effect was modelled by a two-dimensional P-spline but in a much smaller number of cases. Obviously the same convergence problems as described in Fahrmeir, Kneib and Lang (2004) appear in a categorical setting. However, the arguments given there still hold and so we again used estimates obtained from the final (100th) iteration.

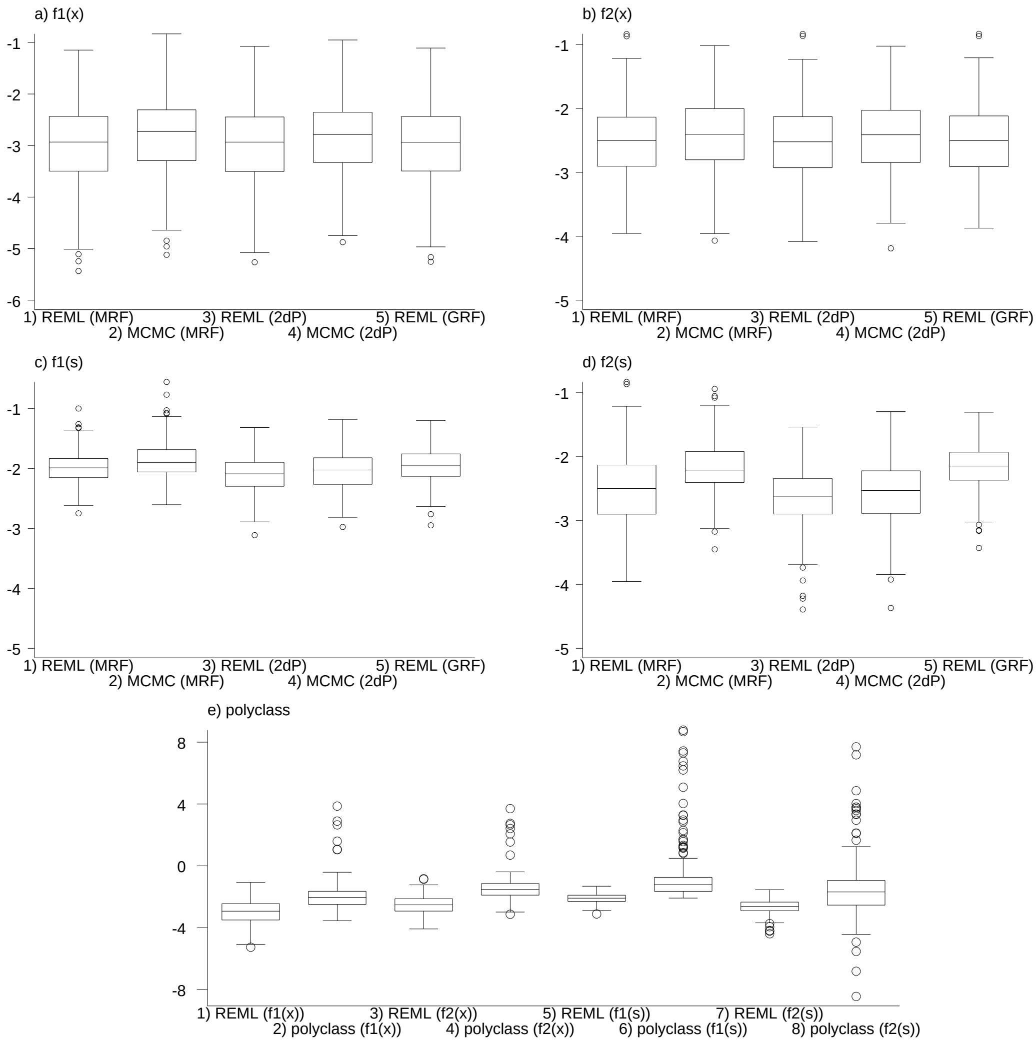


Figure 3: Comparison of different modelling approaches: Boxplots of log(MSE).

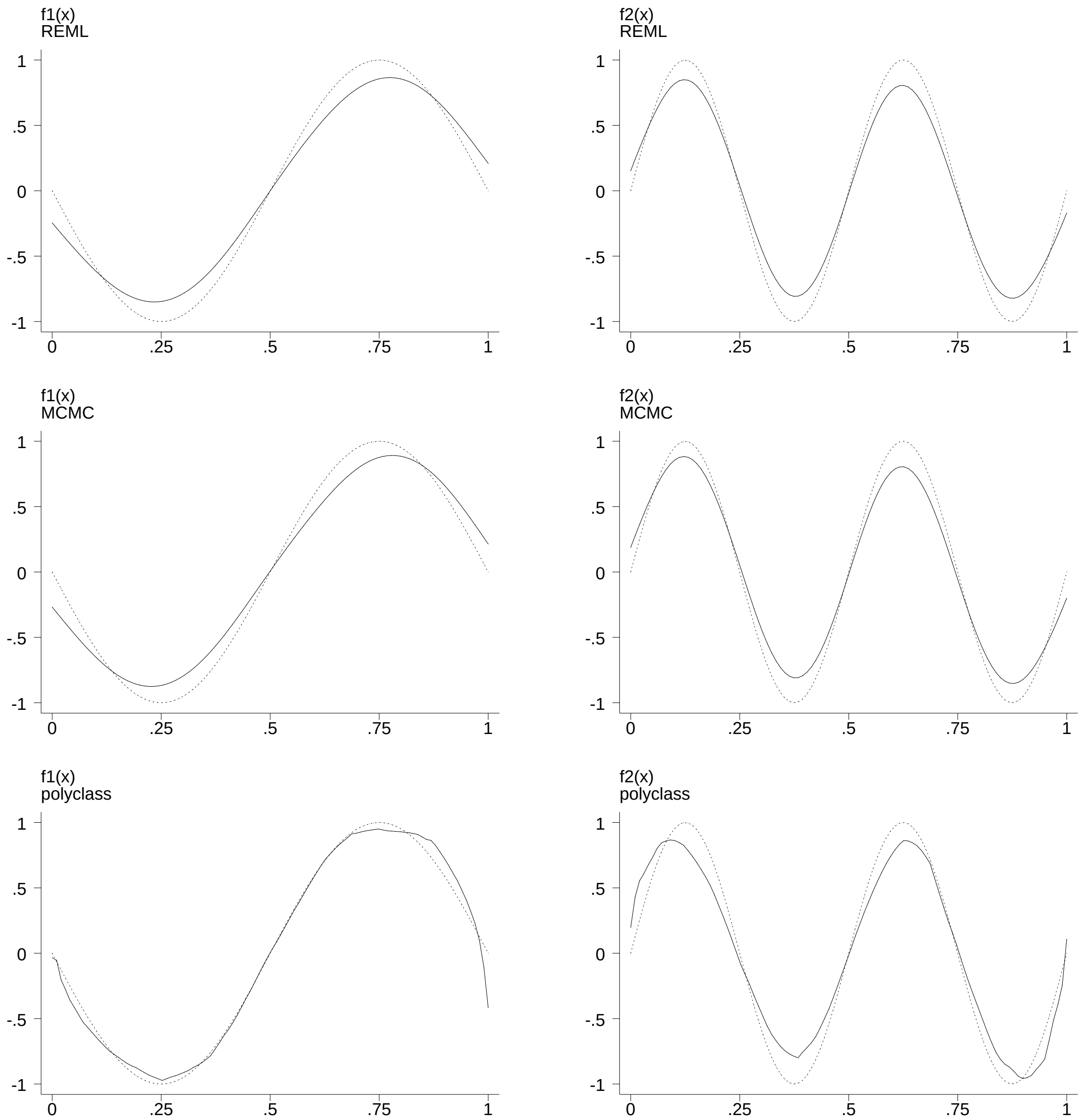


Figure 4: Comparison of different modelling approaches: Bias of nonparametric estimates.

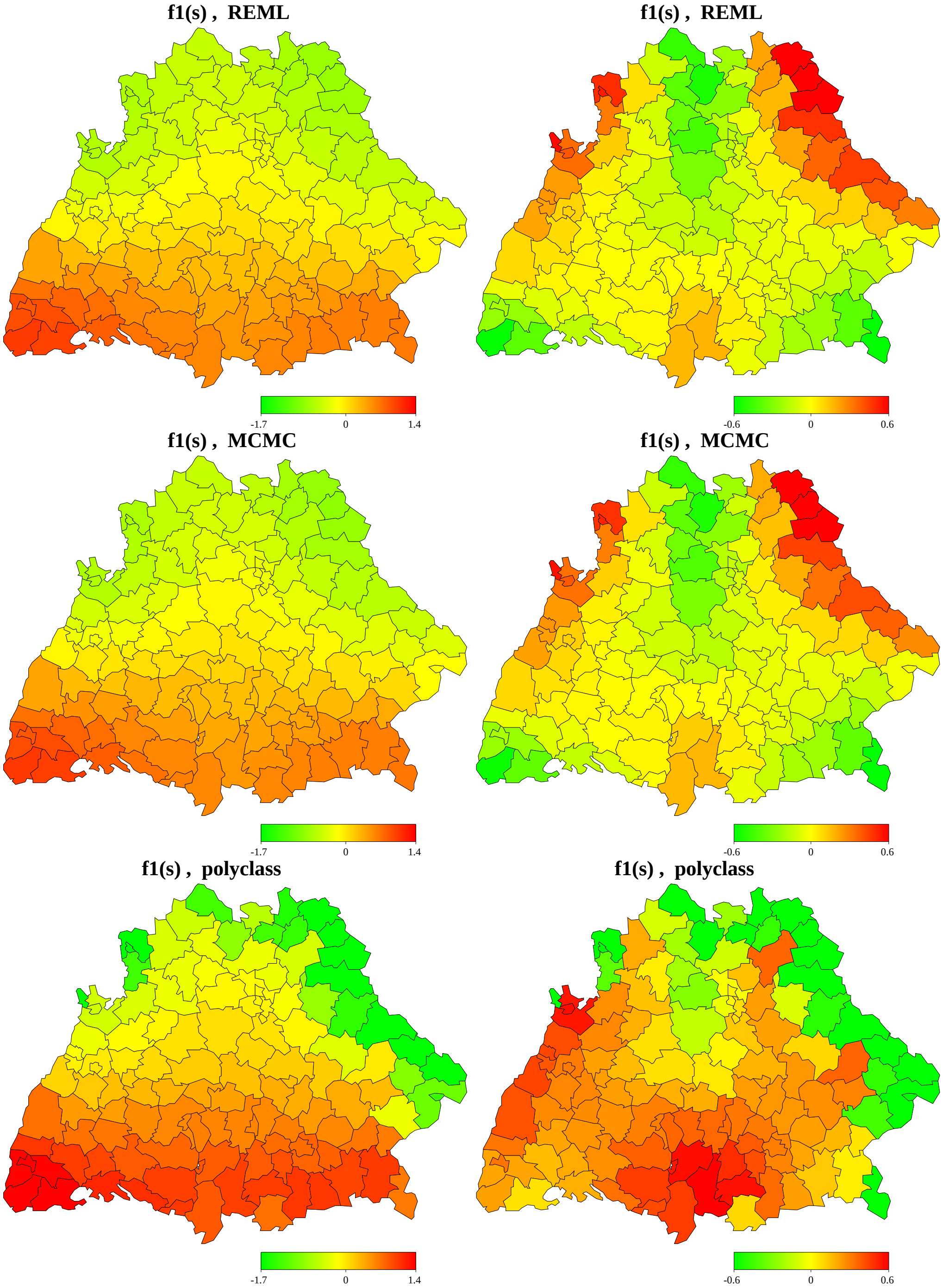


Figure 5: Comparison of different modelling approaches: Mean (left panel) and bias (right panel) of estimates for $f_1(s)$.

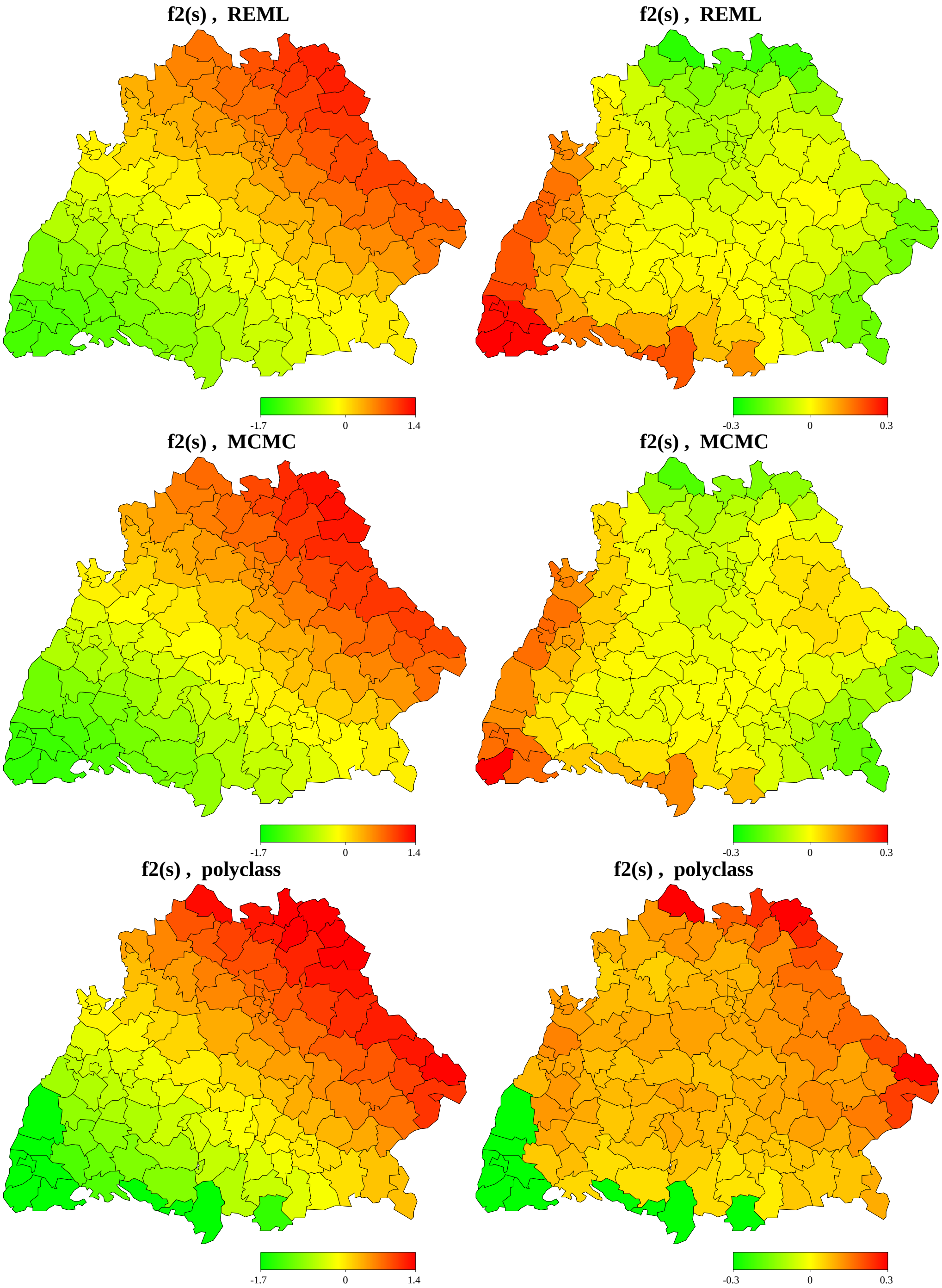


Figure 6: Comparison of different modelling approaches: Mean (left panel) and bias (right panel) of estimates for $f_2(s)$.

4.2 Bias of REML estimates

It is frequently argued that results from REML estimation procedures in GLMMs tend to be biased due to the Laplace approximation involved, especially in sparse data situations (compare e.g. Lin and Breslow (1996)). Therefore, as a second aim, we investigated whether this observation holds in a categorical setting in a second simulation study, based on the models described in Figures 1 and 2 with different sample sizes, namely $n = 500$, $n = 1000$ and $n = 2000$. Results from the REML estimation procedure were compared to their fully Bayesian counterparts since these estimates do not use any approximations but work with the exact posterior. For both approaches, the spatial effect was estimated by a MRF while nonparametric effects were again modelled by cubic P-splines with second order random walk penalty and 20 inner knots.

The results of the simulation lead to the following conclusions:

- In general, bias is smaller for MCMC estimates, most noticeably for more wiggly functions. For increasing sample sizes, differences almost vanish and both approaches give nearly unbiased estimates (Figures 7 to 12).
- REML estimates perform superior to MCMC estimates in terms of MSE (Figures 13 and 14).

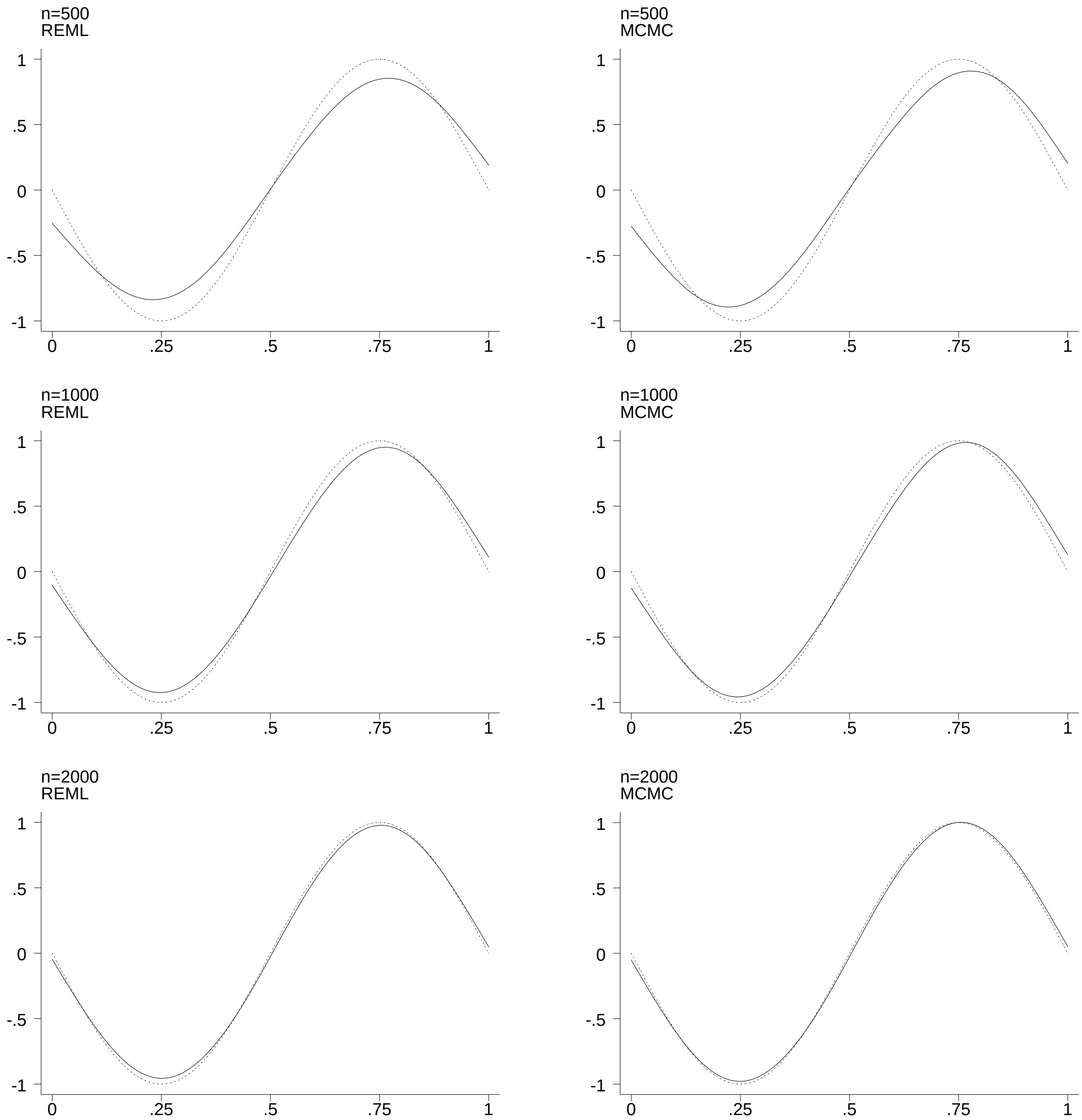


Figure 7: Multinomial logit model: Bias of nonparametric estimates for $f_1(x)$.

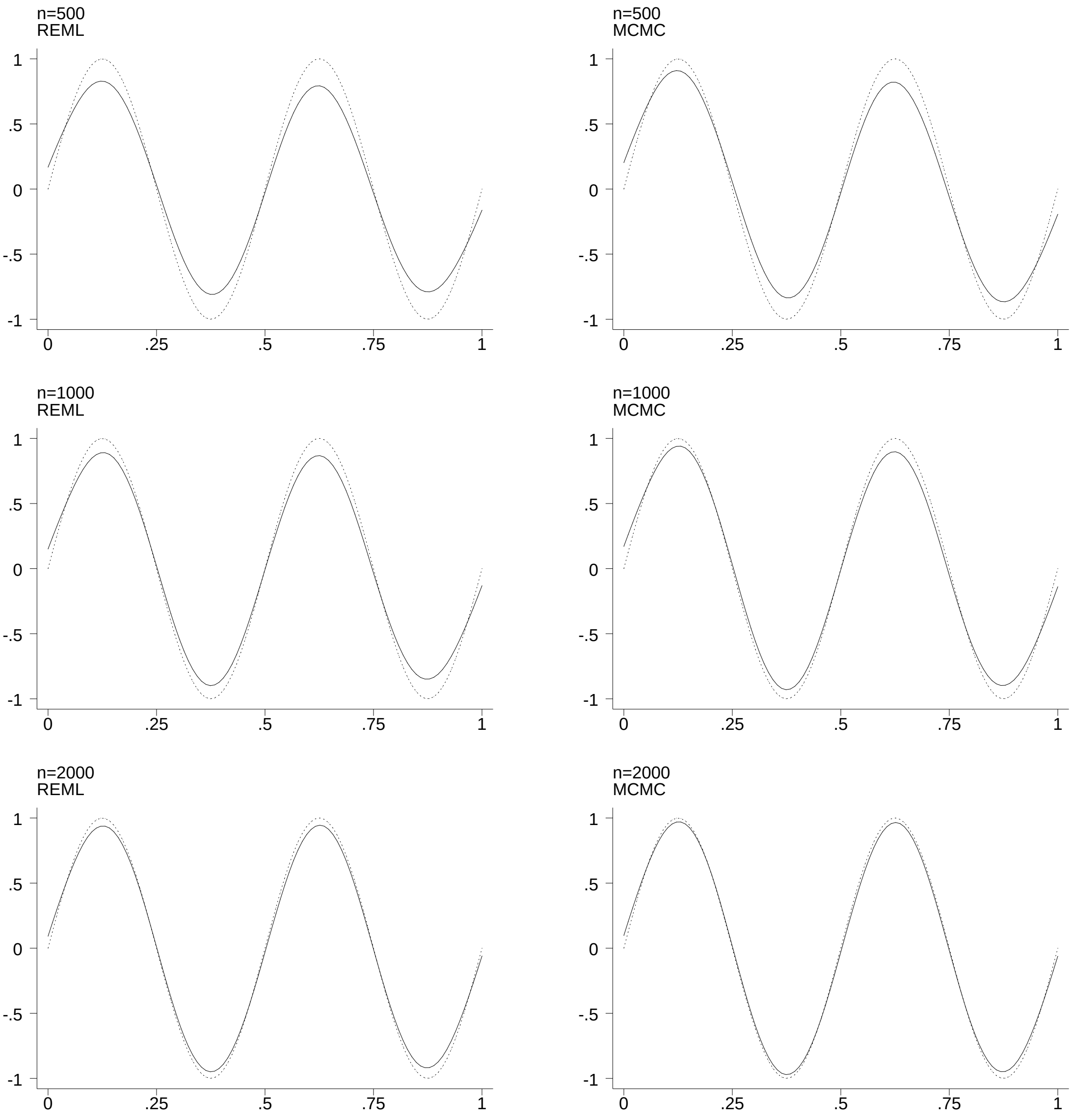


Figure 8: Multinomial logit model: Bias of nonparametric estimates for $f_2(x)$.

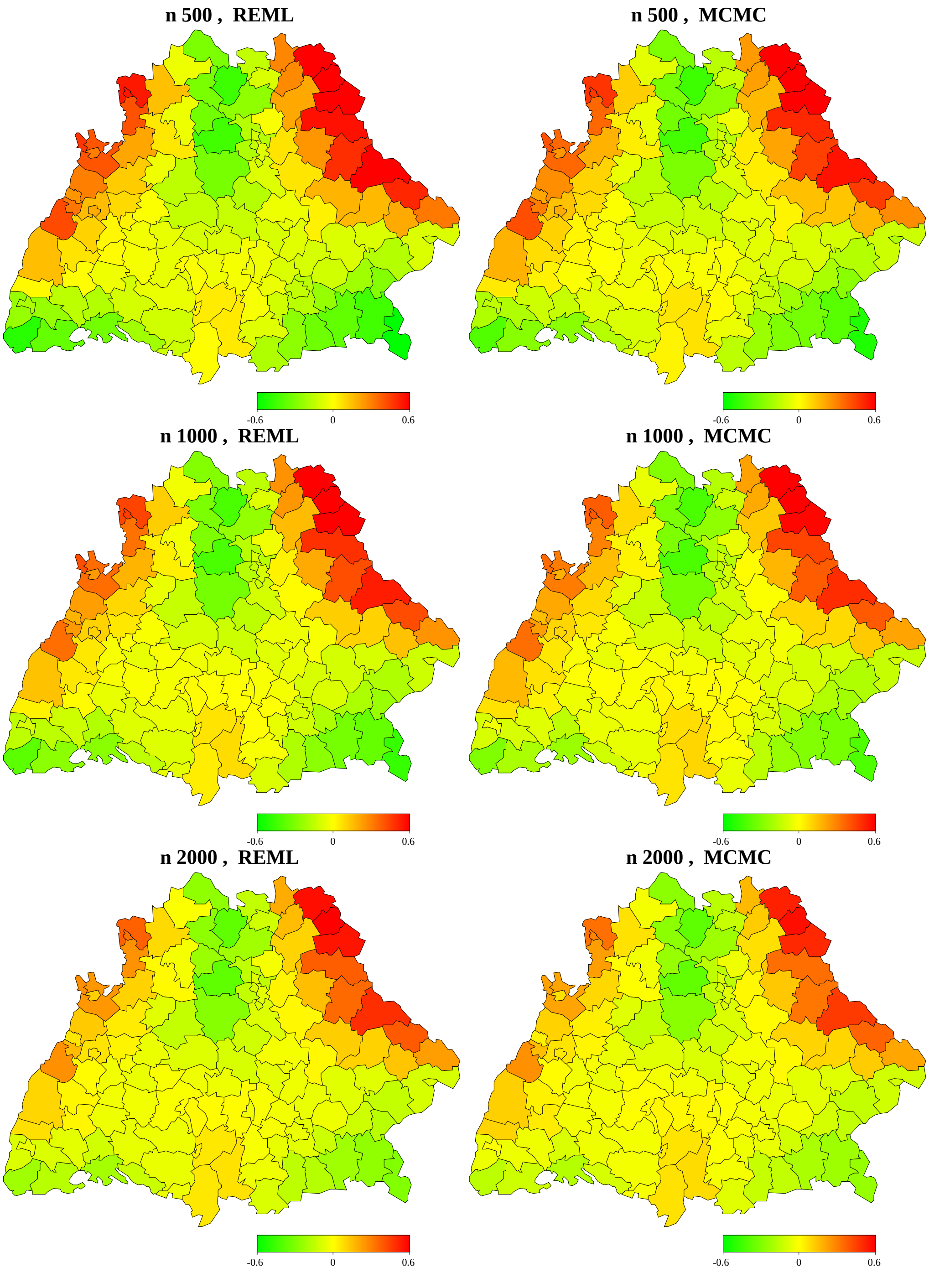


Figure 9: Multinomial logit model: Bias of spatial estimates for $f_1(s)$.

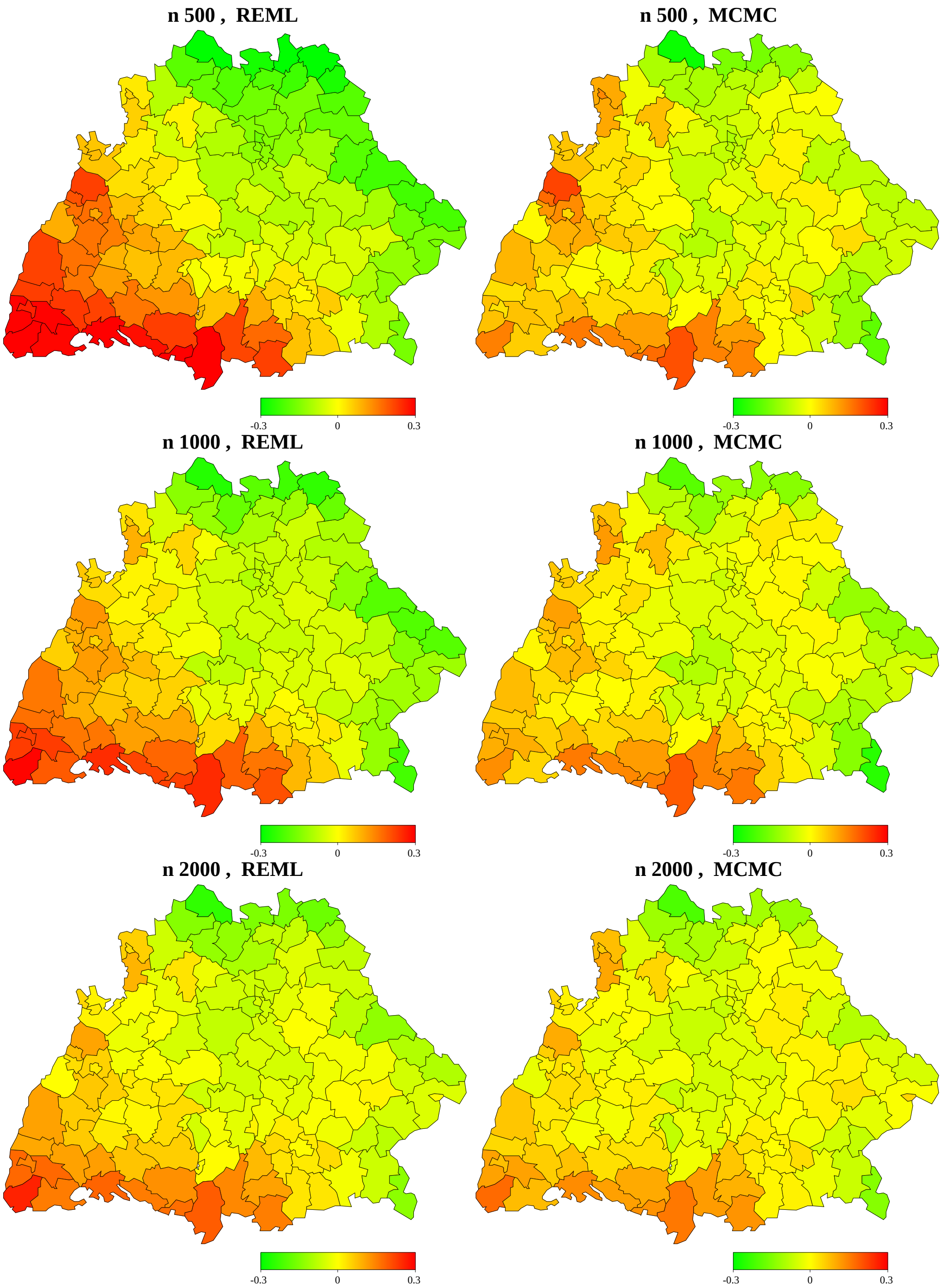


Figure 10: Multinomial logit model: Bias of spatial estimates for $f_2(s)$.

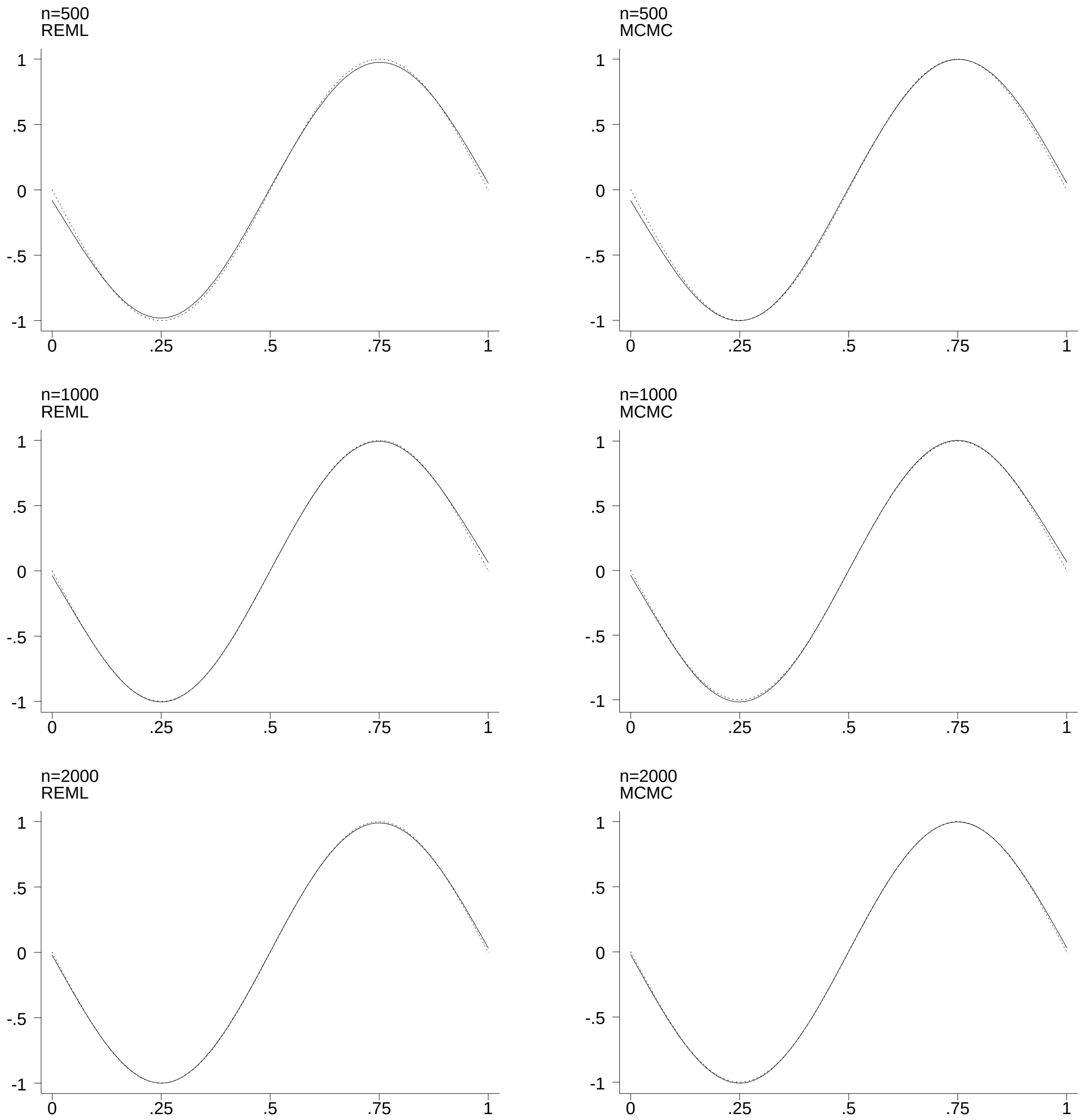


Figure 11: Cumulative probit model: Bias of nonparametric estimates for $f(x)$.

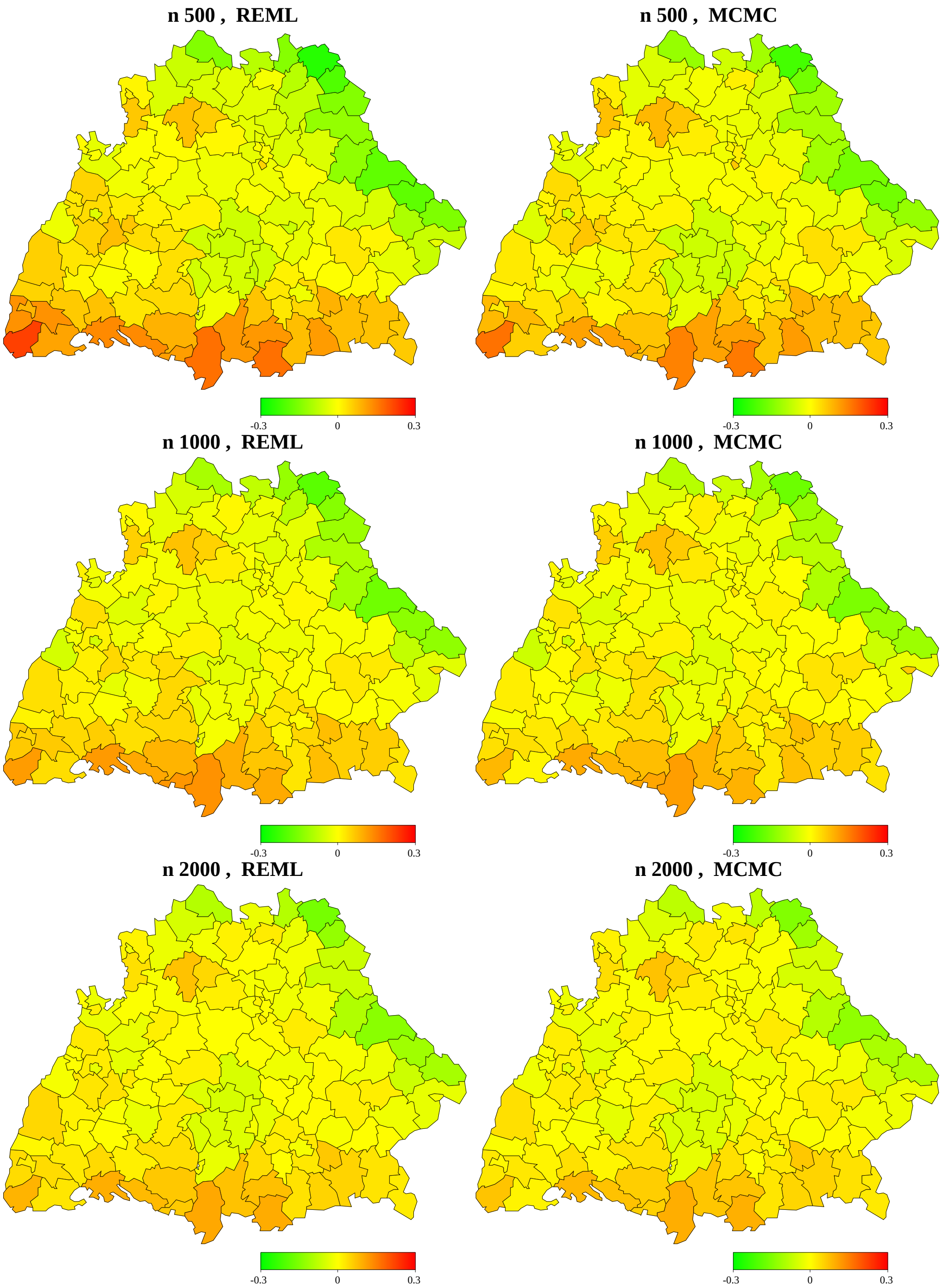


Figure 12: Cumulative probit model: Bias of spatial estimates for $f(s)$.

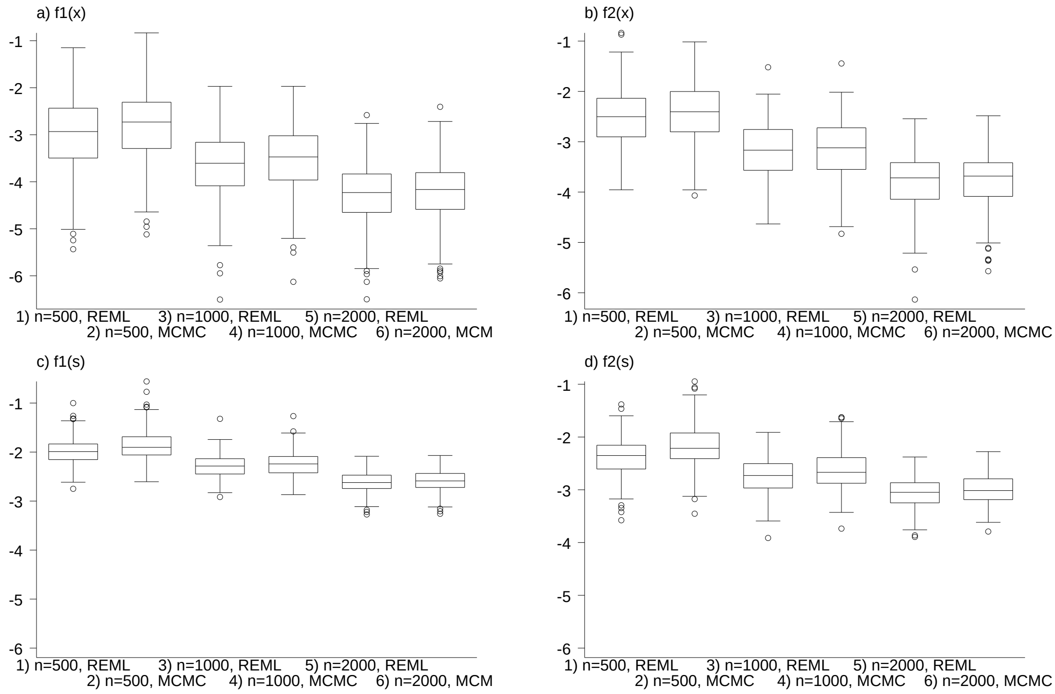


Figure 13: Multinomial logit model: Boxplots of log(MSE) for different sample sizes.

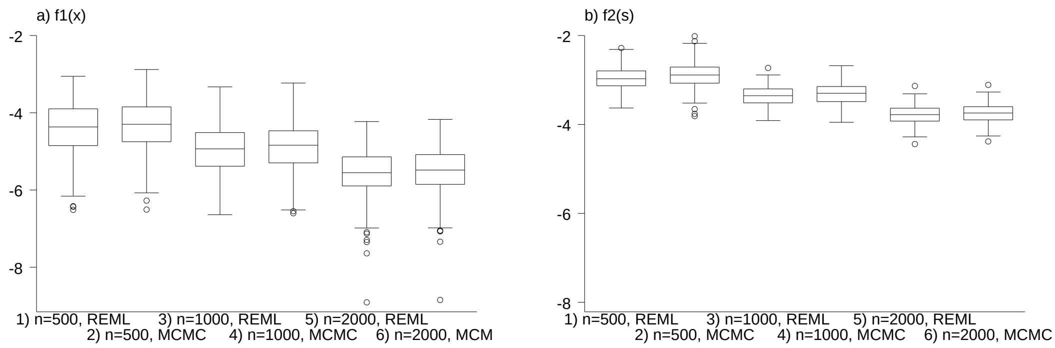


Figure 14: Cumulative probit model: Boxplots of log(MSE) for different sample sizes.

References

Brezger, A. and Lang, S., 2005: Generalized additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis* (to appear).

Fahrmeir, L., Kneib, T. and Lang, S., 2004: Penalized Structured Additive Regression for Space-Time Data: a Bayesian Perspective. *Statistica Sinica*, 14, 731–761.

Fahrmeir, L. and Lang, S., 2001b: Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data. *Annals of the Institute of Statistical Mathematics*, 53, 10–30

Kooperberg, C., Bose, S. and Stone, C.J., 1997: Polychotomous Regression. *Journal of the American Statistical Association*, 92, 117–127.

Lin, X. and Breslow, N. E., 1996: Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association*, 91, 1007–1016.