

Czado, Claudia; Prokopenko, Sergiy

Working Paper

Modeling transport mode decisions using hierarchical binary spatial regression models with cluster effects

Discussion Paper, No. 406

Provided in Cooperation with:

Collaborative Research Center (SFB) 386: Statistical Analysis of discrete structures - Applications in Biometrics and Econometrics, University of Munich (LMU)

Suggested Citation: Czado, Claudia; Prokopenko, Sergiy (2004) : Modeling transport mode decisions using hierarchical binary spatial regression models with cluster effects, Discussion Paper, No. 406, Ludwig-Maximilians-Universität München, Sonderforschungsbereich 386 - Statistische Analyse diskreter Strukturen, München, <https://doi.org/10.5282/ubm/epub.1774>

This Version is available at:

<https://hdl.handle.net/10419/31084>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Modeling Transport Mode Decisions Using Hierarchical Binary Spatial Regression Models with Cluster Effects

Claudia Czado ¹

Sergiy Prokopenko ¹

Center for Mathematical Sciences, Munich University of Technology

Boltzmannstr.3, D-85747 Garching, Germany

email: cczado@ma.tum.de

Abstract

This work is motivated by a mobility study conducted in the city of Munich, Germany. The variable of interest is a binary response, which indicates whether public transport has been utilized or not. One of the central questions is to identify areas of low/high utilization of public transport after adjusting for explanatory factors such as trip, individual and household attributes. The goal is to develop flexible statistical models for a binary response with covariate, spatial and cluster effects. One approach for modeling spatial effects are Markov Random Fields (MRF). A modification of a class of MRF models with proper joint distributions introduced by Pettitt et al. (2002) is developed. This modification has the desirable property to contain the intrinsic MRF in the limit and still allows for efficient spatial parameter updates in Markov Chain Monte Carlo (MCMC) algorithms. In addition to spatial effects, cluster effects are taken into consideration. Group and individual approaches for modeling these effects are suggested. The first one models heterogeneity between clusters, while the second one models heterogeneity within clusters. A naive approach to include individual cluster effects results in an unidentifiable model. It is shown how an appropriate reparametrization gives identifiable parameters. This provides a new approach for modeling heterogeneity within clusters. For hierarchical spatial binary regression models with individual cluster effects two MCMC algorithms for parameter estimation are developed. The first one is based on a direct evaluation of the likelihood. The second one is based on the representation of binary responses with Gaussian latent variables through a threshold mechanism, which is particularly useful for probit models. Simulation results show a satisfactory behavior of the MCMC algorithms developed. Finally the proposed model classes are applied to the mobility study and results are interpreted.

Key words: binary regression, spatial effects, group and individual cluster effects, MCMC

¹The first author was supported by Sonderforschungsbereich 386 *Statistische Analyse Diskreter Strukturen*, and the second author by the Graduiertenkolleg *Angewandte Algorithmische Mathematik* at the Center for Mathematical Sciences, Munich University of Technology, Germany

1 Introduction

This work has been motivated by a German mobility study investigating the usage of public transport options. The variable of interest is a binary indicator, whether public transport has been utilized or not. One central question of the investigators is to identify areas of low/high utilization of public transport after adjusting for explanatory factors such as trip, individual and household attributes. Therefore the goal is to develop flexible statistical models for a binary response with covariate, spatial and cluster effects. There are a large number of statistical models discussed in the literature which incorporate covariates together with spatial information. We provide now a short overview. In the context of general additive models, the simplest possibility to account for spatial information would be to use an additional nominal covariate indicating the region if there are multiple responses per region. But such an approach does not give a model for spatial dependence. This property is especially desired if the data volume is not large with respect to the number of covariates. In this case the assumption of a spatial structure (such as spatial smoothness) is especially helpful to be used as additional prior information.

There are two general approaches to incorporate spatial effects in a model. The first one is appropriate for data collected at specified point locations, while the other one uses data regions. The first approach is known as *generalized linear kriging* (see for example Diggle et al. 1998). It is based on generalized linear mixed models (Breslow and Clayton 1993), where spatial random effects are modeled as realizations of a stationary Gaussian process with zero mean and a parameterized covariance structure. For binary data this approach models the success probability p_i as follows:

$$p_i = \mathbf{E}(Y_i | \mathbf{x}_i, b_i) = h(\eta_i) \quad \text{and} \quad \eta_i = \mathbf{x}_i' \boldsymbol{\alpha} + b_i, \quad i = 1, \dots, n, \quad (1.1)$$

where \mathbf{x}_i is the design vector of the random variable Y_i and $b_i, i = 1, \dots, n$, are realizations of a zero mean stationary Gaussian process \mathbf{b} at the locations of the Y_i 's. The parameterization of the covariance structure by a covariance parameter $\boldsymbol{\delta}$ is usually based on distances between the observed locations. Even in the case of normal responses $Y_i, i = 1, \dots, n$, maximizing the likelihood over $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ becomes analytically intractable as soon as independence of the spatial effects $b_i, i = 1, \dots, n$, cannot be assumed. One general approach therefore is to maximize the reduced log-likelihood $l(\mathbf{Y}; \hat{\boldsymbol{\alpha}}(\boldsymbol{\delta}), \boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$, where $\hat{\boldsymbol{\alpha}}(\boldsymbol{\delta})$ is the maximum likelihood estimate of $\boldsymbol{\alpha}$ for fixed $\boldsymbol{\delta}$, and profile over $\boldsymbol{\delta}$. But such estimation is computationally expensive for large data sets. For arbitrary responses parameter estimation is carried out by Markov Chain Monte Carlo (MCMC) methods such as Gibbs sampling (see Diggle et al. 1998). For large data sets the updating of the covariance parameter $\boldsymbol{\delta}$ is difficult, since it requires to compute the determinant and inverse of a large dimensional variance-covariance matrix at each iteration. Heagerty and Lele (1998) remark (p.1104) that this step is computationally prohibitive already for sample sizes larger than 500. To overcome this problem they assume local independence between spatial effects which have a distance longer than some fixed value R . Heagerty and Lele (1998) use this idea for an iterative approach to determine the local conditional posterior mode of the spatial effect for the prediction at a new location. In contrast to Diggle et al. (1998), Heagerty and Lele (1998) estimate spatial effects $b_i, i = 1, \dots, n$ using a composite likelihood approach. Gelfand et al. (2000), which analyze a binary kriging model for the probit link function $h(\cdot)$ in (1.1), propose to apply MCMC with a suitably selected *importance sampling density*. Their method replaces a $n \times n$ matrix inversion with sampling from an n -dimensional normal, which for large values of n can be carried out much faster using a Cholesky decomposition. They also do not need to compute the determinant of the variance-covariance matrix. It allows to determine the posterior distribution of the regression parameter $\boldsymbol{\alpha}$ and the covariance parameter $\boldsymbol{\delta}$, but the posterior distribution for the spatial effects $b_i, i = 1, \dots, n$ cannot be calculated this way.

The other approach to incorporate a spatial model is appropriate when spatial effects are associated with data regions. These do not need to be on a regular lattice. The model equation is similar as in (1.1), but now data are assumed to be aggregated over regions and spatial effects are individual for each region instead for each observation, as before. Therefore the linear predictors are modeled as

$$\eta_i = \mathbf{x}_i' \boldsymbol{\alpha} + b_{j(i)}, \quad i = 1, \dots, n, \quad j = 1, \dots, J,$$

where J denotes the number of regions and $j(i)$ indicates the region associated with the i^{th} observation. The spatial effects b_j , $j = 1, \dots, J$, are modeled as a realization from some Gaussian *Markov random field* (MRF) (Besag and Green 1993). Gaussian MRF's are zero mean Gaussian process. The name *Gaussian conditional autoregression* (Gaussian CAR) is also used, since such a distribution is typically given through its full conditionals. This last fact allows fast individual updating of $J \ll n$ spatial effects in a Gibbs sampler scheme. More precisely, the distribution of the spatial effect b_j given all the other spatial effects depends only on the spatial effects of the neighbors of the j^{th} region. Therefore this approach requires some spatial neighborhood structure. This modeling is appropriate for our mobility application, since data are aggregated over postal codes of Munich, Germany. It is natural to consider two postal codes as neighbors if they have a joint border. In contrast to stationary Gaussian processes used in kriging, in Gaussian CAR models the explicit form of its precision matrix (inverse covariance matrix) is available. Therefore we do not need to compute the inverse of the variance-covariance matrix when updating the covariance parameter δ . Moreover this precision matrix is usually sparse, which allows to compute its determinant much faster than in the kriging approach. Further, Pettitt et al. (2002) use this fact and propose a specific dependence structure which provides even an analytical computation of its determinant. The next difference to stationary Gaussian processes, used in kriging, is that some Gaussian CAR models possess an improper joint density. The simplest example is the intrinsic CAR model (Besag and Green 1993), whose precision matrix is only semi positive. Fahrmeir and Lang (2000) use improper intrinsic CAR models as a prior for a Bayesian semi parametric regression model for multi categorical time-space data, while Knorr-Held and Rue (2000) applied intrinsic CAR priors for Poisson models used in disease mapping. For our application we study more advanced proper Gaussian CAR models with a parameterized correlation matrix. In particular, we develop a modification of the Pettitt's CAR model, which includes in the limit a specific intrinsic CAR model. The modification we propose still has all nice properties of the Pettitt et al. (2002) CAR models: proper joint distributions, a similar interpretation of parameters, the same conditional correlations and more important allows for fast computation of the determinant of the precision matrix, providing fast Gibbs sampling. An alternative proper Gaussian CAR model was also discussed in Sun et al. (2000). It also includes the intrinsic CAR model and allows for fast computation of the determinant of the precision matrix. It has been used to develop hierarchical spatio-temporal Poisson models for disease mapping data, but not for binary spatial responses. Gaussian CAR models will be considered in more detail in Section 2.

A principally different, well-known approach, also developed for spatial regression binary data over the region lattice is the *auto logistic regression model*. Huffer and Wu (1998), which use this method for the analysis of the distribution of plant species in Florida, U.S.A., propose to extend the auto logistic modeling of the success probability for each species by incorporating a fixed effect term $\mathbf{x}'_i \boldsymbol{\alpha}$:

$$\log \left(\frac{p_i}{1 - p_i} \right) = \mathbf{x}'_i \boldsymbol{\alpha} + \gamma y_i^*, \quad y_i^* := \sum_{j: i \sim j} y_j,$$

where " $i \sim j$ " indicates that sites i and j are neighbors. They work with a regular rectangular lattice and one-observation-per-site data. But in spite of this simplicity Huffer and Wu (1998) note that exact MLE is not tractable, except when the number of sites is quite small, while two other estimation methods, namely the coding method (Besag (1974)) and the maximum pseudo-likelihood method (Besag (1975)), seem to be not sufficiently efficient. For their application Huffer and Wu (1998) investigate a MCMC MLE approach, which produces the likelihood function via Monte Carlo simulations. Note they do not give any idea, how to take into consideration possible interactions between different species. For the Gaussian CAR approach Pettitt et al. (2002) solve this problem by modeling the correlation between several Gaussian CAR models for each species applied to tree biodiversity data. Also Carlin and Banerjee (2002) develop this approach for multiple cancer survival data with 3 types of cancer. Further a multivariate extension of the proper Gaussian CAR model developed in Sun et al. (2000) is considered by Gelfand and Vounatsou (2003) for multivariate continuous and multinomial response data.

We close our short overview on spatial modeling for binary data by mentioning a *non parametric binary regression* approach, which was proposed by Kelsall and Diggle (1998) for the analysis of spatial variation in risk of disease. The idea is to model logits through spatially dependent intensities $\lambda_1(x)$

(cases) and $\lambda_2(x)$ (controls), where $x \in A \subset \mathbb{R}^2$ is the response point location:

$$p(x) := P(Y_i|X_i = x) = \frac{q_1 \lambda_1(x)}{q_1 \lambda_1(x) + q_2 \lambda_2(x)} \Rightarrow \text{logit}(p(x)) = \log\left(\frac{\lambda_1(x)}{\lambda_2(x)}\right) + c \quad (1.2)$$

Kelsall and Diggle (1998) consider non parametric kernel estimators $\hat{p}_h(x)$ for $p(x)$, where h denotes the corresponding bandwidth. Cross-validation is used to optimize h . While in the Gaussian CAR approach we test the significance of the spatial effects b_j , $j = 1, \dots, J$, Kelsall and Diggle (1998) construct tolerance contours, which indicate for each x whether $\hat{p}_h(x)$ is consistent with the proportional hazard assumption, which is given by the null hypothesis $H_0 : \frac{\lambda_1(x)}{\lambda_2(x)} = \text{const}$. These are determined by generating m times new data which are consistent with H_0 but otherwise similar in distribution to the original data. Finally, they construct new estimates $\hat{p}_h^{sim}(x)$ for each of the m generating data sets. The authors note that, since covariates are not included in (1.2), the collection of a stratified sample of controls can be very difficult, particularly when the number of covariates is large. Therefore they extend their model by including fixed effects $\mathbf{u}^t \boldsymbol{\alpha}$:

$$\text{logit}(p(x, u)) = \mathbf{u}^t \boldsymbol{\alpha} + g(x) \quad (1.3)$$

This extension allows to collect a simple random sample of controls and to take into account covariates by modeling their effects within (1.3). The authors however note that kernel regression estimation based on Model (1.3) is substantially more computer intensive, so that the simpler kernel regression method, based on (1.2) will sometimes be preferable.

In addition to spatial effects we extend our modeling of the linear predictor η_i by cluster random effects. It allows us to obtain more flexible models taking into account possible overdispersion caused by unobserved heterogeneity. Examples of clusters are age groups or the household types. We consider two approaches, namely group and individual cluster effects. The first one, which models heterogeneity between clusters, implies the usual idea of having the same random effect within a cluster. These random effects are assumed to be some realization from the multivariate normal $N_K(0, \sigma_c^2 I_K)$ with usually unknown cluster variance σ_c^2 . Here K denotes the number of clusters and I_K stands for the K -dimensional identity matrix. A different choice of the variance-covariance matrix is possible.

The second approach allows for heterogeneity within a cluster, i.e. we model cluster effects within a cluster as independent normally distributed random variables with zero mean and a cluster specific variance. Therefore we have to estimate K cluster specific variances instead of K cluster effects as before. In this paper we will show how an unidentifiability problem occurring in the second case can be overcome. For this hierarchical spatial binary regression model with individual cluster effects we develop two estimating MCMC algorithms. The first one is useful if the likelihood of the data, given covariates and unknown parameters, can be easily computed as for binary logistic models. Markov Chains are then generated using Metropolis-Hastings steps. The second approach, which is particularly useful for probit model is based on latent variables, where the observed binary responses are generated through a threshold mechanism. For latent Gaussian variables this leads to binary probit models (see for example Albert and Chib 1993). For MCMC inference, Gaussian latent variables are considered as unknown additional “parameters” and are generated with the other parameters in a Gibbs sampling scheme. We note that block updating for the regression parameters $\boldsymbol{\alpha}$ and the spatial parameters \mathbf{b} is available in this estimating algorithm. Therefore we achieve considerably better mixing than in the direct algorithm, where parameters are updated individually. This allows us to reduce the number of iterations in the corresponding Markov chains. Further, this method reduces the number of parameters, which require a numerically more complicated Metropolis-Hastings step to only one.

The remainder of this paper is organized as follows. In Section 2 we discuss spatial effects modeling using Gaussian CAR processes. We propose some modification of the Gaussian CAR models developed by Pettitt et al. (2002), which allows to achieve an intrinsic CAR model in the limit. Using this modification we develop in Section 3 a hierarchical spatial binary regression model with group cluster effects, while in Section 4 we present individual cluster effects modeling. In Section 5 we present the results of a comprehensive simulation study investigating the performance of the MCMC algorithms developed in Sections 3 and 4 in small samples. In Section 6 we apply our approach to data from a German mobility study. Finally Section 7 gives a discussion and presents an outlook for further research.

2 Modeling of Spatial Effects Using Gaussian CAR Models

The most popular kind of Markov random fields (MRF) are Gaussian MRF's (Besag and Green 1993), or Gaussian conditional autoregressive processes (Gaussian CAR) (see for detailed discussion Pettitt et al. 2002), where a random vector $\mathbf{b} \in R^J$ is defined through its full conditionals as follows:

$$b_j | \mathbf{b}_{-j} \sim N \left(\mu_j + \sum_{j' \neq j} c_{jj'} (b_{j'} - \mu_{j'}), \kappa_j \right), \quad j = 1, \dots, J.$$

Here $\mathbf{b}_{-j} = (b_1, \dots, b_{j-1}, b_{j+1}, \dots, b_J)^t$ and $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2 . Besag and Green (1993) show that the joint distribution of a zero-mean Gaussian CAR is given by

$$\mathbf{b} \sim N_J(0, (I_J - C)^{-1}M), \quad (2.1)$$

where $C = (c_{jj'})$ with $c_{jj} = 0$, $j = 1, \dots, J$, and $M = \text{diag}(\kappa_1, \dots, \kappa_J)$. Here $N_J(\mu, \Sigma)$ denotes a J -dimensional normal distribution with mean vector μ and covariance matrix Σ . The precision matrix is given by $Q = M^{-1}(I_J - C)$. Below we present examples of Gaussian CARs. Further we will always assume that the neighborhood structure in the Gaussian CAR has no isolated regions or groups of regions.

Example 1: The *intrinsic Gaussian CAR* (Besag and Green 1993) is defined by:

$$b_j | \mathbf{b}_{-j} \sim N(\bar{b}_j, \frac{\tau^2}{N_j}), \quad j = 1, \dots, J, \quad \text{and} \quad \bar{b}_j = \frac{\sum_{j \sim j'} b_{j'}}{N_j}, \quad (2.2)$$

where $N_j = \#$ of neighbors of the region j , and " $j \sim j'$ " denotes contiguous regions. In particular, we have $j \not\sim j$. The precision matrix of \mathbf{b} is equal to $\frac{Q_0}{\tau^2}$, where

$$Q_0 = (q_{jj'}) = \begin{cases} N_j, & \text{if } j = j' \\ -1, & \text{if } j \sim j' \\ 0, & \text{otherwise} \end{cases}. \quad (2.3)$$

Q_0 is positive semi-definite with $\text{rank}(Q_0) = J - 1$, therefore \mathbf{b} has an improper density, but can be characterized (see Prokopenko 2004).

Example 2: Pettitt et al. (2002) use a particular Gaussian CAR, where

$$b_j | b_{j'}, j \neq j' \sim N \left(\frac{\phi}{1 + |\phi|N_j} \sum_{j \sim j'} b_{j'}, \frac{\tau^2}{1 + |\phi|N_j} \right). \quad (2.4)$$

Here the parameter ϕ measures the strength of the spatial dependency. There is no spatial dependency, if $\phi = 0$. Since maximum likelihood estimation is intractable for this model MCMC methods have been used to estimate ϕ and τ^2 . It was shown (Pettitt et al. 2002) that a fast and simple update of ϕ for a Gibbs Step given the vector \mathbf{b} and τ^2 is available. Note that in contrast to the intrinsic CAR, the joint distribution of \mathbf{b} based on conditionals specified in (2.4) is a proper distribution, which leads to a proper posterior when used as a prior distribution. This will circumvent any problems in the Gibbs sampler arising from using an improper prior.

Example 3: In this paper we introduce now a modified Pettitt's CAR model, where the full conditionals for \mathbf{b} are given as follows:

$$b_j | b_{j'}, j \neq j' \sim N \left(\frac{\phi}{1 + |\phi|N_j} \sum_{j \sim j'} b_{j'}, \frac{(1 + |\phi|)\tau^2}{1 + |\phi|N_j} \right). \quad (2.5)$$

This (also proper) distribution differs from Pettitt's CAR (2.4) by the additional term $1 + |\phi|$ in the numerator of the conditional variance. This modification allows us to have the intrinsic CAR (2.2) in the limit, when $\phi \rightarrow \infty$ (for details see Prokopenko 2004). Note that the conditional variance of $b_j | b_{j'}$ decreases to τ^2/N_j as $|\phi|$ increases to infinity, while in the original model (2.4) this quantity decreases to

zero, which is a very restrictive assumption. Further, this model has the same behavior as Pettitt's CAR (2.4) when ϕ goes to zero (no spatial dependency), and also all partial correlations between b_j and b_i given all the other sites are the same, as in (2.4). Finally, in the modified Pettitt's model we can also achieve a simple update for ϕ . To indicate the dependency on ϕ we write now $Q^{m.P}(\phi)$ for the precision matrix of the modified Pettitt's model (2.5). Each update of ϕ requires the computation of the determinant of $Q^{m.P}(\phi)$. With the reparametrization $\psi = \frac{\phi}{1+|\phi|}$ we can apply the same procedure as in Pettitt et al. (2002). More precisely, if we define the diagonal matrix

$$D = \text{diag}(N_1 - 1, \dots, N_J - 1) \quad \text{and} \quad \Gamma = (\gamma_{jj'})_{j,j'=1,\dots,J} = \begin{cases} 1, & \text{if } j \sim j' \\ 0, & \text{if } j \not\sim j', j = j' \end{cases},$$

then $Q^{m.P}(\phi)$ can be written in the form $Q^{m.P}(\psi) = I_J + |\psi|D - \psi\Gamma$. If $(\lambda_1, \dots, \lambda_J)$ are the eigenvalues of $\Gamma - D$ and (ν_1, \dots, ν_J) are the eigenvalues of $\Gamma + D$, then the determinant of $Q^{m.P}(\psi)$ is equal to

$$|Q^{m.P}(\psi)| = \begin{cases} \prod_j (1 - \psi\lambda_j), & \text{if } \psi > 0 \\ 1, & \text{if } \psi = 0 \\ \prod_j (1 - \psi\nu_j), & \text{if } \psi < 0 \end{cases}. \quad (2.6)$$

and can be computed quickly for any value of ψ . Finally we like to note that the conditional variance of $b_j|b_{j'}, j \neq j'$ is independent of the spatial dependence parameter for the proper Gaussian CAR model considered by Sun et al. (2000) in contrast to the modified Pettitt's CAR model (2.5). It is more reasonable to assume that this conditional variance increases as dependence among the spatial effects decreases. If $\phi = 0$, then the conditional variance in Sun et al. (2000) still depends on N_j , while this is not the case for the modified Pettitt's CAR model. Therefore we prefer the modified proper Pettitt's CAR model over the proper CAR model studied by Sun et al. (2000) for modeling spatial effects.

3 Spatial Binary Regression with Group Cluster Effects

3.1 Model Formulation

For the data from the mobility study we use a binary response vector $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ with

$$Y_i = \begin{cases} 1 & \text{if trip } i \text{ used individual transport} \\ 0 & \text{if trip } i \text{ used public transport} \end{cases}, \quad i = 1, \dots, n, \quad (3.1)$$

where Y_i 's follow a Bernoulli distribution with the success probabilities p_i and assume that Y_i given p_i are independent for $i = 1, \dots, n$. In this model we specify p_i through their logits as follows:

$$\theta_i := \log\left(\frac{p_i}{1-p_i}\right) = \underbrace{\mathbf{x}_i^t \boldsymbol{\alpha}}_{\text{fixed effect}} + \underbrace{b_{j(i)}}_{\text{random spatial effect}} + \underbrace{c_{m(i)}}_{\text{random group cluster effect}}. \quad (3.2)$$

Here the design vector \mathbf{x}_i multiplied with the regression parameter vector $\boldsymbol{\alpha} \in \mathbb{R}^p$ represents the fixed effects. With the vector $\mathbf{b} = (b_1, \dots, b_J)$ we attempt to take into consideration possible random spatial effects. As sites we take $J = 74$ postal code areas of the city of Munich. Therefore, the index $j(i)$ denotes the residence postal code of the person who takes trip i . In order to be able to take into account possible spatial smoothness we assume, that b_j 's arise from the modified Pettitt's CAR (2.5).

To model heterogeneity between the clusters we allow also for random cluster effects, which are represented by the vector $\mathbf{c} = (c_1, \dots, c_M)$. We assume that each of the M clusters (say age groups or household types) induces a group specific random effect, which we denote by $c_m, m = 1, \dots, M$, respectively. Therefore we speak of group cluster effects. The index $m(i)$ denotes the cluster of trip i . Finally, we assume that $c_m \sim N(0, \sigma_c^2)$ *i.i.d.* for $m = 1, \dots, M$. This completes the description of Model (3.2). Note that the likelihood of the response vector \mathbf{Y} is proportional to

$$[\mathbf{Y} | \boldsymbol{\alpha}, \mathbf{b}, \mathbf{c}] \propto \prod_{i=1}^n \frac{\exp(Y_i(\mathbf{x}_i^t \boldsymbol{\alpha} + b_{j(i)} + c_{m(i)}))}{1 + \exp(\mathbf{x}_i^t \boldsymbol{\alpha} + b_{j(i)} + c_{m(i)})}.$$

Finally, we remark that Model (3.2) is similar to a family of semi parametric models for multi categorical time-space data (with time- instead cluster effects) as discussed in Fahrmeir and Lang (2000).

3.2 Bayesian Inference Using MCMC Methods

Markov Chain Monte Carlo (MCMC) methods allow us to draw an arbitrary large number of joint samples from the posterior distribution $[\boldsymbol{\alpha}, \mathbf{b}, \mathbf{c}, \phi, \tau^2, \sigma_c^2 | \mathbf{Y}]$ approximately. With these samples we can easily make inference for these parameters using for example estimated posterior means or density estimates of the marginal posterior. Readers unfamiliar with MCMC methods can for example consult Chib (2001) for an introduction and Gilks et al. (1996) for applications of MCMC methods. We denote further the density of a random variable X by $[X]$ and the conditional density of X given Y by $[X|Y]$. We assume independent prior distributions for the fixed effect $\boldsymbol{\alpha}$, the spatial parameters \mathbf{b} given their dependence parameter ϕ and variance scalar τ^2 , the cluster parameters \mathbf{c} given their random variance σ_c^2 and the hyperparameters ϕ, τ^2, σ_c^2 respectively, i.e.

$$[\boldsymbol{\alpha}, \mathbf{b}, \mathbf{c}, \phi, \tau^2, \sigma_c^2] = [\boldsymbol{\alpha}] \times [\mathbf{b}|\phi, \tau^2] \times [\phi] \times [\tau^2] \times [\mathbf{c}|\sigma_c^2] \times [\sigma_c^2].$$

A more detailed description of the following updates can be found in Prokopenko (2004).

3.2.1 Regression Parameter Update

The full conditional for an individual update of α_l , $l = 1, \dots, p$, has the following form:

$$[\alpha_l | \mathbf{Y}, \boldsymbol{\alpha}_{-l}, \mathbf{b}, \mathbf{c}, \phi, \tau^2, \sigma_c^2] \propto \prod_{i=1}^n \frac{\exp(Y_i(x_{il}\alpha_l))}{1 + \exp(\mathbf{x}_i^t \boldsymbol{\alpha} + b_{j(i)} + c_{m(i)})} [\alpha_l], \quad (3.3)$$

where $\boldsymbol{\alpha}_{-l} = (\alpha_1, \dots, \alpha_{l-1}, \alpha_{l+1}, \dots, \alpha_p)^t$. Individual updates are used, since a good joint proposal density for *Metropolis-Hastings (MH) step* is difficult to find. As prior for α_l , $l = 1, \dots, p$ a normal distribution with zero-mean and large standard deviation was taken. As proposal density in the r^{th} iteration for α_l^r a normal density with the mean equal to the previous value α_l^{r-1} and a fixed value for the standard error was chosen. The value for this standard error was found using pilot runs. In particular, we used pilot runs to determine standard error values which give an acceptance rate between 30-60% (as proposed in Bennett et al. (1996) or Besag et al. (1995)). The pilot runs also served as “burn in” phase.

3.2.2 Spatial Parameter Update

For the prior density of \mathbf{b} we use the modified Pettitt’s conditional autoregression (2.5). Since the computational effort for the joint update of \mathbf{b} are rapidly increasing with the dimension of \mathbf{b} , we use individual updates here as well. The full conditional densities are proportional to:

$$[b_j | \mathbf{Y}, \mathbf{b}_{-j}, \boldsymbol{\alpha}, \mathbf{c}, \phi, \tau^2, \sigma_c^2] \propto [\mathbf{Y} | \boldsymbol{\alpha}, \mathbf{b}, \mathbf{c}] \times [b_j | \mathbf{b}_{-j}, \boldsymbol{\alpha}, \phi, \tau^2] \\ \propto \prod_{i:j(i)=j} \frac{\exp(Y_i b_j)}{1 + \exp(\mathbf{x}_i^t \boldsymbol{\alpha} + b_j + c_{m(i)})} \exp \left\{ -\frac{1+|\phi| N_j}{2(1+|\phi|)\tau^2} \left(b_j - \frac{\phi}{1+|\phi| N_j} \sum_{j \sim j'} b_{j'} \right)^2 \right\}, \quad (3.4)$$

where $\mathbf{b}_{-j} = (b_1, \dots, b_{j-1}, b_{j+1}, \dots, b_J)^t$. For the required MH step, we used as proposal density for b_j a similar proposal density as for the regression parameters and we also used pilot runs in order to find a good proposal standard error for each spatial parameter $b_j, j = 1, \dots, J$.

3.2.3 Spatial Dependence Parameter Update

The full conditional density for ϕ is given by

$$[\phi | \mathbf{Y}, \boldsymbol{\alpha}, \mathbf{b}, \mathbf{c}, \tau^2, \sigma_c^2] = [\phi | \mathbf{b}, \tau^2] \propto [\mathbf{b} | \phi, \tau^2] \times [\phi]. \quad (3.5)$$

Since $\mathbf{b} | \phi, \tau^2 \sim N_J(0, \tau^2 Q^{m.P}(\phi)^{-1})$, the determinant of the matrix $Q^{m.P}(\phi)$ must be calculated for each ϕ -iteration. Since once the eigenvalues of $\Gamma - D$ and $\Gamma + D$ are known, we can write the determinant of $Q^{m.P}(\phi)$ analytically as a function of $\psi = \frac{\phi}{1+|\phi|}$ as in (2.6). Since $\psi \in (-1, 1)$, it is reasonable to take a uniform distributed prior for ψ on $(-1, 1)$. Such a prior choice corresponds to a heavy-tailed prior for ϕ , namely a Pareto distribution with the density $\sim \frac{1}{(1+|\phi|)^2}$. This density has no finite moments, but it is unimodal and symmetric with mode at 0. To generalize the link between the priors for ψ and ϕ we note

that the prior for ψ proportional to $\frac{1}{(1-|\psi|)^{1+a}}$, $\psi \in (-1, 1)$ corresponds to a prior for ϕ proportional to $\frac{1}{(1+|\phi|)^{1+a}}$, $\phi \in (-\infty, +\infty)$. It is a proper prior for $a > 0$. The last term in the conditional $[\phi|\mathbf{b}, \tau^2]$ which depends on ψ is proportional to $\exp(-\frac{1}{2\tau^2} \mathbf{b}' Q^{m,P}(\psi) \mathbf{b})$ and can also be calculated fast in each iteration. As proposal distribution for the r^{th} iteration ψ^r we also take a normal density with mean equal to the previous iteration ψ^{r-1} , but now in contrast to the parameters α and \mathbf{b} truncated to the interval $(-1, 1)$. Again pilot runs are used to initialize proposal standard error.

3.2.4 Spatial Variance Parameter Update

For this update, an inverse gamma prior for τ^2 is used with density given by

$$[\tau^2] = \frac{1}{b_\tau^{(a_\tau)} \Gamma(a_\tau) (\tau^2)^{a_\tau+1}} \exp\left(-\frac{1}{b_\tau \tau^2}\right), \quad (3.6)$$

where $a_\tau > 0$ and $b_\tau > 0$ are known hyperparameters. We denote this prior by $\tau^2 \sim IG(a_\tau, b_\tau)$. Since the full conditional for τ^2 depends solely on \mathbf{b} and ϕ we can write

$$[\tau^2|\mathbf{Y}, \alpha, \mathbf{b}, \mathbf{c}, \phi, \sigma_c^2] = [\tau^2|\mathbf{b}, \phi] \propto [\mathbf{b}|\phi, \tau^2] \times [\tau^2].$$

We have that the conditional distribution of $[\tau^2|\mathbf{Y}, \alpha, \mathbf{b}, \mathbf{c}, \phi, \sigma_c^2]$, is again $IG(a_\tau^*, b_\tau^*)$ with

$$a_\tau^* = a_\tau + \frac{J}{2} \text{ and } b_\tau^* = \left\{ \frac{1}{b_\tau} + \frac{\mathbf{b}' Q(\phi) \mathbf{b}}{2} \right\}^{-1}.$$

When a flat improper prior for τ^2 is used (as we have chosen), the posterior $[\tau^2|\mathbf{b}, \phi]$ is $IG(a_\tau^*, b_\tau^*)$ with

$$a_\tau^* = \frac{J}{2} - 1 \text{ and } b_\tau^* = \left\{ \frac{1}{2} \mathbf{b}' Q(\phi) \mathbf{b} \right\}^{-1}.$$

3.2.5 Cluster Parameter Update

The cluster effect $\mathbf{c} = (c_1, \dots, c_m)$ is taken as a random effect with prior $c_m \sim N(0, \sigma_c^2)$ for each cluster $m = 1, \dots, M$, *i.i.d.* Again an MH step is needed. The individual full conditionals can be written as follows

$$\begin{aligned} [c_m|\mathbf{Y}, \mathbf{c}_{-m}, \alpha, \mathbf{b}, \phi, \tau^2, \sigma_c^2] &\propto [\mathbf{Y}|\alpha, \mathbf{b}, \mathbf{c}] \times [c_m|\sigma_c^2] \\ &\propto \prod_{i:m(i)=m} \frac{\exp(Y_i c_m)}{1 + \exp(\mathbf{x}_i^t \alpha + b_{j(i)} + c_m)} \exp\left\{-\frac{1}{2\sigma_c^2} c_m^2\right\}, \end{aligned} \quad (3.7)$$

where $\mathbf{c}_{-m} = (c_1, \dots, c_{m-1}, c_{m+1}, \dots, c_M)^t$. As proposal density for the r^{th} iteration c_m^r we chose a normal density with mean equal to c_m^{r-1} from the previous iteration.

3.2.6 Cluster Variance Parameter Update

The full conditional density of the cluster variance parameter σ_c^2 has a similar form as the spatial variance parameter τ^2 , namely:

$$[\sigma_c^2|\mathbf{Y}, \alpha, \mathbf{b}, \mathbf{c}, \phi, \tau^2] = [\sigma_c^2|\mathbf{c}] \propto [\mathbf{c}|\sigma_c^2] \times [\sigma_c^2] \propto \frac{1}{(\sigma_c^2)^{M/2}} \exp\left\{-\frac{1}{2\sigma_c^2} \mathbf{c}' \mathbf{c}\right\} [\sigma_c^2]. \quad (3.8)$$

A direct Gibbs step is available by choosing as prior density an inverse gamma prior (see (3.6)) or an improper prior for σ_c^2 . In particular, if $\sigma_c^2 \sim IG(a_c, b_c)$, then $\sigma_c^2|\mathbf{Y}, \alpha, \mathbf{b}, \mathbf{c}, \phi, \tau^2 \sim IG(a_c^*, b_c^*)$ with

$$a_c^* = a_c + \frac{M}{2} \text{ and } b_c^* = \left\{ \frac{1}{b_c} + \frac{\mathbf{c}' \mathbf{c}}{2} \right\}^{-1}.$$

For an improper prior it follows that the full conditional density for σ_c^2 is a $IG(a_c^*, b_c^*)$ density with

$$a_c^* = \frac{M}{2} - 1 \text{ and } b_c^* = \left\{ \frac{1}{2} \mathbf{c}' \mathbf{c} \right\}^{-1}.$$

This density has a finite expectation for $M \geq 5$ and a finite variance for $M \geq 7$.

4 Spatial Binary Regression with Individual Cluster Effects

4.1 Model Formulation

Now we introduce for our binary transport response (3.1) a more advanced model where individual cluster effects are modeled by a normal distribution with fixed variance inside each cluster given by:

$$Y_i|p_i \sim \text{Bernoulli}(p_i) \text{ conditionally independent with} \\ \theta_i := \log\left(\frac{p_i}{1-p_i}\right) = \underbrace{\mathbf{x}_i^t \boldsymbol{\alpha}}_{\text{fixed effect}} + \underbrace{b_{j(i)}}_{\text{random spatial effect}} + \underbrace{c_{m(i),k(i)}}_{\text{random individual cluster effect}}, \quad (4.1)$$

where for fixed $m = 1, \dots, M$, $c_{m,k} \sim N(0, \sigma_m^2)$, $k = 1, \dots, K_m$, *i.i.d.* As in Model (3.2), M denotes the number of clusters and $m(i)$ denotes the cluster of trip i . K_m stands for the number of trips, which belong to cluster m (i.e. $K_1 + \dots + K_M = n$) and the index $k(i)$ gives the number of trip i in its cluster. The specification of the fixed effects $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^t$ and the spatial effects $\mathbf{b} = (b_1, \dots, b_J)^t$ remain as before. In contrast to (3.2), the cluster effects are now not the same for each trip in cluster m , namely c_m , but random realizations $c_{m,k}$, $k = 1, \dots, K_m$ from the same cluster distribution $N(0, \sigma_m^2)$ denoted by $\mathbf{c} = (c_{1,1}, \dots, c_{1,K_1}, \dots, c_{M,1}, \dots, c_{M,K_M})^t$. This allows for modeling heterogeneity within each cluster. In Model (4.1) we have to estimate in addition to the parameters $\boldsymbol{\alpha}, \mathbf{b}$ the cluster effect variances $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_M^2)^t$ instead of the cluster effects $\mathbf{c} = (c_1, \dots, c_M)^t$ and their variance σ_c^2 for Model (3.2).

One problem with Model (4.1) is that even without an intercept term α_0 the model is unidentifiable. To understand the nature of the unidentifiability we first substitute in (4.1) the logit link function with the probit link function, i.e. we assume for $i = 1, \dots, n$:

$$Y_i|p_i \sim \text{Bernoulli}(p_i) \text{ conditionally independent with} \\ p_i = \mathbf{P}\{Y_i = 1|\mathbf{x}_i, \boldsymbol{\alpha}, b_{j(i)}, c_{m(i),k(i)}\} = \Phi(\mathbf{x}_i^t \boldsymbol{\alpha} + b_{j(i)} + c_{m(i),k(i)}), \quad (4.2)$$

where $\Phi(\cdot)$ is the standard normal distribution function. This allows for the following latent variable representation:

$$Y_i = 1|\mathbf{x}_i, \boldsymbol{\alpha}, b_{j(i)}, \sigma_{m(i)}^2 \Leftrightarrow Z_i \leq 0, \quad \text{where} \\ Z_i = -\eta_i + \epsilon_i^*, \quad \epsilon_i^* \sim N(0, 1 + \sigma_{m(i)}^2) \text{ independent and } \eta_i = \mathbf{x}_i^t \boldsymbol{\alpha} + b_{j(i)}. \quad (4.3)$$

This is a similar representation as discussed in Albert and Chib (1993) for binary probit models. Therefore we have for $i = 1, \dots, n$

$$\mathbf{P}\{Y_i = 1|\mathbf{x}_i, \boldsymbol{\alpha}, b_{j(i)}, \sigma_{m(i)}^2\} = \mathbf{P}\{Z_i \leq 0|\mathbf{x}_i, \boldsymbol{\alpha}, b_{j(i)}, \sigma_{m(i)}^2\} = \Phi\left(\frac{\mathbf{x}_i^t \boldsymbol{\alpha} + b_{j(i)}}{\sqrt{1 + \sigma_{m(i)}^2}}\right). \quad (4.4)$$

Equation (4.4) shows that the parameters $\boldsymbol{\alpha}, \mathbf{b}$ and $\boldsymbol{\sigma}^2$ are not jointly identifiable in Model (4.2), since it is invariant with respect to the parameter vectors $\left\{k \times (\boldsymbol{\alpha}^t, \mathbf{b}^t, \sqrt{1 + \sigma_1^2}, \dots, \sqrt{1 + \sigma_M^2})^t, k \in \mathbb{R}\right\}$. If we define now

$$\boldsymbol{\alpha}' := \frac{\boldsymbol{\alpha}}{\sqrt{1 + \sigma_1^2}}, \quad \mathbf{b}' := \frac{\mathbf{b}}{\sqrt{1 + \sigma_1^2}}, \quad \sigma_m'^2 := \frac{1 + \sigma_m^2}{1 + \sigma_1^2}, \quad m = 2, \dots, M, \quad \sigma_1'^2 = 1, \quad (4.5)$$

then the marginal distributions (4.4) of $Y_i|\mathbf{x}_i, \boldsymbol{\alpha}, b_{j(i)}, \sigma_{m(i)}^2$ from Model (4.2) will coincide with the marginal distributions from the following model:

$$Y_i|p_i \sim \text{Bernoulli}(p_i) \text{ conditionally independent with} \\ p_i = \mathbf{P}\{Y_i = 1|\mathbf{x}_i, \boldsymbol{\alpha}', b'_{j(i)}, \sigma_{m(i)}'^2\} = \begin{cases} \Phi\left(\mathbf{x}_i^t \boldsymbol{\alpha}' + b'_{j(i)}\right) & \text{if } m(i) = 1 \\ \Phi\left(\frac{\mathbf{x}_i^t \boldsymbol{\alpha}' + b'_{j(i)}}{\sigma_{m(i)}'}\right) & \text{if } m(i) = 2, \dots, M. \end{cases} \quad (4.6)$$

Using (4.3) it follows, that also the joint distribution of $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ in both Models (4.2) and (4.6) are equal. Therefore Model (4.6) is an equivalent reparametrization of Model (4.2). But this representation (4.6) has one parameter less and is therefore identifiable. Further, Model (4.6) can also be represented using normal latent variables:

$$\begin{aligned} Y_i = 1 | \mathbf{x}_i, \boldsymbol{\alpha}', b'_{j(i)}, \sigma_{m(i)}'^2 &\Leftrightarrow Z'_i \leq 0, \quad \text{where} \\ Z'_i = -\eta'_i + \epsilon'_i, \quad \epsilon'_i &\sim N(0, \sigma_{m(i)}'^2) \text{ ind.} \quad \text{and} \quad \eta'_i = \mathbf{x}_i^t \boldsymbol{\alpha}' + b'_{j(i)}. \end{aligned} \quad (4.7)$$

We will show that the latent variable representation allows us in the probit case to reduce significantly the number of variables, which need MH-updates. MH-updates are especially computational expensive for probit models because of the numerical complexity of the computations for $\Phi(\cdot)$ in the tails.

Note that large values of cluster parameter $\sigma_m'^2$ indicate the large heterogeneity within cluster m . Such interpretation of cluster parameters follows from the reparametrization (4.5).

The above discussion helps us also to understand the unidentifiability of Logit Model (4.1), since the behavior of both probit and logit link functions is quite similar and becomes significantly different only in the tails. So we use the same idea to construct an identifiable logit model, which is approximately equivalent in distribution to (4.1). In particular we assume for $i = 1, \dots, n$

$$\begin{aligned} Y_i | p_i &\sim \text{Bernoulli}(p_i) \text{ conditionally independent with} \\ \log\left(\frac{p_i}{1-p_i}\right) &= \begin{cases} \mathbf{x}_i^t \boldsymbol{\alpha}' + b'_{j(i)} & \text{if } m(i) = 1 \\ \frac{\mathbf{x}_i^t \boldsymbol{\alpha}' + b'_{j(i)}}{\sigma_{m(i)}'} & \text{if } m(i) = 2, \dots, M \end{cases}, \end{aligned} \quad (4.8)$$

where $p_i = \mathbf{P}\{Y_i = 1 | \mathbf{x}_i, \boldsymbol{\alpha}', b'_{j(i)}, \sigma_{m(i)}'^2\}$ and $\boldsymbol{\alpha}', \mathbf{b}', \boldsymbol{\sigma}'^2 := (\sigma_1'^2, \dots, \sigma_M'^2)^t$ are defined as in (4.5).

4.2 Bayesian Inference for Hierarchical Spatial Binary Regression Models with Individual Cluster Effects

4.2.1 Bayesian Inference for Logit Model (4.8)

We begin with the logit case (4.8). From (4.8) it follows that the likelihood of the response vector \mathbf{Y} is proportional to

$$[\mathbf{Y} | \boldsymbol{\alpha}', \mathbf{b}', \boldsymbol{\sigma}'] \propto \prod_{i=1}^n \frac{\exp(Y_i \frac{\mathbf{x}_i^t \boldsymbol{\alpha}' + b'_{j(i)}}{\sigma_{m(i)}'})}{1 + \exp(Y_i \frac{\mathbf{x}_i^t \boldsymbol{\alpha}' + b'_{j(i)}}{\sigma_{m(i)}'})},$$

where $\boldsymbol{\sigma}' := (1, \sigma_2', \dots, \sigma_M')^t$, $\sigma_m' := \sqrt{\sigma_m'^2}$, $m = 2, \dots, M$. We again assume independent prior distributions for the fixed effect $\boldsymbol{\alpha}'$, the spatial parameters \mathbf{b}' given their dependence parameter ϕ' and the variance scalar τ'^2 and the cluster parameters $\boldsymbol{\sigma}'$. Finally we assume independence between the hyperparameters ϕ' and τ'^2 . Therefore the joint prior distribution is given by $[\boldsymbol{\alpha}', \mathbf{b}', \boldsymbol{\sigma}', \phi', \tau'^2] = [\boldsymbol{\alpha}'] \times [\mathbf{b}' | \phi', \tau'^2] \times [\phi'] \times [\tau'^2] \times [\boldsymbol{\sigma}']$. The MCMC update procedure for the parameters $\boldsymbol{\alpha}'$ and \mathbf{b}' remains the same as in Subsection 3.2, with the full conditionals replaced by

$$[\alpha'_l | \mathbf{Y}, \boldsymbol{\alpha}'_{-l}, \mathbf{b}', \boldsymbol{\sigma}', \phi', \tau'^2] \propto \prod_{i=1}^n \frac{\exp(Y_i \frac{x_{il} \alpha'_l}{\sigma_{m(i)}'})}{1 + \exp(Y_i \frac{\mathbf{x}_i^t \boldsymbol{\alpha}' + b'_{j(i)}}{\sigma_{m(i)}'})} [\alpha'_l]$$

and

$$\begin{aligned} [b'_j | \mathbf{Y}, \mathbf{b}'_{-j}, \boldsymbol{\alpha}', \boldsymbol{\sigma}', \phi', \tau'^2] &\propto [\mathbf{Y} | \boldsymbol{\alpha}', \mathbf{b}', \boldsymbol{\sigma}'] [b'_j | \mathbf{b}'_{-j}, \phi', \tau'^2] \\ &\propto \prod_{i: j(i)=j} \frac{\exp(Y_i \frac{b'_j}{\sigma_{m(i)}'})}{1 + \exp(\frac{\mathbf{x}_i^t \boldsymbol{\alpha}' + b'_j}{\sigma_{m(i)}'})} \exp\left\{-\frac{1+|\phi'| N_j}{2(1+|\phi'|)\tau'^2} \left(b'_j - \frac{\phi'}{1+|\phi'| N_j} \sum_{j \sim j'} b'_{j'}\right)^2\right\}. \end{aligned}$$

Since the full conditionals of the spatial hyperparameters τ'^2 and ϕ' given the data and other parameters depend only on the spatial effects \mathbf{b}' , their MCMC updates have the same form as described in Subsection

3.2. We update the parameter σ' in a similar way as the other main parameters α' and \mathbf{b}' . The full conditional of σ'_m , $m = 2, \dots, M$, can be written as

$$[\sigma'_m | \mathbf{Y}, \alpha', \mathbf{b}', \sigma'_{-m}, \phi', \tau^{2'}] \propto [\mathbf{Y} | \alpha', \mathbf{b}', \sigma'] [\sigma'_m] \propto \prod_{i:m(i)=m}^n \frac{\exp(Y_i \frac{\mathbf{x}_i^t \alpha' + b'_{j(i)}}{\sigma'_m})}{1 + \exp(Y_i \frac{\mathbf{x}_i^t \alpha' + b'_{j(i)}}{\sigma'_m})} [\sigma'_m].$$

It is reasonable to take a prior for σ'_m , $m = 2, \dots, M$, which is distributed around 1. Such a choice is clear from the reparametrization (4.5). According to (4.5) large deviations from 1 for some σ'_m , $m = 2, \dots, M$, correspond to large values for some σ_m^2 , $m = 1, \dots, M$, in the primary model (4.1), what would correspond to insignificance of the regression and spatial effects in these clusters. Therefore we use a normal distribution $N(1, 1)$ truncated on the interval $[0.2, +\infty)$ as prior for σ'_m , $m = 2, \dots, M$.

4.2.2 Bayesian Inference for Probit Model (4.6) Based on Representation (4.7)

In the hierarchical structure of (4.7) the latent variables Z'_i 's link the data \mathbf{Y} with the model parameters α' , \mathbf{b}' and $\sigma^{2'}$. The full conditionals of these parameters do not depend on the binary vector \mathbf{Y} given the vector $\mathbf{Z}' = (Z'_1, \dots, Z'_n)^t$ and therefore represent some standard distributions, which do not need computational expensive MH-steps. Moreover, in contrast to the previous models joint updates for the parameter vectors α' and \mathbf{b}' are available here. Further, the full conditional of \mathbf{Z}' given the data \mathbf{Y} and all the parameters has also a simple form, which is suitable for the direct joint updating. More precisely, since the latent Z'_i 's are conditionally independent given η'_i , $i = 1, \dots, n$, we can immediately reduce the joint update of $[\mathbf{Z}' | \mathbf{Y}, \alpha', \mathbf{b}', \sigma^{2'}, \phi', \tau^{2'}]$ to the individual updates of $[Z'_i | \mathbf{Y}, \alpha', \mathbf{b}', \sigma^{2'}, \phi', \tau^{2'}]$ for $i = 1, \dots, n$. Each of these univariate conditional distributions is equivalent to $[Z'_i | \mathbf{Y}, \alpha', \mathbf{b}', \sigma^{2'}]$, since given \mathbf{b}' the information contained in ϕ' and $\tau^{2'}$ has no influence on \mathbf{Z}' . Moreover, we have $[Z'_i | \mathbf{Y}, \alpha', \mathbf{b}', \sigma^{2'}] = [Z'_i | Y_i, \alpha', b'_{j(i)}, \sigma_{m(i)}^{2'}]$, $i = 1, \dots, n$, due again to the conditional independence. It is easy to see that these distributions are univariate truncated normal with mean $-\mathbf{x}_i^t \alpha' - b'_{j(i)}$ and variance $\sigma_{m(i)}^{2'}$. The truncation interval is $(-\infty, 0]$ (or $[0, \infty)$) when $Y_i = 1$ (or $Y_i = 0$). We use rejection sampling for the generation of truncated univariate normal random variables in the numerical implementation as proposed by Robert (1995).

We proceed now with the parameter updates. First define a design matrix X of $p \times n$ dimension as $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)^t$ and assume a full column rank. Further we define $B = (b_{ij})$ as a $n \times J$ spatial incidence matrix with $b_{ij} = 1$, if $j(i) = j$ and $b_{ij} = 0$, if $j(i) \neq j$. This implies that $B \cdot \mathbf{b}' = (b'_{j(1)}, \dots, b'_{j(n)})^t$. Finally let $\Sigma := \text{cov}(\mathbf{Z}' | \alpha', \mathbf{b}', \sigma^{2'}) = \text{diag}(\sigma_{m(1)}^{2'}, \dots, \sigma_{m(n)}^{2'})$.

For the regression parameter α' update with prior $N_p(\mu_0, \Sigma_0)$ we obtain, that $\alpha' | \mathbf{Z}', \mathbf{b}', \sigma^{2'} \sim N_p(\mu_{\alpha'}, \Sigma_{\alpha'})$ with

$$\begin{aligned} \Sigma_{\alpha'} &= (X \Sigma^{-1} X^t + \Sigma_0^{-1})^{-1} \quad \text{and} \\ \mu_{\alpha'} &= -\Sigma_{\alpha'} (X \Sigma^{-1} (\mathbf{Z}' + B \mathbf{b}') - \Sigma_0^{-1} \mu_0). \end{aligned} \quad (4.9)$$

For a flat improper prior of α' (4.9) can be simplified by replacing the parameters $\mu_0 = \mathbf{0}$ and $\Sigma_0^{-1} = 0$, which gives also a proper distribution.

For the spatial parameter vector \mathbf{b}' its joint full conditional $[\mathbf{b}' | \mathbf{Y}, \mathbf{Z}', \alpha', \sigma^{2'}, \phi', \tau^{2'}] = [\mathbf{b}' | \mathbf{Z}', \alpha', \sigma^{2'}, \phi', \tau^{2'}]$ can be found in a similar way as for α' . Under the J -variate normal $N_J(\mathbf{0}, (Q^{m.P}(\phi'))^{-1})$ prior for $[\mathbf{b}' | \phi', \tau^{2'}]$, the full conditional $[\mathbf{b}' | \mathbf{Z}', \alpha', \sigma^{2'}, \phi', \tau^{2'}] \propto [\mathbf{Z}' | \alpha', \mathbf{b}', \sigma^{2'}] [\mathbf{b}' | \phi', \tau^{2'}]$ is also J -variate normal $N_J(\mu_{\mathbf{b}'}, \Sigma_{\mathbf{b}'})$ with

$$\begin{aligned} \Sigma_{\mathbf{b}'} &= \left(B^t \Sigma^{-1} B + \frac{1}{\tau^{2'}} Q^{m.P}(\phi') \right)^{-1} \quad \text{and} \\ \mu_{\mathbf{b}'} &= -\Sigma_{\mathbf{b}'} B^t \Sigma^{-1} (\mathbf{Z}' + X^t \alpha'). \end{aligned} \quad (4.10)$$

Note that in (4.10) for each update we need to invert the $J \times J$ -dimensional precision matrix of \mathbf{b}' , which may be computationally very expensive, if the number of regions J is large. Since the band structure of the precision matrix $\Sigma_{\mathbf{b}'}^{-1} = B^t \Sigma^{-1} B + \frac{1}{\tau^{2'}} Q^{m.P}(\phi')$ coincides with the band structure of Q (note that

$B^t \Sigma^{-1} B$ is a diagonal matrix) and is therefore known and stable, it might be more efficient to simulate a J -variate normal $N_J(\boldsymbol{\mu}_{\mathbf{b}'}, \Sigma_{\mathbf{b}'})$ vector using the *Cholesky decomposition* of $\Sigma_{\mathbf{b}'}$, which requires an order J^2 operation. It is significantly cheaper for large J , than the inversion of the matrix $\Sigma_{\mathbf{b}'}$ (an order J^3 operation) and more stable (see Gelfand et al. (2000), p.380). Further it is useful to reorder \mathbf{b}' in a way, which provides minimal bandwidth for its precision matrix. One method to facilitate this is the Cuthill-McKee Algorithm (George and Liu 1981) which is implemented in MATLAB by the function *symrcm*. In contrast to Model (4.8) we update here the cluster variance parameters $\sigma_m^{2'}$, instead of σ'_m , $m = 2, \dots, M$, since it can be updated directly under a suitable choice of prior. More precisely, the individual full conditionals for $\sigma_m^{2'}$, $m = 2, \dots, M$, are given by

$$\begin{aligned} [\sigma_m^{2'} | \mathbf{Y}, \mathbf{Z}', \boldsymbol{\alpha}', \mathbf{b}', \boldsymbol{\sigma}_{-m}^{2'}, \phi', \tau^{2'}] &= [\sigma_m^{2'} | \mathbf{Z}', \boldsymbol{\alpha}', \mathbf{b}', \boldsymbol{\sigma}_{-m}^{2'}] \\ &\propto [\mathbf{Z}' | \boldsymbol{\alpha}', \mathbf{b}', \boldsymbol{\sigma}^{2'}] [\sigma_m^{2'}] \propto \frac{1}{(\sigma_m^{2'})^{\frac{K_m}{2}}} \exp \left\{ -\frac{1}{2\sigma_m^{2'}} \sum_{i:m(i)=m} (Z'_i + \mathbf{x}_i^t \boldsymbol{\alpha}' + b'_{j(i)})^2 \right\} \times [\sigma_m^{2'}], \end{aligned} \quad (4.11)$$

where $\boldsymbol{\sigma}_{-m}^{2'} = (1, \sigma_2^{2'}, \dots, \sigma_{m-1}^{2'}, \sigma_{m+1}^{2'}, \dots, \sigma_M^{2'})^t$. If $[\sigma_m^{2'}]$ is $IG(a_0, b_0)$ (see (3.6)), we immediately obtain in (4.11) a $IG(a_{\sigma'}, b_{\sigma'})$ density function, up to a constant, where

$$\begin{aligned} a_{\sigma'} &= \frac{K_m}{2} + a_0 \quad \text{and} \\ b_{\sigma'} &= \left\{ \frac{1}{b_0} + \frac{\sum_{i:m(i)=m} (Z'_i + \mathbf{x}_i^t \boldsymbol{\alpha}' + b'_{j(i)})^2}{2} \right\}^{-1}. \end{aligned} \quad (4.12)$$

As mentioned before in this subsection, it is reasonable to choose a prior for $\sigma_m^{2'}$, $m = 2, \dots, M$, which is distributed around 1. Therefore we took $[\sigma_m^{2'}] \sim IG(3, 0.5)$ for $m = 2, \dots, M$. This gives $\mathbf{E}(\sigma_m^{2'}) = \mathbf{Var}(\sigma_m^{2'}) = 1$. We note finally, that the update of the spatial hyperparameters ϕ' and $\tau^{2'}$ remains the same as described in Subsection 3.2 for Models (3.2) and (4.8). So we have only one parameter, namely ϕ' , that needs a MH-step.

5 Simulation Studies

5.1 Study 1: Hierarchical Spatial Binary Regression with Group Cluster Effects

The simulation study based on the Logit Model (3.2) has the following mean structure:

$$\Theta_i := \log \left(\frac{p_i}{1 - p_i} \right) = x_{1i} \alpha_1 + x_{2i} \alpha_2 + b_{j(i)} + c_{m(i)}, \quad i = 1, \dots, n, \quad j = 1, \dots, J, \quad m = 1, \dots, M.$$

We simulated $n = 2100$ binary responses residing in $J = 70$ regions arranged on a 7×10 regular lattice (i.e. 30 observations per region) and in $M = 5$ clusters (i.e. 420 observations per cluster) so, that each cluster is represented in each region with 6 responses. The number of responses, the number of regions and clusters approximately corresponds to our real mobility data. The regression effect is simulated identically for each of the 4 data sets. More precisely, we chose x_{i1} as categorical covariate with possible values 0 or 1 and x_{i2} as continuous covariate taking cycled integer values between 1 and 23, i.e. $x_{1,2} = 1, x_{2,2} = 2, \dots, x_{23,2} = 23, x_{24,2} = 1, \dots$. With this choice we achieved a good data mixing inside both regions and clusters. The true values for the regression parameters were taken as $\alpha_1 = -1$ and $\alpha_2 = 0.05$. Spatial effects b'_j , $j = 1, \dots, 70$, are simulated from the modified Pettitt's Model (2.5) with $\phi = 25$ giving strong spatial smoothing. We chose $\tau^2 = 0.64$ which gives a similar range of the observed spatial effects and the regression effects. For the neighborhood structure we chose a first order neighborhood dependence defined by joint borders. Finally, for each data set we simulated group cluster effects from $\mathbf{c} \sim N_5(0, \sigma_c^2)$ with $\sigma_c^2 = 1$.

For the MCMC estimation we chose $\alpha_1 \sim N(0, 100^2)$ and $\alpha_2 \sim N(0, 10^2)$ reflecting a diffuse prior choice. For the variance hyperparameters τ^2, σ_c^2 we chose flat priors, while for $\psi = \frac{\phi}{1+\phi}$, $\psi \in [0, 1)$ we chose as prior density $[\psi] \sim \frac{1}{(1-\psi)^{1+a}}$ with $a = 0.5$. This corresponds to a Pareto distribution for $\phi = \frac{\psi}{1-\psi}$, $\phi \in [0, +\infty)$, namely $[\phi] \sim \frac{1}{(1+\phi)^{1+a}}$.

MCMC algorithm of Section 3.2 was implemented in MATLAB and was run for 50,000 iterations with every 50th iteration recorded. As "burn in" phase served 10 pilot runs with 300 iterations per each pilot run, which we used to determine optimal proposal standard error values for the MH-step (i.e. an acceptance rate of 30 – 60%). The resulting trace plots (not shown) show that such a length of "burn in" phase is enough. The autocorrelation plots (not shown) indicate, that the autocorrelations between recorded iterations are below 0.1. Figure 5.1 shows marginal posterior density estimates of the parameters $\alpha_0, \alpha_1, \psi = \frac{\phi}{1+\phi}, \tau^2$ and σ_c^2 from the 4 data sets, where the vertical fat dashed lines correspond to the true parameter value. For each density curve its mode is also marked by a thin vertical line. From this we see that in all four cases the true values are well inside 90% credible intervals. Further we see that ψ is overestimated, however other simulation studies show that when using $a = 1$ (i.e. $\psi \sim Uni[0, 1]$) the parameter ψ is underestimated. We note that another simulation study, conducted using an improper prior choice with $a = -1 (\Leftrightarrow [\phi] \propto 1)$ indicates that in this case the posterior for ψ (and for ϕ) is improper. Despite the overestimation of ψ , posterior mode estimates of the spatial and cluster effects (not shown) are quite precise. We also calculated absolute and relative errors between the true parameter values and their posterior mode estimators. These estimates as well as graphical checks with estimated posterior densities for $J = 70$ spatial and $M = 5$ cluster effects can be found in Prokopenko 2004 (Table 5.1 and Figure 5.5 respectively).

5.2 Study 2: Hierarchical Spatial Binary Regression with Individual Cluster Effects using Logit Model (4.8)

The simulation study based on the Logit Model (4.8) has the following mean structure:

$$\Theta'_i := \log \left(\frac{p_i}{1 - p_i} \right) = \frac{x_{1i}\alpha'_1 + x_{2i}\alpha'_2 + b'_{j(i)}}{\sigma'_{m(i)}}, \quad i = 1, \dots, n, \quad j = 1, \dots, J, \quad m = 1, \dots, M.$$

As before we chose $\alpha'_1 = -1, \alpha'_2 = 0.05, \tau^{2'} = 0.64, \phi' = 25$. As true values for the cluster parameters $\sigma'_m, m = 2, \dots, M$, we take 4 samples from a $Uni[0.75, 1.25]$ distribution. According to Model (4.8) we set $\sigma'_1 = 1$. In this way we have as true cluster parameter values: $\sigma' = (1, 1.2251, 0.8656, 1.0534, 0.9930)^t$. We generated 4 data sets with $n = 2100$ binary responses using the same $J = 70$ regions and $M = 5$ clusters. The same specification of the fixed regression effects $x_{1i}\alpha'_1 + x_{2i}\alpha'_2, i = 1, \dots, n$, and the spatial effects $b_{j(i)}, i = 1, \dots, n$, is used. Prior choices for $\alpha', \mathbf{b}', \tau^{2'}, \psi'$ remain the same, while for the prior distribution of the cluster parameters $\sigma'_m, m = 1, \dots, M$, we used $N(1, 1)$ truncated on the interval $[0.2, +\infty]$.

Figure 5.2 gives posterior density estimates of all parameters using the MCMC algorithm of Section 4.2.1 based on 50000 iterations with every 50th iteration recorded. Estimated posterior location measures and quantiles can be found in Prokopenko 2004 (Table 5.2). In contrast to Simulation Study 1 we note that for this model ψ' is no longer overestimated.

5.3 Study 3: Hierarchical Spatial Binary Regression with Individual Cluster Effects using Model (4.6)

The only difference between this model and the previous Model (4.8) is that now we use a probit link instead of a logit one as in Model (4.8). Therefore the generation of the data sets is similar and uses the same parameter values $\alpha', \mathbf{b}', \sigma', \tau^{2'}, \phi'$ as before. However, the MCMC algorithm is based now on the utilization of the latent variables $Z'_i, i = 1, \dots, n$ (for details see Section 4.2.3) which allow a direct Gibbs step for all parameters except ψ' . Moreover, now in contrast to Model (4.8) a joint update for α' and \mathbf{b}' is possible. One further difference is that in this model we update the parameter $\sigma^{2'}$ instead of σ' as before. As a prior for the parameter $\sigma_m^{2'}, m = 2, \dots, M$, we use an $IG(3, 0.5)$ distribution (with $\mathbf{E}(\sigma_m^{2'}) = \mathbf{Var}(\sigma_m^{2'}) = 1$), while the other parameters have the same priors as before.

Since the corresponding MCMC algorithm requires double computation time compared to the previous simulation studies we run 25,000 instead of 50,000 iterations with every 25th iteration recorded. However the corresponding autocorrelation plots (not shown here) indicate even better mixing as by the previous models. This effect is due to joint updates. Unfortunately, the precision of the estimators in this study is not as good as before. This holds especially for the variance parameters $\sigma_m^{2'}, m = 2, \dots, M$. Also for the

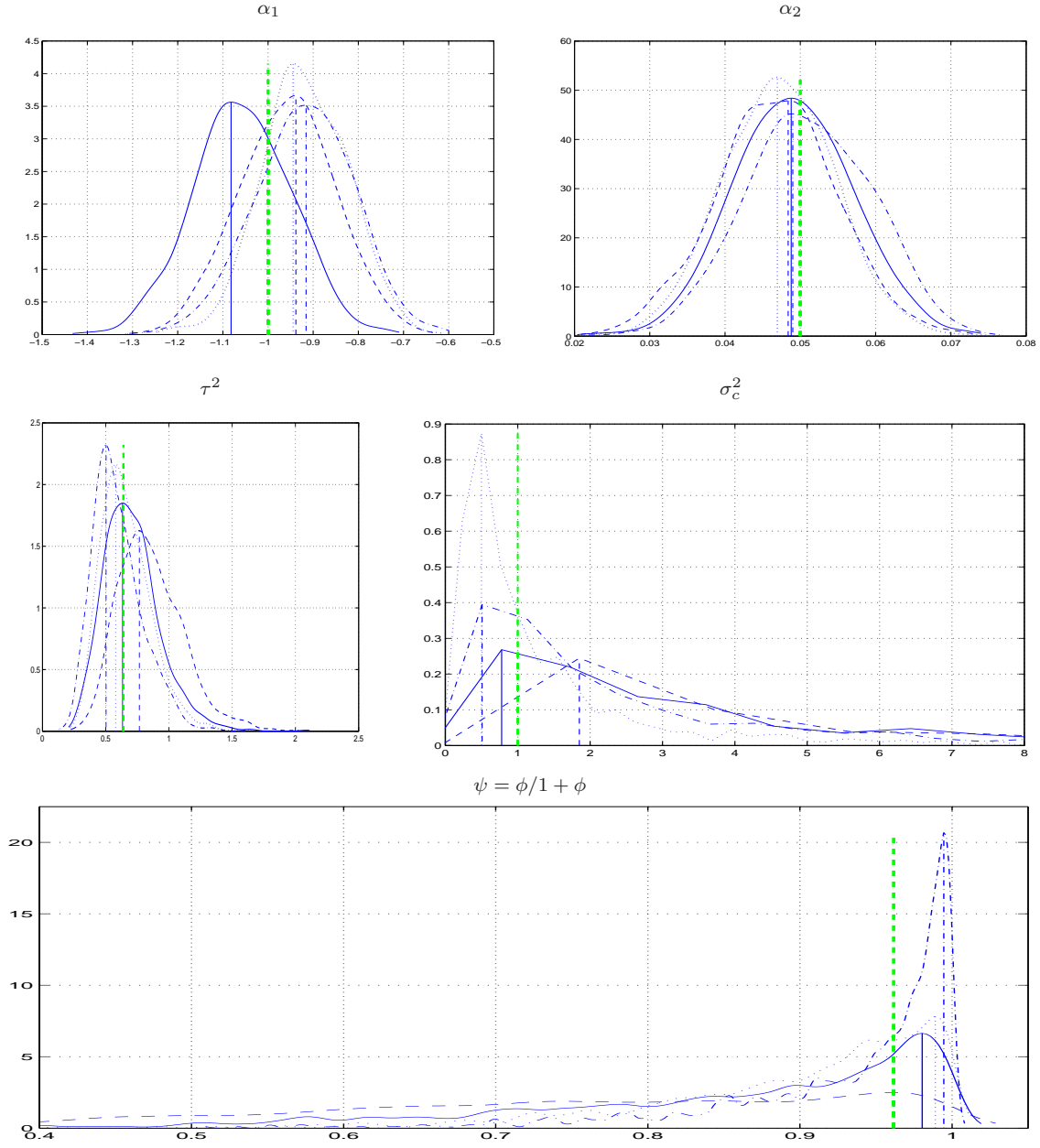


Figure 5.1: Study 1: Estimated Marginal Posterior Densities for Parameters $\alpha_1, \alpha_2, \tau^2, \psi, \sigma_c^2$ in Model (3.2) (solid for Data Set 1, dashed for Data Set 2, dash-dot for Data Set 3, dotted for Data Set 4)

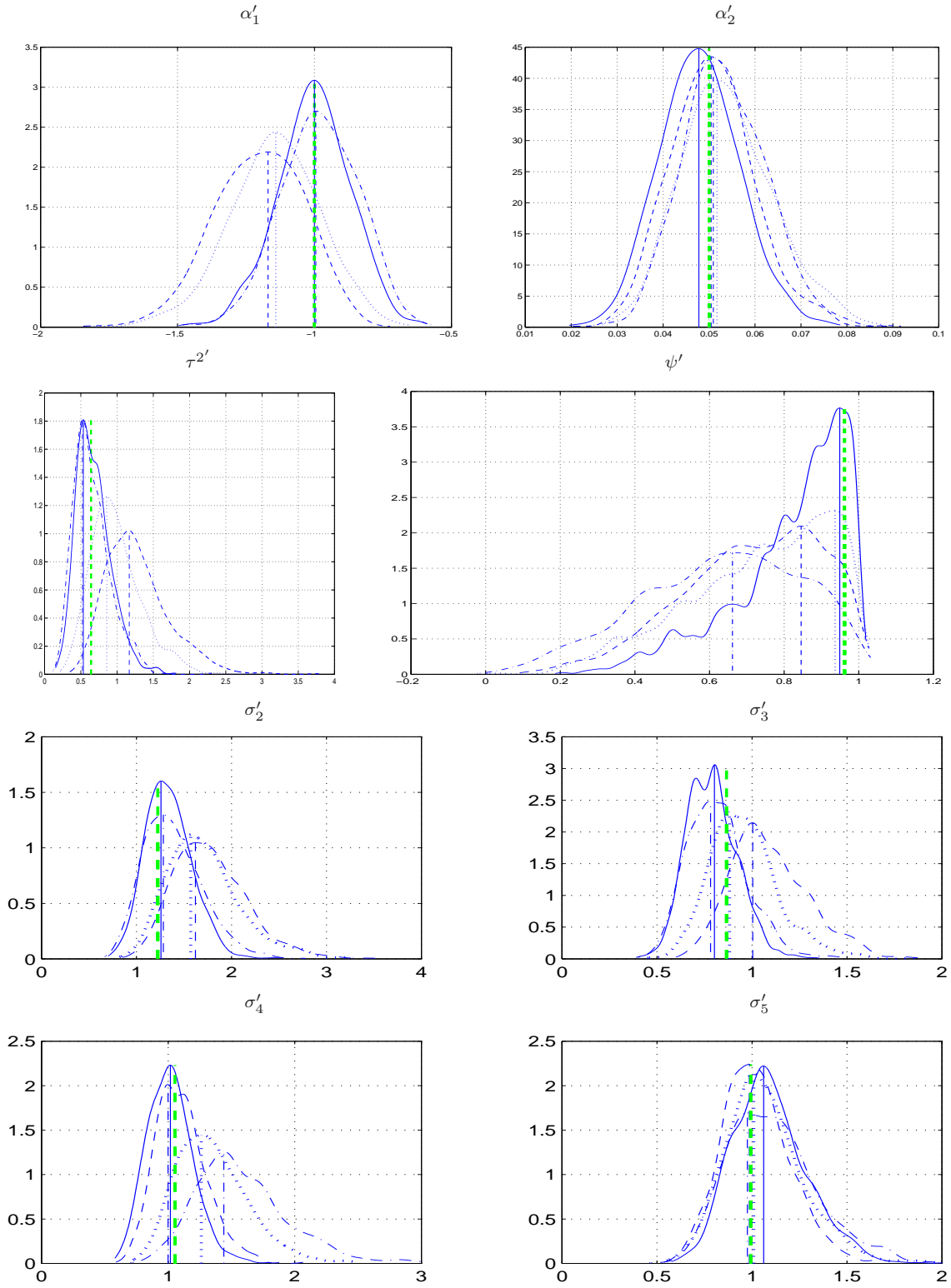


Figure 5.2: Study 2: Estimated Marginal Posterior Densities for Parameters α' , $\tau^{2'}$, ψ' , σ' in Model (4.8) (solid for Data Set 1, dashed for Data Set 2, dash-dot for Data Set 3, dotted for Data Set 4)

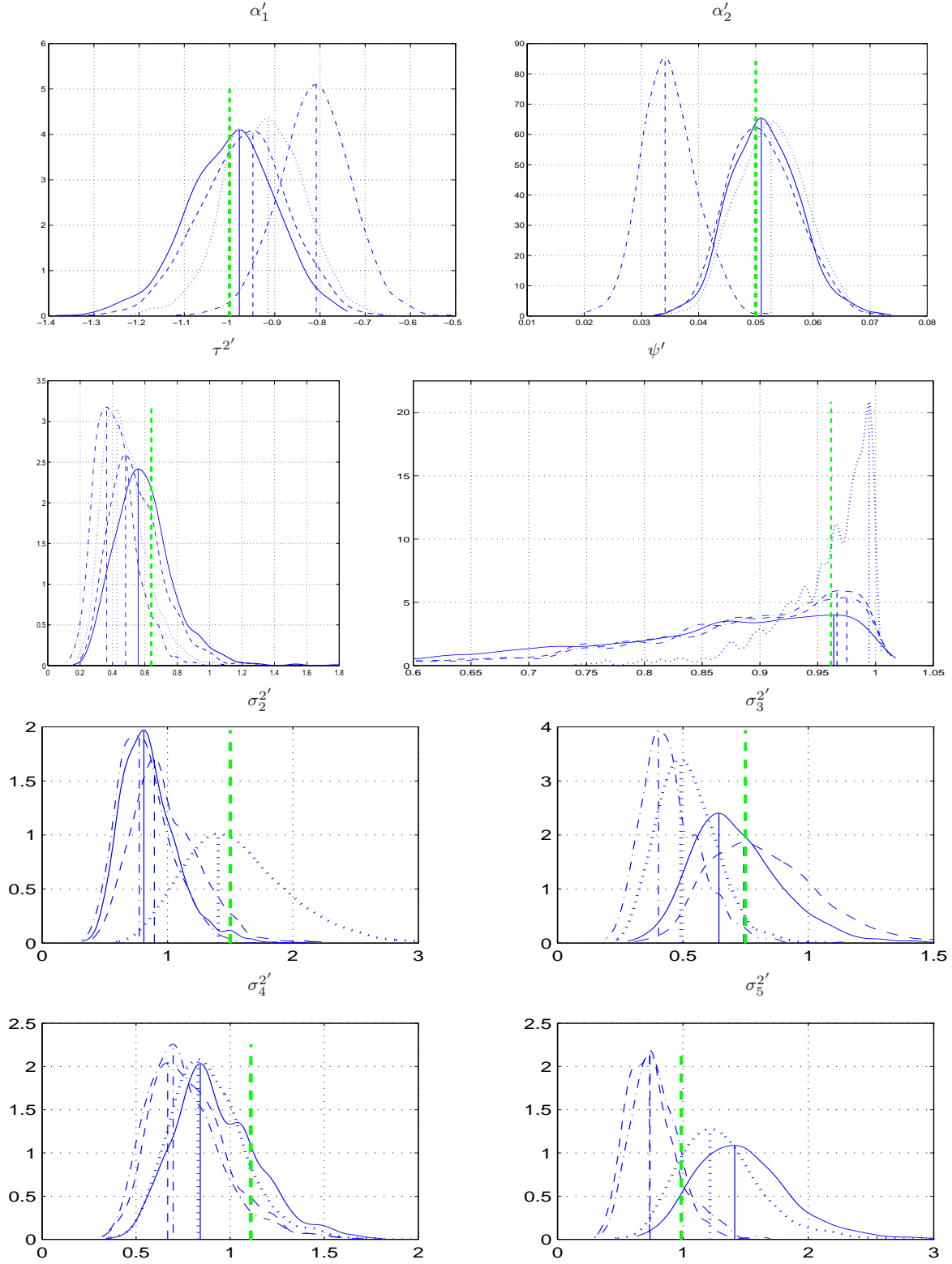


Figure 5.3: Estimated Marginal Posterior Densities for Parameters α' , $\tau^{2'}$, ψ' , $\sigma^{2'}$ in Model (4.6) (solid for Data Set 1, dashed for Data Set 2, dash-dot for Data Set 3, dotted for Data Set 4)

regression parameter α' the estimator based on the third data set has relatively large deviations from the corresponding true values.

5.4 Summary of Simulation Results

First of all, trace plots of the realized MCMC Chains from Models (4.8) and (4.6) with individual cluster effects indicate identifiability.

To consider the random variability we simulated 4 data sets, which provided 4 posterior estimates of the corresponding parameter vector. For group cluster Model (3.2) all 4 posterior estimates for the fixed α , spatial \mathbf{b} and cluster \mathbf{c} parameters lie quite closely around the corresponding true values. The true values are also well inside 90% credible intervals. However the spatial hyperparameter ψ , if close to 1, is often overestimated when a prior $[\psi] \sim \frac{1}{(1-|\psi|)^{1-a}}$ with $a = 0.5$ (see Figure 5.1) and underestimated when $a = 1$ is chosen. For the Logit Model (4.8) the hyperparameter ψ is not overestimated by a prior choice of $[\psi] \sim \frac{1}{(1-|\psi|)^{1-a}}$ with $a = 0.5$, while the estimates for the other parameters remain close to the corresponding true values (see Figure 5.2, and Table 5.2 in Prokopenko (2004)). Finally for the Probit Model (4.6) the estimates for α' , \mathbf{b}' , $\tau^{2'}$ and ψ' are usually still sufficient, but not as precise, as for former models. This is especially true for the cluster parameter $\sigma^{2'}$ (see Figure 5.3).

The MCMC estimating algorithms are fast enough so, that we were able to simulate long Markov Chains (50,000 iterations for Models (3.2), (4.8) and 25,000 iterations for Model (4.6)), with every 50th or 25th iteration respectively recorded. The running time was between 10 and 15 hours. This provided a good mixing for each model, which were evident by low observed autocorrelations. This is especially true for Model (4.6), where the estimating MCMC algorithm allows block updating and reduces the number of parameters, whose updates require a MH-step to only one. Finally we note, that the length of the “burn in” phase, which consists of 10 pilot runs is enough to achieve the stationary phase of the corresponding Markov Chains.

6 Application: Mobility Data

6.1 Data Description

We analyze a data set studying mobility behavior of private households in Munich. One central question is to identify areas of low/high utilization of public transport after adjusting for explanatory factors such as trip, individual and household related attributes. The goal is to find flexible statistical models which incorporate covariates together with spatial and cluster information.

The data was collected within the study “Mobility 97” (see Zängler 2000). The participants of the survey are German-speaking persons not younger than 10 years, which live in a private household in the state of Bavaria. In order to take into consideration seasonal fluctuations in mobility behavior of the participants, the survey was carried out in three waves in March, June and October of 1997. Each participant reported all his or her trips conducted by public or individual transport during a period of two or three days. We consider part of the data which includes 1375 trips taken by 296 persons in 167 households in the city of Munich, Germany.

For each trip the binary variable of interest Y has value 1, if individual transport was used and value 0, if public transport was used. In addition person, household and trip related covariates were recorded. Neglecting spatial and cluster effects standard model selection techniques for logistic regression selected the following covariates. Person related covariates are age (metric), sex, personal income, car usage (main, secondary or not user) and whether the person possesses or not a public transport net card. We retain only one household related covariate, namely household type (single, single parent or not single). Trip related covariates are day type (work day or weekend), day time (day or night), distance and whether the person took the trip alone or not alone. Table 6.1 shows the chosen covariates. For the covariate **USAGE**, note that both main and secondary users must be not younger than 18 years and must have a driver license and a car available in the household. We use standard model selection procedures for binary regression models ignoring spatial and cluster effects to identify a starting model for our analysis.

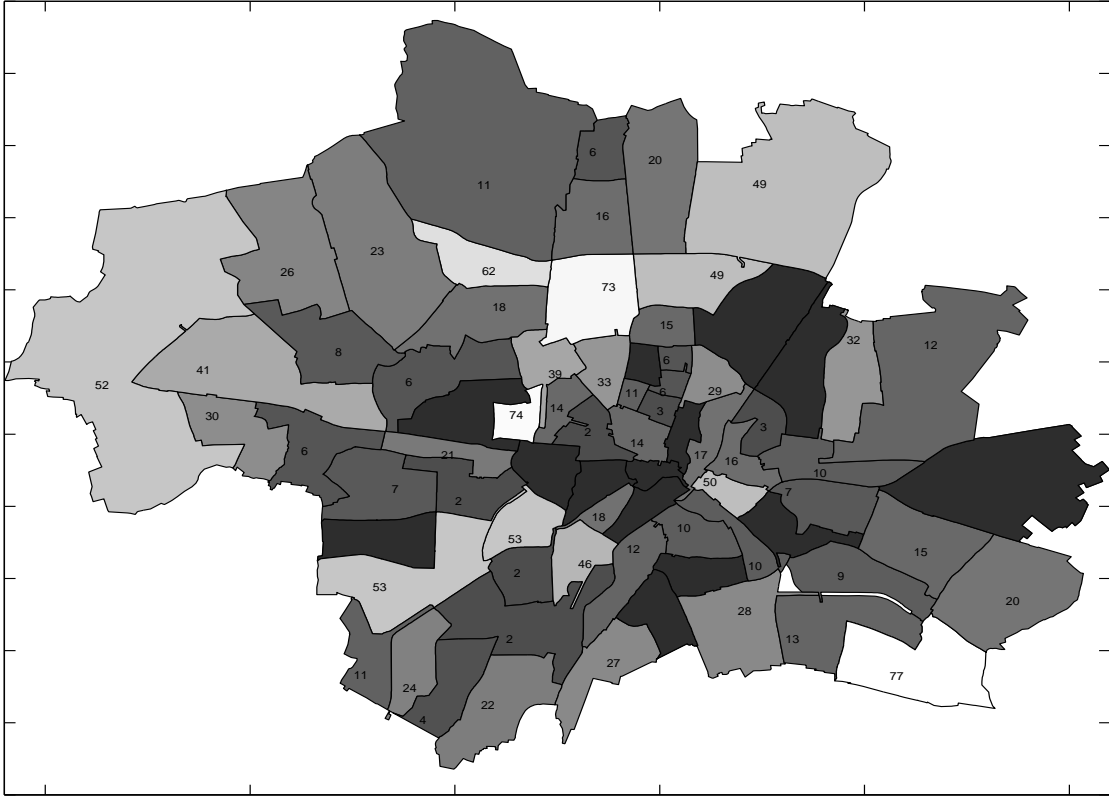


Figure 6.1: Number of Available Trips over Postal Codes of Munich, Germany

This approach resulted in selecting main effects and 10 significant interactions:

| | |
|--------------------|--------------------------|
| WAY ALONE:NET CARD | USAGE:SEX |
| WAY ALONE:USAGE | DISTANCE:USAGE |
| DAY TYPE:NET CARD | USAGE:DAY TIME |
| SEX:DAY TIME | PERSONAL INCOME:NET CARD |
| DISTANCE:AGE | DAY TYPE:AGE |

6.2 Results

We present the results for 11 different model specifications. Model 1 is a spatial logit regression model with no cluster effects, while Models 2 — 5 are spatial logit regression models with group cluster effects. Finally Models 6 — 11 are binary spatial regression models with individual cluster effects. Here both logit (based on (4.8)) and probit (based on (4.6)) link specifications are used. For the first 5 models as well as for Models 6, 8, 10 the 25000 MCMC iterations were run and every 25th iteration was recorded, giving acceptable low autocorrelations (not shown). Models 6, 8, 10 are based on Logit Model (4.8). Finally, Models 7, 9, 11 have the same cluster choice as Models 6, 8, 10, respectively, but are based on Probit Model (4.6). For probit models we run 20000 iterations (and recorded every 20th iteration) of the MCMC algorithm based on latent variable representation (4.7). We found, that 10 pilot runs (5 pilot runs for probit models) with 300 iterations per pilot run are sufficient as “burn in”.

As a starting point for the choice of fixed effects for each of the 11 models we used the covariates identified in Table (6.1) involving a total of 36 regression parameters. The intercept effect is modeled within the spatial and cluster part. As prior distributions for $\alpha_1, \dots, \alpha_{36}$ we chose independent normal

| Covariable type | Variable | Levels | Number of Trips Using | | | Most frequently obs. value |
|-------------------------|---|------------------------|-----------------------|------------------|-------|----------------------------|
| | | | Individual Transport | Public Transport | Total | |
| PERSON related | PERSONAL INCOME | NO INCOME (< 200 DM) | 24 | 31 | 55 | 0 |
| | | MIDDLE (200 – 3000 DM) | 475 | 193 | 668 | 1 |
| | | HIGH (> 3000 DM) | 521 | 131 | 652 | 0 |
| | USAGE | MAIN USER | 731 | 100 | 831 | 1 |
| | | SECONDARY USER | 213 | 99 | 312 | 0 |
| | | NOT USER | 76 | 156 | 232 | 0 |
| NET CARD | YES | 235 | 247 | 482 | 0 | |
| | NO | 785 | 108 | 893 | 1 | |
| SEX | MALE | 549 | 172 | 721 | 1 | |
| | FEMALE | 471 | 183 | 654 | 0 | |
| AGE | metric (quadratic, normalized with Splus function $\text{poly}(\text{age},2)$) | | | | | median 42 years |
| HOUSEHOLD related | HOUSEHOLD TYPE | SINGLE | 156 | 125 | 281 | 0 |
| | | SINGLE PARENT | 84 | 10 | 94 | 0 |
| | | NOT SINGLE | 780 | 220 | 1000 | 1 |
| | | | | | | |
| TRIP related | DAY TYPE | WORK DAY | 595 | 297 | 892 | 1 |
| | | WEEKEND | 425 | 58 | 483 | 0 |
| | DISTANCE | SHORT (≤ 3.5 km) | 294 | 71 | 365 | 0 |
| | | MIDDLE (3.6 – 21.5 km) | 571 | 257 | 828 | 1 |
| | | FAR (> 21.5 km) | 155 | 27 | 182 | 0 |
| | WAY ALONE | ALONE | 507 | 267 | 774 | 1 |
| | | NOT ALONE | 513 | 88 | 601 | 0 |
| | DAY TIME | DAY (6 a.m. - 9 p.m.) | 905 | 336 | 1241 | 1 |
| NIGHT (9 p.m. - 6 a.m.) | | 115 | 19 | 134 | 0 | |
| T O T A L | | | 1020 | 355 | 1375 | |

Table 6.1: Significant covariates identified in logistic regression model selection without spatial and cluster effects

distributions with zero mean and standard error equal to 5. We consider an interaction as insignificant when the corresponding estimated 90% credible interval contains the zero value for all interaction terms. If an interaction is found to be insignificant, then the corresponding terms will be removed and the model parameters will be estimated again using the appropriate MCMC algorithm. Continuing with this procedure we arrive at a model where all interactions are significant.

For Model 1 we chose as prior for $\psi = \frac{\phi}{1+|\phi|}$ a uniform distribution on $(-1, 1)$, while for τ^2 we chose an non-informative prior, i.e. $[\tau^2] \propto 1$.

The top row of Figure 6.2 presents estimated posterior densities for the hyperparameters τ^2 and ψ , respectively. Note that the spatial dependence parameter ψ is negative, which indicates that large positive spatial effects in an area can be surrounded by negative spatial effects and vice versa. This behavior is seen on the bottom maps of Figure 6.2, where posterior means and their 90% credible intervals (*CI*) of the spatial effects are given. Here we use the following color code: white if 0 is below 90% *CI*, gray if 0 is contained in 90% *CI* and black if 0 is above 90% *CI*. We note that spatial effects in postal codes with no observations are insignificant.

We now consider the Logit Models 2 — 5 with group cluster effects. In Model 2 we chose as cluster groups the 74 postal codes, while the cluster groups specification for Models 3 and 5 are given in Table (6.2). Model 4 uses 12 cluster groups formed by number of trips a household has taken. For the cluster variance σ_c^2 we choose $\sigma_c^2 \sim IG(3, 0.5)$, while prior choices for fixed and spatial parameters remain the same as in Model 1. Only in Model 2, in order to avoid numerical problems (clustering around border values -1 and 1) we chose $[\psi] \propto (1 - |\psi|)^{0.5}$ instead of $[\psi] \propto 1$ on the interval $(-1, 1)$. The posterior centrality estimates of the hyperparameters and their 90% credible intervals are given in Table 6.3.

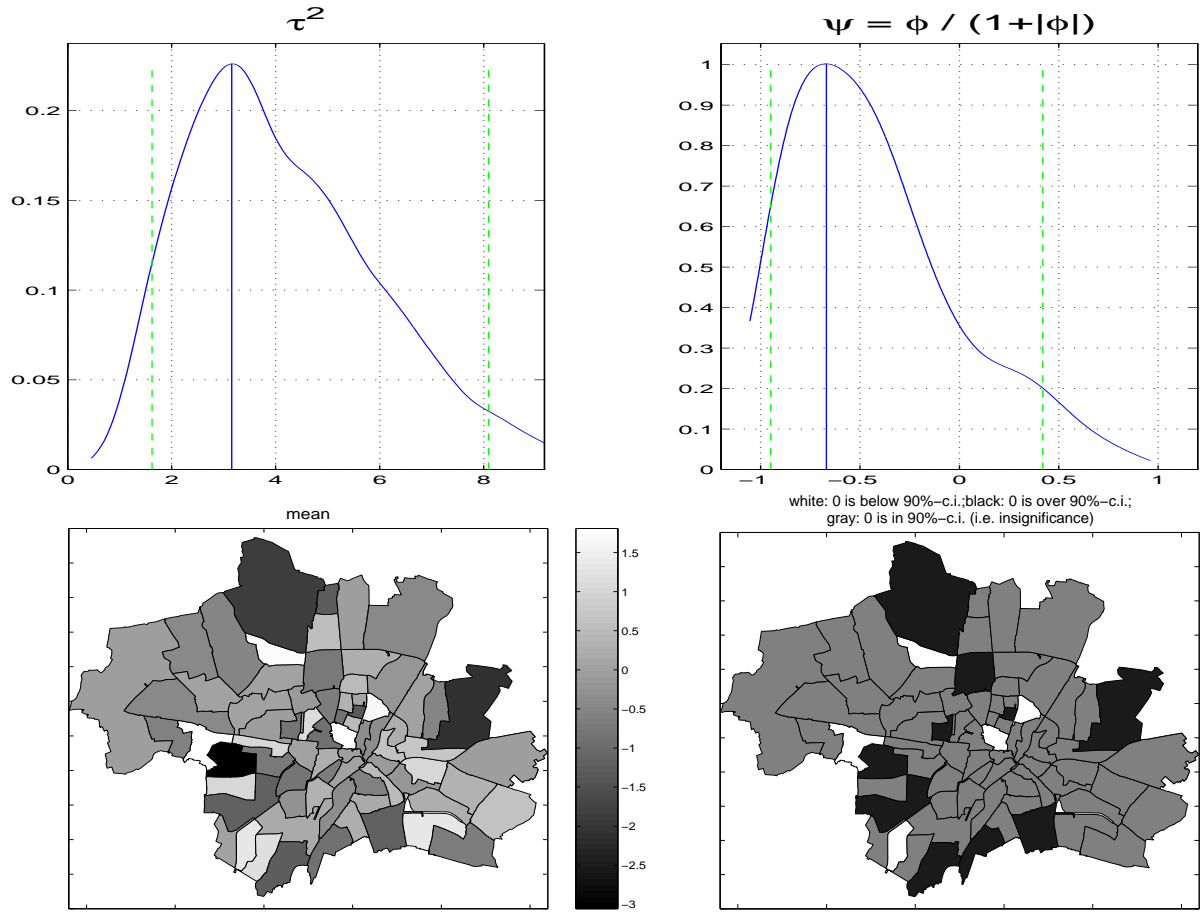


Figure 6.2: Results for Model 1: Top: Estimated Posterior Densities of Spatial Hyperparameters (Solid Line = Estimated Posterior Mode, Dashed Line = 90% *CI*). Bottom: Estimated Spatial Effects \hat{b}_j , $j = 1, \dots, 74$ and 90% *CI*

| Model | Cluster | Description | Total |
|----------|-----------------|--|-----------|
| 3 | 1 st | Households which conducted ≥ 23 trips | 275 trips |
| | 2 nd | Households which conducted 16 – 22 trips | 296 trips |
| | 3 rd | Households which conducted 12 – 15 trips | 250 trips |
| | 4 th | Households which conducted 8 – 11 trips | 275 trips |
| | 5 th | Households which conducted ≤ 7 trips | 279 trips |
| Model | Cluster | Description | Total |
| 5 | 1 st | Persons which conducted ≥ 12 trips | 309 trips |
| | 2 nd | Persons which conducted 9 – 11 trips | 301 trips |
| | 3 rd | Persons which conducted 7 – 8 trips | 240 trips |
| | 4 th | Persons which conducted 5 – 6 trips | 285 trips |
| | 5 th | Persons which conducted ≤ 4 trips | 240 trips |

Table 6.2: Distribution of Trips into Clusters in Models 3 and 5

| Model | Number of Clusters | Parameter | Mode | Mean | Median | 90% 5% | CI 95% |
|-------|---|--------------|--------|--------|--------|-----------|-----------|
| 2 | 74 formed by postal codes | ψ | -0.500 | -0.271 | -0.372 | -0.857 | 0.646 |
| | | τ^2 | 3.628 | 4.777 | 4.313 | 0.981 | 10.335 |
| | | σ_c^2 | 0.554 | 0.836 | 0.678 | 0.315 | 1.912 |
| 3 | 5 formed by # of trips per household | ψ | -0.541 | -0.422 | -0.446 | -0.930 | 0.149 |
| | | τ^2 | 6.262 | 9.124 | 8.233 | 3.358 | 18.417 |
| | | σ_c^2 | 0.802 | 1.270 | 1.076 | 0.486 | 2.797 |
| 4 | 12 formed by # of trips per household | ψ | -0.507 | -0.516 | -0.538 | -0.954 | 0.031 |
| | | τ^2 | 6.293 | 8.299 | 7.452 | 3.194 | 16.067 |
| | | σ_c^2 | 0.880 | 1.272 | 1.122 | 0.589 | 2.398 |
| 5 | 5 formed by # of trips per person | ψ | -0.874 | -0.543 | -0.594 | -0.956 | 0.058 |
| | | τ^2 | 4.025 | 5.298 | 4.777 | 2.020 | 9.685 |
| | | σ_c^2 | 0.526 | 0.753 | 0.646 | 0.324 | 1.585 |

Table 6.3: Point and Interval Posterior Estimates for the Hyperparameters in Models 2 - 5 (with Group Cluster Effects)

In Model 2 we have as cluster groups the 74 postal codes. Therefore both structured (b_j , $j = 1, \dots, 74$) and unstructured (c_j , $j = 1, \dots, 74$) spatial effects are included in Model 2. In Figure 6.3 we present spatial maps with estimated posterior means for the structured spatial effects b_j (top left) and unstructured spatial effects c_j (top middle). On the top right map we present estimated posterior means of the sum $b_j + c_j$ of structured and unstructured spatial effects. Corresponding 90% CI are given in the middle row of Figure 6.3. It is remarkable that both structured and unstructured effects are insignificant, while their sum is, and form a similar spatial pattern as in Model 1. Therefore it is not surprising that the posterior density of ψ , in particular the posterior mode estimate in Models 1 and 2 are also similar (see bottom row of Figure 6.3).

In Figure 6.4 we present for Model 3 estimated posterior densities of the group cluster effects c_m , $m = 1, \dots, 5$. A cluster effect is significant (marked with *), if its 90% credible interval does not include zero. Note that cluster effects for households with large numbers of trips are positive and cluster effects for households with few numbers of trips are negative. Finally the maps on the bottom row of Figure 6.5 give estimated spatial effects.

Similarly to Models 3 and 4 in Model 5 only the higher cluster effects (i.e. with fewest numbers of trips), namely the 4th and the 5th are significant. Again, both have negative values, namely around -1 , i.e. the probability to use public transport for the corresponding trips is higher. We omit the corresponding density plots to save space. For the last 2 models we also omit figures with the estimated spatial effects maps since their spatial patterns are similar to the ones of Models 1 or 3 and Model 2 when the joint effect of structured and unstructured spatial components is considered. Therefore the posterior density of the spatial dependence parameter ψ , in particular the estimate (the posterior mode) also remains similar (not shown for Models 3, 4 and 5). This can be seen in Table 6.3 where posterior centrality estimates and 90% credible intervals for the hyperparameters are given.

We consider now the model specifications with individual cluster effects. Their exact model specification is given by the second and third column of Table 6.4. As before, we chose a flat prior $[\tau^{2'}] \propto 1$. We take the prior $[\psi'] \propto (1 - |\psi'|)^{0.5}$ to avoid numerical problems (clustering around border values -1 and 1). In Probit Models 7, 9, 11 the cluster parameters $\sigma_2^{2'}, \dots, \sigma_M^{2'}$ have an inverse gamma prior given by $IG(3, 0.5)$ (with expectation and variance equal to 1). This choice allows direct Gibbs sampling for

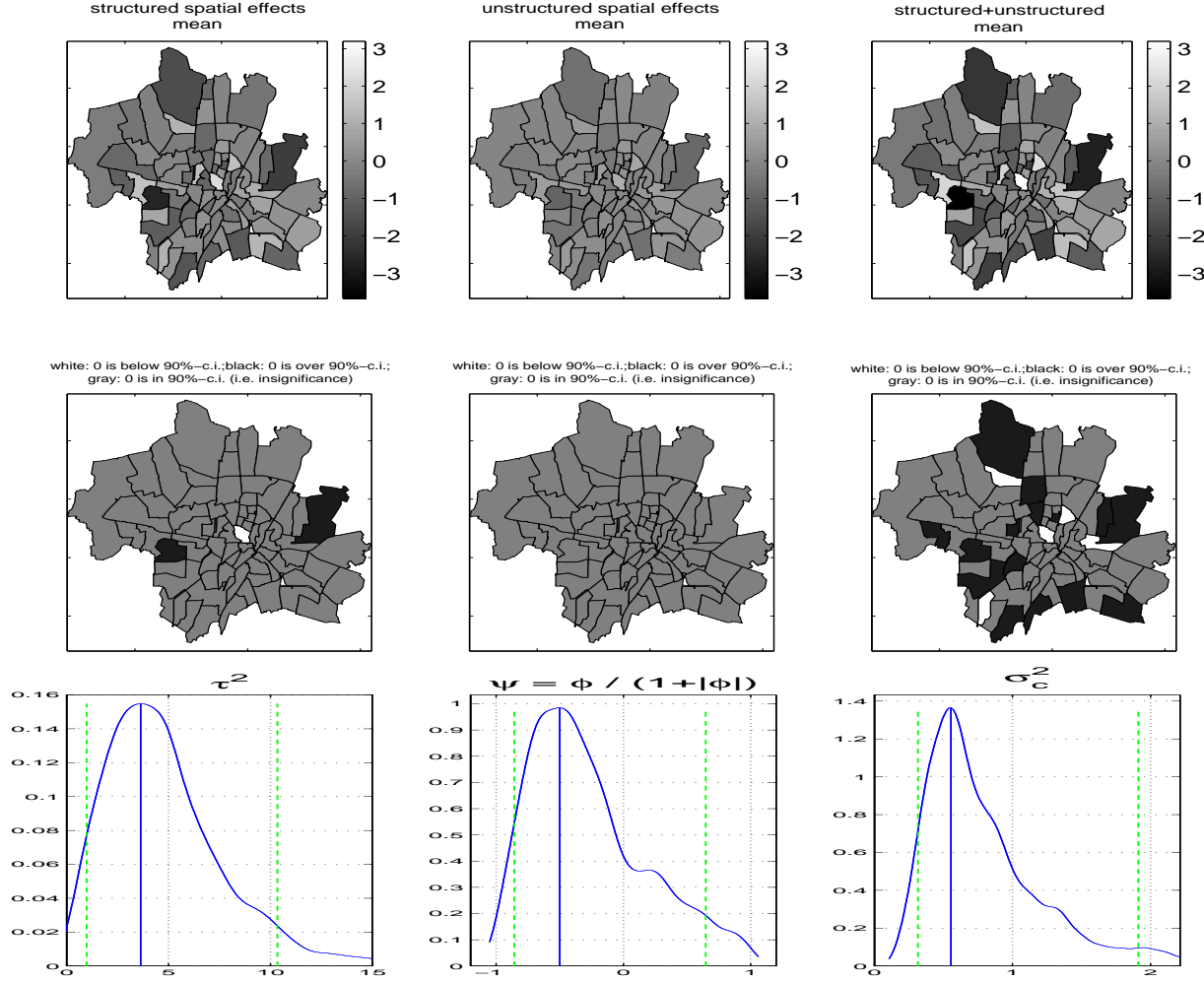


Figure 6.3: Results for Model 2: Top: Estimated Spatial Effects: Structured $\hat{b}_j, j = 1 : 74$ (left), Unstructured $\hat{c}_j, j = 1 : 74$ (middle) and their Sum $\widehat{b_j + c_j}, j = 1 : 74$ (right). Middle: 90% Credible Intervals for Structured Effects, Unstructured Effects and their Sum. Bottom: Estimated Posterior Densities of Hyperparameters (Solid Line = Estimated Posterior Mode, Dashed Line = 90% CI)

these parameters. In Logit Models 6, 8, 10 we assume for $\sigma'_2, \dots, \sigma'_M$ a normal $N(1, 1)$ prior truncated to $(0.2, +\infty)$. The posterior centrality estimates and their 90% CI's of the hyperparameters for models with individual cluster effects are given in Table 6.4. From Table 6.4 we see that cluster components of the higher clusters are significant, i.e. $1 \notin 90\% \text{ CI}$. It means (Models 8 — 11) that the heterogeneity within the group with the fewest numbers of trips per household (or per person) is the largest. Further we see from Table 6.4, that more cluster components are significant for individual cluster effects formed by household type or number of trips per household than by the number of trips per person.

In all models with individual cluster effects the spatial dependence hyperparameter ψ is negative and about the same size. The link specification (logit versus probit) however influences the size of τ^2 , which is expected since the link influences the size of the regression parameters as well.

The estimates for the fixed effects α' for all 11 models are given in Table 6.5. Posterior mode estimates are marked with *, when the corresponding parameter is insignificant, i.e. the 90% credible interval contains zero. If all terms of an interaction effect were insignificant, the model was reduced on this interaction and the model parameters were estimated again using the appropriate MCMC algorithm. Those interactions are marked with “n.r.”, correspond to “not represented” in the model.

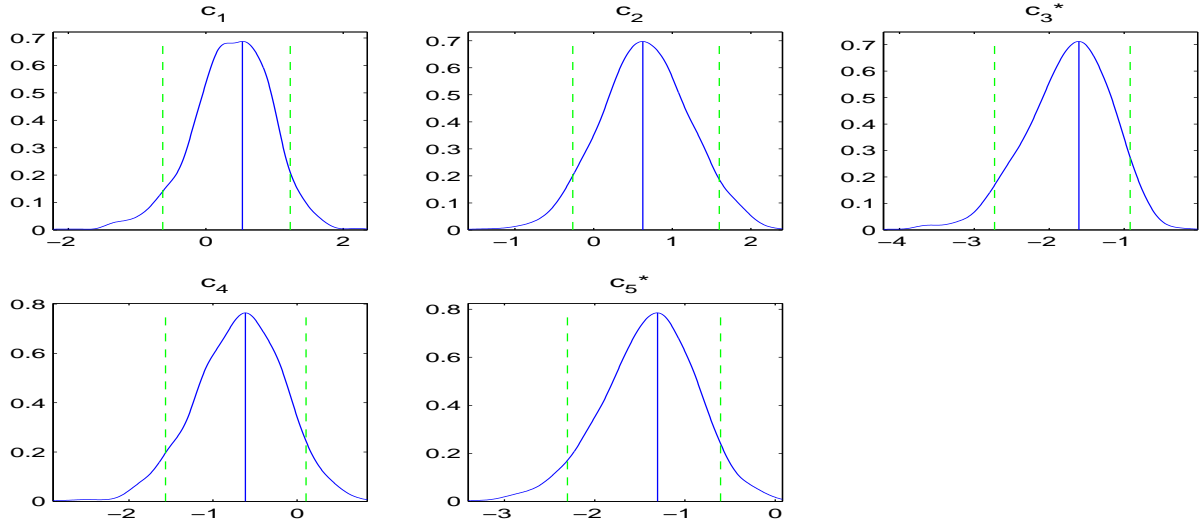


Figure 6.4: Estimated Posterior Densities of Group Cluster Effects c_m , $m = 1, \dots, 5$ in Model 3. (Solid Line = Estimated Posterior Mode, Dashed Line = 90% CI)

| Model | Link | Number of Clusters | Parameter | Mode | Mean | Median | 90% 5% | CI 95% |
|-------|--------|--|-----------------|--------|--------|--------|-----------|-----------|
| 6 | logit | 3 formed by household type | ψ | -0.468 | -0.396 | -0.418 | -0.870 | 0.181 |
| | | | $\tau^{2'}$ | 4.861 | 6.854 | 5.931 | 2.553 | 14.196 |
| | | | σ_2' | 0.277 | 0.484 | 0.430 | 0.226 | 0.921 |
| | | | σ_3' | 1.439 | 1.461 | 1.443 | 1.068 | 1.943 |
| | | | | | | | | |
| 7 | probit | 3 formed by household type | ψ | -0.525 | -0.408 | -0.444 | -0.874 | 0.111 |
| | | | $\tau^{2'}$ | 1.685 | 2.283 | 2.095 | 0.903 | 4.418 |
| | | | $\sigma_2^{2'}$ | 0.404 | 0.540 | 0.497 | 0.252 | 0.973 |
| | | | $\sigma_3^{2'}$ | 1.672 | 1.831 | 1.763 | 1.062 | 2.763 |
| | | | | | | | | |
| 8 | logit | 5 formed by # of trips per household | ψ | -0.410 | -0.413 | -0.422 | -0.865 | 0.075 |
| | | | $\tau^{2'}$ | 10.769 | 17.101 | 14.799 | 6.002 | 36.512 |
| | | | σ_2' | 0.922 | 1.010 | 0.973 | 0.648 | 1.464 |
| | | | σ_3' | 2.842 | 2.951 | 2.913 | 2.240 | 3.734 |
| | | | σ_4' | 1.430 | 1.486 | 1.459 | 1.078 | 2.019 |
| | | | σ_5' | 1.822 | 1.797 | 1.789 | 1.313 | 2.343 |
| | | | | | | | | |
| 9 | probit | 5 formed by # of trips per household | ψ | -0.512 | -0.382 | -0.412 | -0.854 | 0.195 |
| | | | $\tau^{2'}$ | 3.588 | 5.380 | 4.678 | 1.817 | 11.032 |
| | | | $\sigma_2^{2'}$ | 0.797 | 0.906 | 0.845 | 0.477 | 1.534 |
| | | | $\sigma_3^{2'}$ | 5.844 | 6.949 | 6.477 | 3.707 | 11.627 |
| | | | $\sigma_4^{2'}$ | 1.543 | 1.958 | 1.837 | 0.996 | 3.370 |
| | | | $\sigma_5^{2'}$ | 2.743 | 3.174 | 2.953 | 1.600 | 5.474 |
| | | | | | | | | |
| 10 | logit | 5 formed by # of trips per person | ψ | -0.476 | -0.403 | -0.439 | -0.876 | 0.199 |
| | | | $\tau^{2'}$ | 7.538 | 9.468 | 8.232 | 3.167 | 19.895 |
| | | | σ_2' | 1.027 | 1.058 | 1.041 | 0.752 | 1.430 |
| | | | σ_3' | 1.168 | 1.180 | 1.166 | 0.797 | 1.610 |
| | | | σ_4' | 1.271 | 1.300 | 1.287 | 0.897 | 1.768 |
| | | | σ_5' | 1.553 | 1.681 | 1.642 | 1.196 | 2.255 |
| | | | | | | | | |
| 11 | probit | 5 formed by # of trips per person | ψ | -0.350 | -0.369 | -0.383 | -0.867 | 0.2220 |
| | | | $\tau^{2'}$ | 1.773 | 2.517 | 2.182 | 0.849 | 5.2690 |
| | | | $\sigma_2^{2'}$ | 0.782 | 0.887 | 0.846 | 0.510 | 1.4190 |
| | | | $\sigma_3^{2'}$ | 0.976 | 1.067 | 1.009 | 0.557 | 1.7630 |
| | | | $\sigma_4^{2'}$ | 1.021 | 1.340 | 1.250 | 0.731 | 2.1950 |
| | | | $\sigma_5^{2'}$ | 1.972 | 2.209 | 2.088 | 1.169 | 3.7700 |
| | | | | | | | | |

Table 6.4: Point and Interval Posterior Estimates for the Spatial Hyperparameters and Cluster Parameters in Models 6 - 11 (with Individual Cluster Effects)

| Main Effect | Model | | | | | | | | | | |
|------------------------------------|--------------|-----------------------|--------|--------|--------|----------------------------|--------|--------|--------|-------|--------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| | spatial only | spatial+group cluster | | | | spatial+individual cluster | | | | | |
| PERSONAL INCOME | | | | | | | | | | | |
| MIDDLE | 0.41* | 0.48* | 1.63 | 1.41 | 0.71* | 1.06 | 0.58 | 1.62 | 0.80 | 0.64* | 0.37* |
| HIGH | 0.25* | 0.42* | 1.27 | 1.14 | 0.12* | 0.76 | 0.35* | 1.46 | 0.56 | 0.24* | 0.05* |
| USAGE | | | | | | | | | | | |
| SECOND.USER | 0.38* | 1.09* | 1.27 | 1.16* | 0.88* | 1.11 | 0.73 | 1.23* | 0.51* | 1.51 | 0.80 |
| NOT.USER | -3.87 | -6.41 | -6.52 | -6.52 | -5.90 | -6.38 | -3.63 | -9.99 | -5.39 | -7.44 | -3.95 |
| NET CARD | | | | | | | | | | | |
| NO | 2.07 | 2.67 | 3.03 | 3.32 | 2.72 | 2.78 | 1.70 | 3.90 | 2.28 | 3.11 | 1.64 |
| SEX | | | | | | | | | | | |
| FEMALE | 0.28* | 0.16* | -0.19* | -0.47* | 0.10* | 0.30* | 0.11* | -0.48* | -0.20* | 0.01* | -0.05* |
| AGE | | | | | | | | | | | |
| POLY.AGE.1 | 16.80 | 8.73 | 11.64 | 11.53 | 9.95 | 6.11 | 5.57 | 7.97 | 8.82 | 9.81 | 7.46 |
| POLY.AGE.2 | -13.07 | -8.96 | -9.03 | -8.64 | -9.67 | -8.93 | -7.85 | -9.69 | -8.65 | -7.63 | -8.03 |
| (orthogonal parametrization taken) | | | | | | | | | | | |
| HOUSEHOLD | | | | | | | | | | | |
| SINGLE.PARENT | 1.61 | 3.15 | 3.42 | 2.92 | 3.31 | n. r. | n. r. | 4.24 | 2.62 | 3.65 | 2.17 |
| NOT.SINGLE | 0.70 | 0.68 | 0.25* | 0.27* | 0.90 | n. r. | n. r. | 0.85* | 0.40* | 0.96 | 0.44 |
| DAY TYPE | | | | | | | | | | | |
| WEEKEND | 1.44 | 2.21 | 2.46 | 2.52 | 2.11 | 2.25 | 1.19 | 3.32 | 1.70 | 2.78 | 1.29 |
| DISTANCE | | | | | | | | | | | |
| MIDDLE | -0.96 | -1.15 | -1.06 | -1.17 | -1.05 | -1.29 | -0.61 | -1.90 | -1.04 | -1.16 | -0.59 |
| FAR | 0.32* | 0.81* | 0.98* | 0.83* | 0.97* | 1.21* | 0.64* | 0.78* | 0.65* | 0.85* | 0.33* |
| WAY ALONE | | | | | | | | | | | |
| NOT.ALONE | 1.82 | 2.09 | 2.07 | 2.30 | 1.93 | 2.17 | 1.36 | 3.21 | 1.76 | 2.30 | 1.20 |
| DAY TIME | | | | | | | | | | | |
| NIGHT | -0.58* | -1.02 | -1.12 | -1.29 | -1.13 | -1.19 | -0.63 | -1.99 | -1.08 | -1.30 | -0.64 |
| Interaction | | | | | | | | | | | |
| WAY ALONE:NET CARD | | | | | | | | | | | |
| NOT.ALONE:NO | -1.86 | -2.39 | -2.37 | -2.76 | -2.37 | -2.54 | -1.54 | -3.10 | -1.69 | -2.53 | -1.48 |
| USAGE:SEX | | | | | | | | | | | |
| SECOND.USER:FEMALE | -1.70 | -2.13 | -2.07 | -1.81 | -2.01 | -2.30 | -1.45 | -2.80 | -1.73 | -2.50 | -1.22 |
| NOT.USER:FEMALE | -0.20* | 0.66* | 0.58* | 0.79* | 0.40* | 0.26* | 0.22* | 1.39* | 0.62* | 0.80* | 0.62* |
| WAY ALONE:USAGE | | | | | | | | | | | |
| NOT.ALONE:SECOND.USER | 0.79 | 1.21 | 0.80 | 0.76* | 1.22 | 1.09 | 0.62 | 1.20* | 0.66* | 1.32 | 0.77 |
| NOT.ALONE:NOT.USER | 1.75 | 3.65 | 4.19 | 3.76 | 3.41 | 4.35 | 2.30 | 5.08 | 3.06 | 4.22 | 2.03 |
| DISTANCE:USAGE | | | | | | | | | | | |
| MIDDLE:SECOND.USER | -0.68* | -1.03 | -1.39 | -0.97 | -1.19 | -1.31 | -0.74 | -1.44* | -0.52* | -1.54 | -0.87 |
| FAR:SECOND.USER | -1.02 | -2.25 | -2.12 | -1.72 | -2.22 | -2.73 | -1.52 | -2.41 | -1.52 | -3.61 | -1.62 |
| MIDDLE:NOT.USER | 0.95 | 1.68 | 1.52 | 1.64 | 1.27 | 1.20 | 0.63 | 2.47 | 1.66 | 1.53 | 0.89 |
| FAR:NOT.USER | -1.19 | -1.19* | -1.55* | -2.01 | -1.51* | -2.31 | -1.27 | -2.68 | -1.93 | -1.94 | -1.15 |
| DAY TYPE:NET CARD | | | | | | | | | | | |
| WEEKEND:NO | n. r. | -0.91 | -1.23 | -1.23 | -1.07 | -0.82* | -0.50* | -1.51 | -0.83 | -1.25 | -0.45 |
| USAGE:DAY TIME | | | | | | | | | | | |
| SECOND.USER:NIGHT | 1.32 | 5.01 | 5.22 | 6.63 | 5.71 | 5.07 | 3.53 | 6.17 | 4.53 | 5.67 | 4.96 |
| NOT.USER:NIGHT | -0.06* | 0.31* | 0.45* | 0.38* | 0.26* | 0.32* | 0.56* | 0.72* | 0.69* | 0.68* | 0.59* |
| SEX:DAY TIME | | | | | | | | | | | |
| FEMALE:NIGHT | 1.70 | 2.88 | 3.36 | 3.55 | 3.49 | 3.02 | 1.11 | 2.94 | 1.65 | 3.30 | 1.22 |
| PERSONAL INCOME: | | | | | | | | | | | |
| NET CARD | | | | | | | | | | | |
| MIDDLE:NO | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. |
| HIGH: NO | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. |
| DISTANCE:AGE | | | | | | | | | | | |
| MIDDLE:POLY.AGE.1 | -12.93 | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. |
| FAR:POLY.AGE.1 | -0.09* | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. |
| MIDDLE:POLY.AGE.2 | -2.41* | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. |
| FAR:POLY.AGE.2 | 0.76* | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. |
| DAY TYPE:AGE | | | | | | | | | | | |
| WEEKEND:POLY.AGE.1 | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. |
| WEEKEND:POLY.AGE.2 | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. | n. r. |

Table 6.5: Posterior Mode Estimates for Main Effect and Interaction Parameters (* = 90% credible interval does not include 0, n.r.= effect was not required in model, since model with effect has a 90% credible interval which includes 0)

6.3 Model Comparison

A general method for model comparison in Bayesian models estimated by MCMC is the DIC criterion suggested by Spiegelhalter et al. (2002). This method is developed for exponential family models and based on the deviance. Even though the binary probit and logit models belong to this class of models, Collett (2002) has shown that the deviance in binary data problems should not be used for model assessment. Our application is based on binary responses. Since they cannot be grouped to binomial responses with sufficient large numbers of trials because of the complexity of the fixed, spatial and cluster effects, we decided not to use the DIC criterion for this application. We like to note that it is in general possible to determine DIC values of the models discussed in Section 3 and 4 as long as the binary regression data can be grouped to binomial regression data with sufficiently large number of trials.

To facilitate model comparison for this application we propose to use the sum of weighted squared residuals. Since in this application it is of special interest to compare the fit for all 11 investigated models with regard to their spatial effects, we chose as measure D_w the sum of weighted squared residuals over all postal codes of Munich defined by

$$D_w(\mathbf{Y}) := \sum_{j=1}^{74} n_j (p_j^{empir} - p_j^{estim})^2, \quad (6.1)$$

where $n_j :=$ number of trips in the j^{th} postal code. Empirical probabilities p_j^{empir} are equal to observed proportion of trips using individual transport in postal code area j , and the posterior probability estimates p_j^{estim} are based on the MCMC run, and defined as:

$$p_j^{estim} := \frac{1}{n_j * R} \sum_{i: j(i)=j} \sum_{r=1}^R h^{-1}(\eta_{ir}), \quad (6.2)$$

where $h(\cdot)$ is logit (for Models 1 - 5, 6, 8, 10) or probit (for Models 7, 9, 11) link function, and

$$\eta_{ir} = \begin{cases} \mathbf{x}_i^\dagger \boldsymbol{\alpha}_r + b_{j(i),r} & \text{for Model 1} \\ \mathbf{x}_i^\dagger \boldsymbol{\alpha}_r + b_{j(i),r} + c_{m(i),r} & \text{for Models 2-5} \\ \frac{\mathbf{x}_i^\dagger \boldsymbol{\alpha}'_r + b'_{j(i),r}}{\sigma'_{m(i),r}} & \text{for Models 6-11.} \end{cases}$$

In Table 6.6 we present value D_w for all 11 models and the number of parameters required in calculating

| Model | 1 spatial only | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|--------------------------------------|-------------------|-------------------------|------|------|------|------------------------------|-----|------|------|------|------|
| | | spatial + group cluster | | | | spatial + individual cluster | | | | | |
| fixed effects | 31 | 28 | 28 | 28 | 28 | 26 | 26 | 28 | 28 | 28 | 28 |
| spatial effects | 74 | 74 | 74 | 74 | 74 | 74 | 74 | 74 | 74 | 74 | 74 |
| cluster effects | 0 | 74 | 5 | 12 | 5 | 2 | 2 | 4 | 4 | 4 | 4 |
| total number of parameters for D_w | 105 | 176 | 107 | 114 | 107 | 102 | 102 | 106 | 106 | 106 | 106 |
| \mathbf{D}_w | 2.35 | 1.23 | 0.95 | 1.02 | 1.44 | 1.9 | 1.7 | 3.25 | 2.88 | 1.84 | 1.72 |

Table 6.6: Model Fit Comparison with Regard to Spatial Probabilities

D_w . The total number of parameters required for D_w will be used as a rough measure for the complexity of the model with regard to the spatial fit. This means we regard these parameters as model parameters and the spatial dependence parameter, spatial variance and the cluster variance parameters in group cluster models as hyperparameters belonging to the prior. This approach is consistent with the approach taken in Spiegelhalter et al. (2002), which point out in their discussion that complexity depends on the focus of the analysis. We already pointed out the difficulties in using the DIC criterion in binary regression models. But now we want to add that in setting our focus on assessing the spatial fit, the corresponding



Figure 6.5: Top right map: Observed Probabilities of Individual Transport Use by Postal Codes in Munich, Germany; Top left map: Posterior Mean Probability Estimates of Individual Transport Use by Postal Codes in Munich, Germany for Model 3; Bottom maps: Estimated Spatial Effects b_j , $j = 1, \dots, 74$ in Model 3.

calculations of the complexity measure p_D suggested by Spiegelhalter et al. (2002) cannot be facilitated since the corresponding deviances are not available in closed form as pointed out by S.P. Brooks in the discussion of the paper of Spiegelhalter et al. (2002). According to Table 6.6 the best fit with regard to spatial probabilities has Model 3 (with group cluster effects). We see that even though the models with individual cluster effects have a lower model complexity with regard to spatial fit, their goodness of fit as measured by D_w is worse than Model 3. Model 4 has a comparable D_w value to Model 3 but the model complexity is higher, therefore we prefer Model 3.

For Model 3 we present a color map with estimated spatial probabilities over postal codes of Munich (see Figure 6.5, top right map), which coincides quite well with the map showing the empirical spatial probabilities (for comparison see Figure 6.5, top left map). This indicates graphically that Model 3 has a reasonably good fit of the data with respect to the spatial resolution. Recall that we presented in Figure 6.1 the map with the number of available trips over postal codes of Munich.

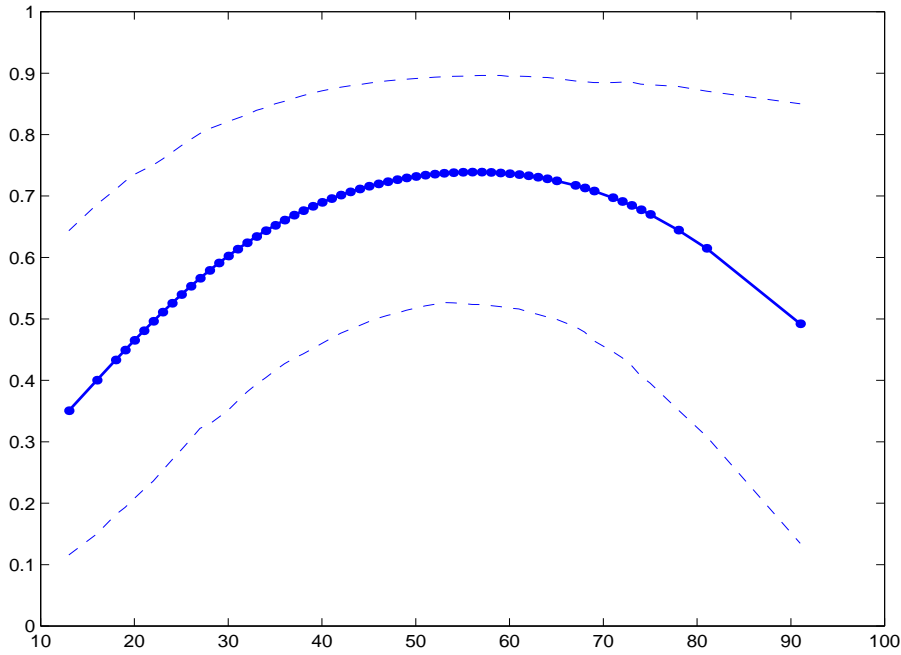


Figure 6.6: Estimated posterior mean probabilities for using an individual transport in Postal code area 81377 and 5th cluster group for different **AGE**, while other covariates are set as in Table 6.1 (dotted lines correspond to 90% credible bounds)

6.4 Model Interpretation

After model fitting and model selection one is interested in what can be learned about the transport behavior. For this we now investigate the implications of Model 3. First we will estimate these probabilities when individual or combinations of two covariates change. The remaining covariates in the model are set to their “most usual values”, corresponding to the modus for categorical covariates and median values for quantitative covariates. These “most usual values” are presented in Table 6.1. The only quantitative covariate **AGE** has a median age of 42 years. Since Model 3 includes spatial effects we have to specify a postal code for which we estimate these probabilities. We have chosen postal code area 81377, since this postal code area has a large observed number of trips and the smallest 90% credible interval for its spatial effect. Finally Model 3 contains group cluster effects with regard to the number of trips a household has taken. Since each cluster group contains the similar number of individual trips, for our investigations we chose the last, i.e. the 5th cluster group corresponding to households with ≤ 7 trips (see Table 6.2), which has the smallest 90% credible interval for its cluster effect c_5 . Posterior mean estimates for individual transport probabilities for a fixed covariate vector \mathbf{x} in Postal code 81377 (corresponding to b_{47} and the 5th group cluster can be calculated as

$$p_{mean}(\mathbf{x}) := \frac{1}{R-B} \sum_{r=B+1}^R p_r(\mathbf{x}) = \frac{1}{R-B} \sum_{r=B+1}^R \left(\frac{\exp(\mathbf{x}^t \hat{\alpha}_r + \hat{b}_{47,r} + \hat{c}_{5,r})}{1 + \exp(\mathbf{x}^t \hat{\alpha}_r + \hat{b}_{47,r} + \hat{c}_{5,r})} \right), \quad (6.3)$$

where $\hat{\alpha}_r$, $\hat{b}_{47,r}$ and $\hat{c}_{5,r}$ are the r^{th} MCMC estimate of α , b_{47} and c_5 respectively. Here R is the total number of MCMC iterations and B is the burn-in. We can also determine 90% credible bounds which are defined as 5% and 95% quantiles of the sample $\{p_r(\mathbf{x})\}_{r=B+1}^R$. For “the most usual” trip, which is associated with postal code 81377 and 5th cluster group, the estimated posterior mean probability for taking individual transport is equal to 0.7.

Figure 6.6 gives the estimated posterior mean probability together with 90% credible bounds for choosing individual transport as age changes in Postal code area 81377 and trips associated with the 5th

group cluster when the remaining covariates are set to their “most usual value” given in Table 6.1. It is not very surprising that the probability of using a car increases rapidly to an age of about 35 years, remains reasonably stable between 35 years and 65 years and decreases slowly after the 65 years. Younger people have a lower probability to own a car, while older people might prefer public transport options.

We can interpret the effect of age directly, since no interaction terms include age. For almost all other covariate effects we have to consider covariate combinations corresponding to interaction terms. In particular note that Model 3 includes 7 interaction terms. In order to interpret effects of the categorical covariates we plot for each of the 7 interactions the estimated posterior mean probabilities (6.3) for using individual transport. For brevity reasons we interpret only 2 of the 7 interaction plots. From top left panel of Figure 6.7 we see that net card users prefer public transport for trips taken alone much more often than when the trip is taken with others. This is to be expected since a net card can only be used in general by a single person. In contrast users without a net card take individual transport options much more often regardless if the trip taken alone or not.

The third left panel shows an interesting behavioral difference between females and males. During the day there is a little difference. However during the night women nearly always use individual transport options, while males choose this option only half as often. An explanation might be that women are afraid to use public transport at night because of low usage and deserted stops.

This shows that some expected behavioral patterns can be captured when interactions are allowed in the model. The remaining panels of Figure 6.7 are interpreted in detail in Section 6.4 of Prokopenko (2004).

We continue now this subsection with the interpretation of spatial effects. There are 24 postal codes whose 90% credible intervals do not include zero and therefore are significant (see bottom right panel of Figure 6.5). We expect that the interpretation of the spatial effects is related to the structure of the subway (U-Bahn) net and suburban railway (S-Bahn) net. Table 6.7 confirms our assumption in general. The left column shows the numbers of postal code areas, which have U- or S-stops inside. The right

| | with U- or S-stops | without U- or S-stops inside PLZ |
|-----------------------|---------------------|--|
| 90% <i>CI</i> over 0 | 2 (80333, 81476) | 5 |
| 90% <i>CI</i> below 0 | 11 | 6 (80999, 80634, 80797, 81243, 80689, 81373) |

Table 6.7: Interpretation of spatial effects in context of presence/absence of the U-or S- stops inside of postal codes; the postal code numbers of 8 untypical postal code areas are given in parentheses.

column contains the numbers of postal code areas without stops. The estimated odds ratio of Table 6.7 is $\frac{2.6}{11.5} \approx 0.22$, which is clearly below 1 (the 90% confidence interval is $[0.044, 1.091]$). This confirms that presence of U- and S-stops are related to significant spatial effects. While there is a general relationship between significant spatial effects and the presence of the U+S-net in these postal areas, 8 areas do not follow this pattern (see Table 6.7). These areas should therefore be of special interest to the city planners, which seek to improve the public transport net, since these areas indicate areas of low/high public transport usage even after adjustment of trip, person and household specific effects.

We already noted that the estimate of the spatial dependency parameter $\hat{\psi} \approx -0.5$ is negative. This can be explained by the specific structure of S- and U-Bahn net of Munich, whose lines run from the center to suburbs like a star. Since the sign of the spatial effects correlates with the presence/absence of the U-or S- stops, it is not surprising, that especially far from the center the neighboring postal codes have often spatial effects with opposite signs.

Finally we mention that cluster effects for households with large numbers of trips are positive and cluster effects for households with few numbers of trips are negative (see Table 6.2 and Figure 6.4). This implies that households with high mobility needs use a car more often than households with low mobility needs.

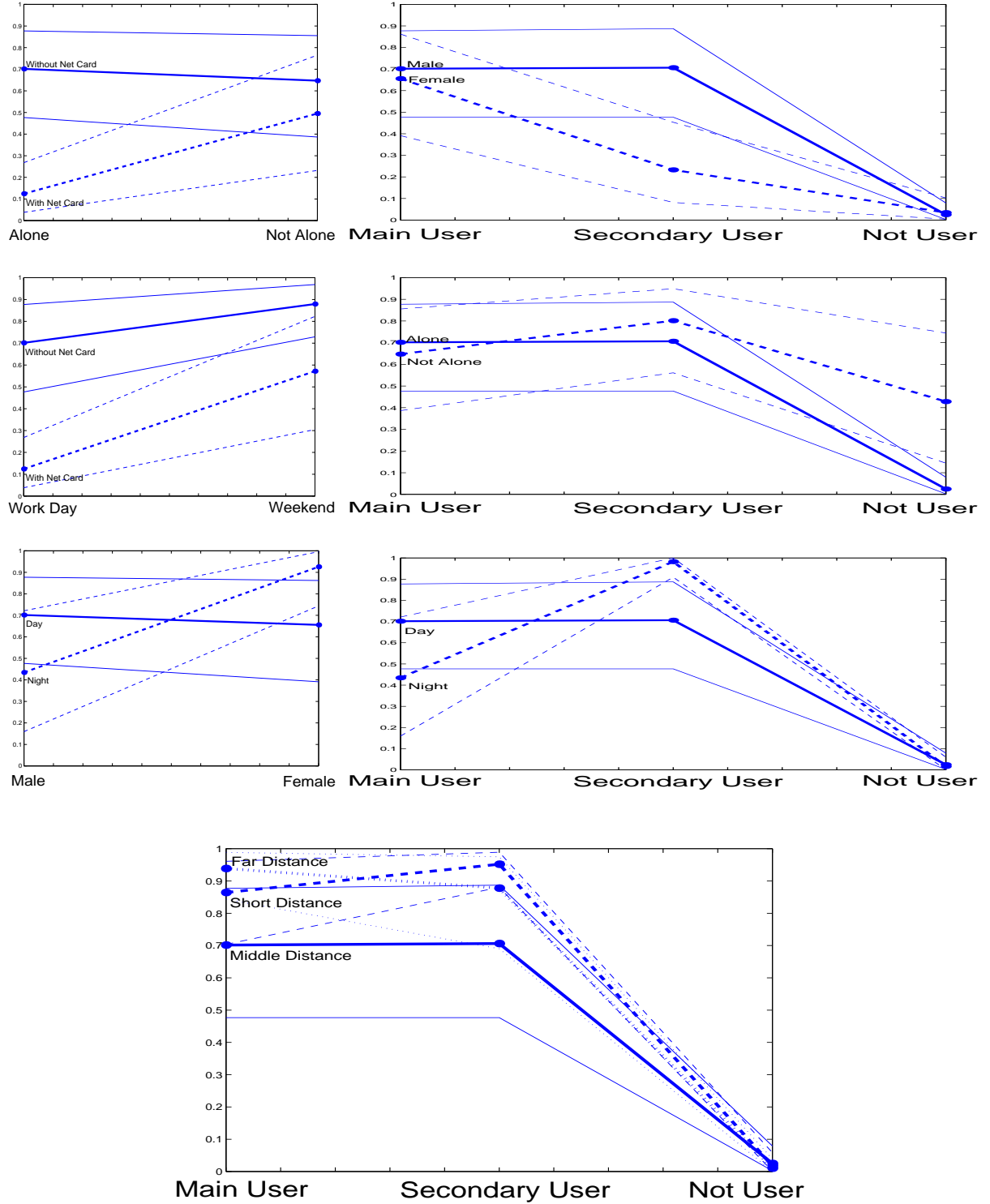


Figure 6.7: Estimated posterior mean probabilities for using an individual transport in Postal code area 81377 and 5th cluster group for different combinations of the covariates which form the interaction, while other covariates are set to the “most usual value” given in Table 6.1. Dotted lines correspond to 90% credible bounds.

Left column: WAY ALONE:NET CARD, DAY TYPE:NET CARD, SEX:DAY TIME.

Right column: USAGE:SEX, USAGE:WAY ALONE, USAGE:DAY.TIME.

Bottom: USAGE:DISTANCE.

7 Discussion: Summary and Discussion

An extended version of the spatial Gaussian CAR model proposed by Pettitt et al. (2002) has been presented, which allows for spatial independence and the intrinsic CAR model as special cases. In addition, this model possesses a proper joint distribution and allows for a fast update of the spatial dependence parameter. As noted before this modification has a more reasonable behavior of the conditional variance of a spatial effect given all other spatial effects than the model considered in Sun et al. (2000).

In a hierarchical setup this extended CAR model has been used for binary spatial regression data. To capture additional heterogeneity, cluster effects have been included. In addition to the conventional modeling of heterogeneity between groups (group cluster effects) through independent random effects, modeling of heterogeneity within groups (individual cluster effects) has also been considered. A naive approach for this yields an unidentifiable model. It is shown how the model can be reparametrized to overcome nonidentifiability. Parameter estimation is facilitated by an MCMC approach. centrality estimates and their credible intervals have asymptotically Separate MCMC algorithms have been developed for the two hierarchical model classes considered: binary regression with spatial and group cluster effects and binary regression with spatial and individual cluster effects. For the models with individual cluster effects, a logit and probit formulation can be used. For a probit link we developed a MCMC algorithm by using a latent variable representation requiring only a single MH updates. This is faster because of better mixing behavior than a corresponding MCMC algorithm based on the logit formulation. All MCMC algorithms presented in this paper are validated through simulation.

The usefulness of these models has been demonstrated by the application to the mobility study which motivated the work. We show that this approach is able to detect spatial regions where public transport options are more/less often used after adjusting for explanatory factors.

For model comparison, we use the sum of weighted squared residuals as a measure of fit and the number of parameters required for estimating spatial probabilities as a rough measure of model complexity. A more theoretical based approach is still needed and of current research interest. Alternatives such as posterior predictive p-values proposed by Gelman et al. (1996) are possible.

Another line of further research is to consider the problem of including interactions between cluster and spatial effects. For this we suggest to use multivariate CAR models mentioned for example by Pettitt et al. (2002). The multivariate CAR model is a model for $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_J)^t$, where the components $\mathbf{b}_j = (b_{j1}, \dots, b_{jM})^t$, $j = 1, \dots, J$ are M -dimensional vectors instead of scalars, as before. The joint distribution of the vector \mathbf{b} is defined as follows:

$$\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_J)^t \sim N_{J \times M}(\mathbf{0}, \tau^2(Q^{-1} \otimes V)), \quad V = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \vdots \\ \vdots & & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix} \in \mathbb{R}^{M \times M}, \quad (7.1)$$

where $A \otimes B$ stands for *Kronecker product* of matrices A and B . In particular for the multivariate modified Pettitt CAR, the conditional distribution is given then as follows (compare with (2.5)):

$$\mathbf{b}_j | \mathbf{b}_{j'}, j \neq j' \sim N_M \left(\frac{\phi}{1 + |\phi|N_j} \sum_{j \sim j'} \mathbf{b}_{j'}, \frac{(1 + |\phi|)\tau^2}{1 + |\phi|N_j} \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \vdots \\ \vdots & & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix} \right).$$

The parameter ρ measures the strength of the cluster dependence. If $\rho = 0$ then all M components of vector \mathbf{b}_j are iid. As before, the parameter ϕ measures the strength of the spatial dependence. If $\phi = 0$ then the vectors $\mathbf{b}_j, j = 1, \dots, J$ are independent and normally distributed with mean zero and covariance matrix $\tau^2 V$. Properties of the multivariate CAR model are studied in Pettitt et al. (2002). The authors use multivariate Gaussian CAR models for a data augmentation approach. In their application the binary data concerns the presence or absence of two tree varieties represented at 469 sites. Since the presence of these two kinds of trees can depend on each other, the authors model this data using a

multivariate CAR model with $J = 469$ and $M = 2$ by a threshold mechanism. Gelfand and Vounatsou (2003) use multivariate extensions of (Sun et al. 2000) proper CAR models for multivariate continuous and multinomial spatial data. Further, Carlin and Banerjee (2002) use multivariate Gaussian CAR models in generalized linear mixed models, namely for spatial survival data analysis, where the Gaussian variable enters in the linear predictor. We now propose to apply the multivariate Gaussian CARs in a new way, namely for modeling spatial-cluster interactions. More precisely, we propose to model spatial and cluster effects jointly as some multivariate CAR. As usually, J denotes the number of regions, while M stands for the number of clusters. Then logits are modeled as follows (compare with (3.2)):

$$\theta_i := \log \left(\frac{p_i}{1 - p_i} \right) = \underbrace{\mathbf{x}_i^t \boldsymbol{\alpha}}_{\text{fixed effect}} + \underbrace{b_{j(i), m(i)}}_{\text{spatial and cluster effect}},$$

where $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_J)^t$, $\mathbf{b}_j = (b_{j1}, \dots, b_{jM})^t$, $j = 1, \dots, J$ is modeled as a realization of the multivariate CAR (7.1). In this model we have to estimate one additional parameter ρ , which measures strength of a space-cluster interaction. The absence of interaction is indicated by $\rho = 0$. In this case the M vectors $(b_{1m}, \dots, b_{Jm})^t$, $m = 1, \dots, M$ are independent identically distributed Gaussian CAR models.

Finally we show how to interpret the modeled interaction present in the multivariate CAR model (7.1) as a product of spatial and cluster effects. By this we mean that the distribution of the multivariate Gaussian CAR vector \mathbf{b} with the variance-covariance matrix $\tau^2(Q^{-1} \otimes V)$ has the same mean and covariance matrix as the random vector $B := (B_{11}, \dots, B_{JM})^t$ with components

$$B_{jm} = B_j \cdot A_m,$$

where $B_j, j = 1, \dots, J$ and $A_m, m = 1, \dots, M$ are independent random vectors. Here

$$(B_1, \dots, B_J)^t \sim N_J(\mathbf{0}, \tau^2 Q^{-1}) \quad \text{is a Gaussian CAR and} \\ \mathbf{A} := (A_1, \dots, A_M)^t \quad \text{has zero mean and covariance } V.$$

If in addition for the vector \mathbf{A} the following distribution is chosen:

$$\begin{aligned} P(A_m = 1) &= P(A_m = -1) = \frac{1}{2}, \\ P(A_{m'} = 1 | A_m = 1) &= P(A_{m'} = -1 | A_m = -1) = \frac{1+\rho}{2}, \\ P(A_{m'} = 1 | A_m = -1) &= P(A_{m'} = -1 | A_m = 1) = \frac{1-\rho}{2}, \end{aligned}$$

then the components B_{jm} , $j = 1, \dots, J$, $m = 1, \dots, M$ have the same distribution as the corresponding spatial-cluster effects b_{jm} , i.e. they are also normal with the same mean and variance. However their joint distributions are different.

Another topic of research would be to consider the modeling of simultaneous heterogeneity within and between clusters, which would combine group and individual cluster approaches. In particular we would assume

$$c_{mk} \sim N(c_m, \sigma_m^2), \quad c_m \sim N(0, \sigma_c^2) \quad \text{for } m = 1, \dots, M, \quad k = 1, \dots, K_m.$$

(compare with 3.2 and 4.1). Here a similar non identifiability problem has to be solved.

References

- Albert, J. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *J. Am. Statist. Ass.* 88, 669–679.
- Bennett, J., A. Racine-Poon, and J. Wakefield (1996). MCMC for nonlinear hierarchical models. In *Markov Chain Monte Carlo in Practice*, pp. 339–57. Chapman & Hall.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc. B*, 36, 192–236.
- Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician* 24, 179–195.

- Besag, J. and P. Green (1993). Spatial statistics and Bayesian computation (with discussion). *J. R. Statist. Soc. B*, 55, 25–37.
- Besag, J., P. Green, D. Higdon, and K. Mengersen (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science* 10, 3–66.
- Breslow, N. and D. Clayton (1993). Approximate inference in generalized linear mixed models. *J. Am. Statist. Ass.* 88, 9–25.
- Carlin, B. and S. Banerjee (2002). Hierarchical multivariate CAR models for spatio-temporally correlated survival data. *Bayesian Statistics 7, to appear*, 7–21.
- Chib, S. (2001). Markov Chain Monte Carlo methods: Computation and inference. In *Handbook of Econometrics, Volume 5*, pp. 381–394. North Holland.
- Collett, D. (2002). *Modelling Binary Data, second Edition*. London: Chapman & Hall.
- Diggle, P., J. Tawn, and R. Moyeed (1998). Model-based geostatistics (with discussion). *Applied Statistics* 47, Part 3, 299–350.
- Fahrmeir, L. and S. Lang (2000). Bayesian semiparametric regression analysis of multicategorical time-space data. <http://www.stat.uni-muenchen.de/sfb386/>, *Discussion Paper 202*.
- Gelfand, A., N. Ravishanker, and M. Ecker (2000). Modeling and inference for point-referenced binary spatial data. In *Generalized Linear Models: A Bayesian Perspective*, pp. 381–394. Marcel Dekker.
- Gelfand, A. E. and P. Vounatsou (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* 4, 11–25.
- Gelman, A., X. Meng, and H. Stern (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 6, 733–807.
- George, A. and J. Liu (1981). *Computer Solution of Large Sparse Positive Definite Systems*. Prentice-Hall.
- Gilks, W., S. Richardson, and D. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. New York: Chapman & Hall.
- Heagerty, J. and S. Lele (1998). A composite likelihood approach to binary spatial data. *J. Am. Statist. Ass.* 93, 1099–1111.
- Huffer, F. and H. Wu (1998). Markov Chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics* 54, 509–524.
- Kelsall, J. and P. Diggle (1998). Spatial variation in risk of disease: a nonparametric binary regression approach. *Appl. Statist.* 47, Part 4, 559–573.
- Knorr-Held, L. and H. Rue (2000). On block updating in Markov random field models for disease mapping. (revised, may 2001). <http://www.stat.uni-muenchen.de/sfb386/>, *Discussion Paper 210*.
- Pettitt, A., I. Weir, and A. Hart (2002). Conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. *Statistics and Computing* 12, 353–367.
- Prokopenko, S. (2004). *Hierarchical binary spatial regression models with cluster effects, Ph.D. thesis*. Munich University of Technology: <http://tumb1.biblio.tu-muenchen.de/publ/diss/ma/2004/prokopenko.html>.
- Robert, C. (1995). Simulation of truncated normal variables. *Statistics and Computing* 5, 121–125.
- Spiegelhalter, D., N. Best, B. Carlin, and A. Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society* 64, 583–639(57).
- Sun, D., K. T. Robert, H. Kim, and Z. He (2000). Spatio-temporal interaction with disease mapping. *Statistics in Medicine* 19, 2015–2035.
- Zängler, T. (2000). *Mikroanalyse des Mobilitätsverhaltens in Alltag und Freizeit*. Berlin: Springer Verlag.