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

Claudia Czado Sergij Prokopenko *

October 26, 2007

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. For the spatial effects a modification of a class of Markov Random Fields (MRF) models with proper joint distributions introduced by Pettitt et al. (2002) is developed. It contains the intrinsic MRF in the limit and allows for efficient Markov Chain Monte Carlo (MCMC) algorithms. Further cluster effects using group and individual approaches are taken into consideration. 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 a re-parametrization gives identifiable parameters. This provides a new approach for modeling heterogeneity within clusters. Finally the proposed model classes are applied to the mobility study.

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

[1]Claudia Czado is Associate Professor at the Center for Mathematical Sciences, Munich University of Technology, Boltzmannstr. 3, D-85747 Garching, Germany (E-mail: cczado@ma.tum.de) and Dr. Sergiy Prokopenko is Manager at AC Nielsen Germany. The 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. Discrete choice models based on maximization of random utilities (McFadden 1984) have often been used in investigating such travel mode decisions (Ben-Akiva and Lerman 1985, McFadden 2001 and Bhat 2006) giving rise to the familiar multinomial logit model. For a binary choice this gives a logit model. Since the variable of interest is a binary indicator, whether public transport has been utilized or not, we base our models on a logit formulation. Early papers (McGillivray 1972 and McFadden 1974) on travel mode decisions also consider binary choices, but they do not include spatial components. The central question is to identify areas of low/high utilization of public transport after adjusting for trip, individual and household attributes. A spatial model is utilized to find such areas. In addition extra heterogeneity is to be expected since trips are taken by the same individual and same household, which will be accounted for by cluster effects. The contribution of this paper is twofold, first we develop such flexible spatial models, provide efficient estimation and model selection methods and secondly we show that these models allow for a comprehensive analysis of the German mobility study.

In our application spatial information is aggregated over postal code areas. Therefore we model spatial effects for each region. Here linear predictors are given as $\eta_i = \mathbf{x}_i'\boldsymbol{\alpha} + b_{j(i)}, i = 1, \ldots, n, j = 1, \ldots, J$, where J denotes the number of regions, $\mathbf{b} = (b_1, \cdots, b_J)'$ are spatial effects and $j(i)$ indicates the region associated with the i^{th} observation. Spatial effects are modeled as a realization from a Gaussian *Markov random field* (MRF) (Besag and Green 1993, Banerjee et al. 2004 and Rue and Held 2005). The notation *Gaussian conditional auto-regression* (Gaussian CAR) is also used. This allows fast individual updating of $J << n$ spatial effects in a Gibbs sampler, but requires a spatial neighborhood structure. Postal codes are neighbors if they have a joint border.

In contrast to kriging, in Gaussian CAR models the explicit form of the spatial effect's precision matrix (inverse covariance matrix) is available. Moreover this precision matrix is usually sparse, which allows one to compute its determinant much faster than in the kriging approach. Pettitt et al. (2002) use this fact and propose a specific dependence structure which provides even an analytical computation of its determinant. 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 (2001) used improper intrinsic CAR models as a prior in a semi parametric regression model for multi categorical space-time data, while Knorr-Held and Rue (2002) applied intrinsic CAR priors for Poisson models used in disease mapping. 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 the intrinsic CAR model. This modification still has all nice properties of the Pettitt's CAR models: proper joint distributions, a similar interpretation of parameters, the same conditional correlations and more importantly allows for fast computation of the determinant of the precision matrix. An alternative proper Gaussian CAR model was also discussed in Sun et al. (2000). It also includes the intrinsic CAR model in the limit 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. Hierarchical multivariate CAR models have been considered in Jin et al. (2005). Gaussian CAR models will be considered in more detail in Section 2.

Another approach to regionally aggregated data is based on specifying the joint distribution of the spatial effects directly yielding simultaneous autoregressive (SAR) Gaussian models as introduced by Whittle (1954) and later studied by Cressie (1993) and Anselin (1988). Especially economists prefer the simultaneous approach for the analysis of spatial regional data (Anselin and Florax 1995 and Anselin, Florax, and Rey 2004). Pinkse and Slade (1998) and McMillen (1992) consider a probit formulation with latent spatial regression following a SAR specification. Different estimation methods for these models are provided by Fleming (2004). Wall (2004) points out difficulties in interpreting the spatial dependence parameter for CAR and SAR models as spatial correlation parameters in non-regular lattices. We agree with Cressie (1993) and Rue and Held (2005) that CAR models are easier to interpret and do not consider SAR models here.

Finally, we mention auto logistic regression models (Besag 1974) with covariate effects considered by Huffer and Wu (1998). However exact MLE's are not tractable in these models for larger data sets, so Huffer and Wu (1998) investigate a MCMC MLE approach.

In addition to spatial effects we extend our modeling of the linear predictor η_i by cluster random effects. This accounts for possible over-dispersion caused by unobserved heterogeneity as to be expected in our mobility application. We consider two approaches, namely group and individual cluster effects. The first one, which models heterogeneity between clusters, follows the usual idea of having the same random effect within a cluster. 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. For K clusters we have to estimate K cluster specific variances instead of K cluster effects as before. We will show how an identifiability problem occurring in the second case can be overcome. Efficient MCMC algorithms will be developed. In this paper we restrict our analysis to logit models with spatial and cluster effects. However Prokopenko (2004) also develops MCMC algorithms for probit formulations using a latent variable representation (Albert and Chib 1993).

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 CAR models (Pettitt et al. 2002), where a random vector $\mathbf{b} \in R^J$ is defined through its full conditionals as follows:

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

Here $\mathbf{b_{-j}} = (b_1, \ldots, b_{j-1}, b_{j+1}, \ldots, b_J)^t$ 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 $\mathbf{b} \sim N_J\left(\mathbf{0},\, (\mathbf{I_J} - \mathbf{C})^{-1}\mathbf{M}\right)$, where $\mathbf{C} = (c_{jj'})$ with $c_{jj} = 0$, $j = 1, \ldots, J$, and $\mathbf{M} = \mathrm{diag}(\kappa_\mathbf{1}, \ldots, \kappa_\mathbf{J})$. The J-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $N_J(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The precision matrix is given by $\mathbf{Q} = \mathbf{M^{-1}}(\mathbf{I_J} - \mathbf{C})$. We assume that the neighborhood structure has no isolated regions or groups of regions.

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

$$b_j|\mathbf{b_{-j}} \sim N(\bar{b}_j, \frac{\tau^2}{N_j})\,, \quad j = 1, \ldots, J, \quad \text{and} \quad \bar{b}_j = \frac{\sum_{j' \sim j} b_{j'}}{N_j}\,, \tag{2.1}$$

where $N_j =$ number of neighbors of the region j, and "$j' \sim j$" denotes contiguous regions. In particular, we have $j \not\sim j$. The corresponding precision matrix is positive semi-definite with rank $= J-1$, therefore $\mathbf{b}$ has an improper density, but can be characterized by a flat improper distribution for b_J and a proper multivariate normal distribution for the remaining components given b_J (see Prokopenko 2004).

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

$$b_j|\mathbf{b_{-j}} \sim N\left(\frac{\phi}{1+|\phi|N_j}\sum_{j'\sim j} b_{j'}\,,\,\frac{\tau^2}{1+|\phi|N_j}\right). \tag{2.2}$$

The parameter ϕ measures the strength of the spatial dependency. There is no spatial dependency, if $\phi = 0$. This corresponds to unstructured random effects. Therefore it is possible to include two sets of random effects, one with $\phi = 0$ (unstructured) and one with $\phi > 0$ (structured) as is often done in disease mapping. For the data considered the separation into structured and nonstructural effects yields nonsignificant effects on both levels. Since ML estimation is intractable for this model MCMC methods have been used to estimate ϕ and τ^2. Pettitt et al. (2002) show that a fast and simple update of ϕ for a Gibbs Step given the vector $\mathbf{b}$ and τ^2 is available. In contrast to the intrinsic CAR, the joint distribution of $\mathbf{b}$ based on conditionals specified in (2.2) is a proper distribution, which leads to a proper posterior when used as a prior distribution.

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

$$b_j|\mathbf{b_{-j}} \sim N\left(\frac{\phi}{1+|\phi|N_j}\sum_{j\sim j'} b_{j'}\,,\,\frac{(1+|\phi|)\tau^2}{1+|\phi|N_j}\right). \tag{2.3}$$

Using for example Besag (1974) or Rue and Held (2005) it is easy to see that the full conditional distributions in (2.3) yield a proper joint distribution. This modification allows us to have the intrinsic CAR (2.1) in the limit, when $\phi \to \infty$. Note that the conditional variance of $b_j|\mathbf{b_{-j}}$ decreases to τ^2/N_j as $|\phi|$ increases to infinity, while in the original model (2.2) this quantity decreases to zero, which is restrictive. This model has the same behavior as Pettitt's CAR (2.2) when ϕ goes to zero (no spatial dependency), and all partial correlations between b_j and b_i given all the other sites are the same. Further the modified Pettitt's CAR model has larger conditional and marginal variance for b_j than the original Pettitt's model for $\phi > 0$, thus allowing for a larger variability. In the modified Pettitt's model we can also achieve a simple update for ϕ. We write now $\tau^{-2} \times \mathbf{Q}^{m.P}(\phi)$ for the precision matrix of the Model (2.3). In particular $\mathbf{Q}^{m.P}(\phi) = \mathbf{M}^{-1}(\phi)(\mathbf{I}_J - \mathbf{C}(\phi))$, where $\mathbf{M}(\phi) = diag(\frac{(1+|\phi|)}{1+|\phi|N_1}, \cdots, \frac{(1+|\phi|)}{1+|\phi|N_J})$ and $\mathbf{C}(\phi) = (c_{jj'}(\phi))_{jj'=1,\cdots,J}$ with

$$c_{jj'}(\phi) = \begin{cases} \frac{\phi}{1+|\phi|N_j}, & \text{if } j \sim j' \\ 0, & \text{if } j \not\sim j',\ j = j' \end{cases}.$$

Each update of ϕ needs the determinant of $\mathbf{Q}^{m.P}(\phi)$. With $\psi = \frac{\phi}{1+|\phi|}$ we follow a similar approach as in Pettitt et al. (2002). Define the diagonal matrix

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

then $\mathbf{Q}^{m.P}(\phi)$ can be written in the form $\mathbf{Q}^{m.P}(\psi) = I_J + |\psi|\mathbf{D} - \psi\mathbf{\Gamma}$. If $(\lambda_1, \ldots, \lambda_J)$ are the eigenvalues of $\mathbf{\Gamma} - \mathbf{D}$ and $(\nu_1, \ldots, \nu_J)$ are the eigenvalues of $\mathbf{\Gamma} + \mathbf{D}$, then the determinant of

$\mathbf{Q}^{m.P}(\psi)$ is equal to

$$|\mathbf{Q}^{m.P}(\psi)| = \begin{cases} \prod_j (1-\psi\lambda_j), & \text{if } \psi > 0 \\ 1, & \text{if } \psi = 0 \\ \prod_j (1-\psi\nu_j), & \text{if } \psi < 0 \end{cases} \tag{2.4}$$

and can be computed quickly for any value of ψ. We note that the conditional variance of $b_j|\mathbf{b_{-j}}$ 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 (2.3). 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 Model (2.3). Therefore we prefer the proper Model (2.3) over the proper CAR model studied by Sun et al. (2000).

3 Spatial Logistic Regression with Group Cluster Effects

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

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

where Y_i is Bernoulli with the success probabilities p_i and assume that Y_i given p_i is independent for $i = 1, \ldots, n$. We specify p_i through their logits :

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

Here the design vector $\mathbf{x_i}$ multiplied with the regression parameter vector $\boldsymbol{\alpha} \in \mathbb{R}^p$ represents fixed effects. With the vector $\mathbf{b} = (b_1, \ldots, b_J)$ we allow for 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 (2.3). To model heterogeneity between clusters we allow for random cluster effects represented by the vector $\mathbf{c} = (c_1, \ldots, c_M)$. Each of the M clusters (say household types) induces a group specific random effect, which we denote by $c_m, m = 1, \ldots, M$, respectively. The index $m(i)$ denotes the cluster of trip i. We assume that $c_m \sim N(0, \sigma_c^2)$ *i.i.d.* for $m = 1, \ldots, M$.

Note that the likelihood of the response vector $\mathbf{Y}$ is proportional to

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

Since we will follow a Bayesian approach we need to complete the model specification by providing the prior specifications for $\boldsymbol{\alpha}, \phi, \tau^2$ and σ_c^2. We denote the density of a random variable X by $[X]$ and the conditional density of X given Y by $[X|Y]$, respectively. We assume independent prior distributions, i.e. $[\boldsymbol{\alpha}, \mathbf{b}, \mathbf{c}, \phi, \tau^2, \sigma_c^2] = [\boldsymbol{\alpha}] \times [\mathbf{b}|\phi, \tau^2] \times [\phi] \times [\tau^2] \times [\mathbf{c}|\sigma_c^2] \times [\sigma_c^2]$. The specific prior choices are given in Table 3.1.

For τ^2 and σ_c^2 we use the common choice of inverse Gamma, which is conjugate to the multivariate normal distribution. More specific we denote by $IG(a, b)$ the inverse gamma density given by $[x] = \frac{b^a}{\Gamma(a)x^{a+1}} \exp(-\frac{b}{x})$ for $x > 0$. For $a, b > 0$ this prior is proper, while for $a \leq 0$ or

Parameter	Prior specification
Regression	$\alpha_l \sim N(0,\sigma^2_{\alpha_l}),\ l=1,\ldots,p\ ind. \quad \sigma^2_{\alpha_l}$ large
Spatial	$b_j \vert \mathbf{b_{-j}} \sim N\left(\frac{\phi}{1+\vert\phi\vert N_j}\sum_{j\sim j'} b_{j'}\,,\ \frac{(1+\vert\phi\vert)\tau^2}{1+\vert\phi\vert N_j}\right) j+1,\cdots J$
Spatial dependence	$\psi := \frac{\phi}{1+\vert\phi\vert} \in (-1,1) \quad \pi(\psi) \propto \frac{1}{(1-\vert\psi\vert)^{1-a}},\ a>0$
Spatial variance	$\pi(\tau^2) = IG(a_\tau, b_\tau)\ a_\tau, b_\tau > 0$ or $a_\tau = -1, b_\tau = 0$
Cluster	$c_m \vert \sigma^2_c \sim N(0,\sigma^2_c)\ i.i.d.$
Cluster variance	$\sigma^2_c \sim IG(a_c, b_c)\ a_c, b_c > 0$ or $a_c = -1, b_c = 0$

Table 3.1: Prior distributions used in the spatial logistic regression with group cluster effects

$b \leq 0$ improper priors result. Further $a = -1, b = 0\,(a = -.5, b = 0)$ corresponds to a flat prior for $X(\sqrt{X})$ and $a = 0, b = 0$ to Jeffrey's prior. Sun et al. (2001) and Fahrmeir and Kneib (2007) investigated the question if these improper choices result in improper posteriors. Since Model (3.1)-(3.2) can be represented as a non Gaussian structured additive regression (STAR) proposed by Fahrmeir et al. (2004), it follows under regularity conditions (see Theorem 2 of Fahrmeir and Kneib 2007) and $J > 2$ and $J - p - 4 > 0$ that a flat prior for τ^2 and σ^2_c results in a proper posterior distribution.

MCMC methods allow us to draw approximately arbitrary large samples from the posterior distribution $[\boldsymbol{\alpha}, \mathbf{b}, \mathbf{c}, \phi, \tau^2, \sigma^2_c \vert \mathbf{Y}]$ on which inference is based. Readers unfamiliar with MCMC methods can consult Chib (2001) for an introduction and Gilks et al. (1996) for applications of MCMC methods.

Individual random walk Metropolis Hastings (MH) updates are used for the regression α_l, the spatial b_j and the cluster c_m parameters, since good joint proposal distributions are difficult to find. As individual proposal distributions we use a normal distribution with mean equal to the previous value and a fixed value for the standard deviation. This value is determined by pilot runs, where the standard deviations is adjusted upwards (downwards) in an automatic fashion, when the acceptance rate was too high (low). This resulted in an acceptance rate between 30-60% (as proposed in Bennett et al. 1996 or Besag et al. 1995). They also serve as burnin phase. The re-parameterized spatial hyper-parameter $\psi = \frac{\phi}{1+\vert\phi\vert}$ also requires an MH update.

The hyper-parameters τ^2 and σ^2_c can be updated in a Gibbs step. For the full conditional of τ^2 we have $[\tau^2 \vert \mathbf{Y}, \boldsymbol{\alpha}, \mathbf{b}, \mathbf{c}, \phi, \boldsymbol{\sigma^2_c}] = [\boldsymbol{\tau^2} \vert \mathbf{b}, \phi] \propto [\mathbf{b} \vert \phi, \tau^2] \times [\boldsymbol{\tau^2}]$. Using an $IG(a_\tau, b_\tau)$ prior for τ^2 it is easy to see that $[\tau^2 \vert \mathbf{Y}, \boldsymbol{\alpha}, \mathbf{b}, \mathbf{c}, \phi, \boldsymbol{\sigma^2_c}]$ is again $IG(a^*_\tau, b^*_\tau)$ with $a^*_\tau = a_\tau + \frac{J}{2}$ and $b^*_\tau = \left\{b_\tau + \frac{\mathbf{b}'\,Q(\phi)\,\mathbf{b}}{2}\right\}$. If $\sigma^2_c \sim IG(a_c, b_c)$, the full conditional $\sigma^2_c \vert \mathbf{Y}, \boldsymbol{\alpha}, \mathbf{b}, \mathbf{c}, \phi, \tau^2$ is $IG(a^*_c, b^*_c)$ with $a^*_c = a_c + \frac{M}{2}$ and $b^*_c = \left\{b_c + \frac{\mathbf{c}'\mathbf{c}}{2}\right\}$. A summary of these updates is given in Table 3.2.

4 Spatial Logistic Regression with Individual Cluster Effects

We consider now a more advanced model with individual cluster effects given by:

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

Parameter	Update
$\alpha_l,\ l=1,\dots,p$	Individual MH with normal RW proposal
$b_j,\ j=1,\dots,J$	Individual MH with normal RW proposal
$\psi=\frac{\phi}{1+\lvert\phi\rvert}$	MH Update with $uniform(-1,1)$ proposal
τ^2	Gibbs Update, FC = IG(a_τ^*, b_τ^*)
$c_m,\ m=1,\dots,M$	Individual MH with normal RW proposal
σ_c^2	Gibbs Update, FC = IG(a_c^*, b_c^*)

Table 3.2: Updating schemes of the MCMC algorithm for a spatial logistic regression with group cluster effects (MH = Metropolis Hastings step, RW = random walk, FC = full conditional)

where for $m=1,\dots,M, c_{m,k}\sim N(0,\sigma_m^2)$ i.i.d., $k=1,\dots,K_m$. As in Model (3.2), M denotes the number of clusters and $m(i)$ denotes the cluster of trip i. K_m counts the number of trips, which belong to cluster m and $k(i)$ gives the number of trip i in its cluster. The specification of $\boldsymbol{\alpha}$ and $\mathbf{b}$ 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 $N(0,\sigma_m^2)$. This allows for heterogeneity within each cluster.

In Model (4.1) we have to estimate in addition to $\boldsymbol{\alpha},\mathbf{b}$ the cluster effect variances $\boldsymbol{\sigma^2}=(\sigma_1^2,\dots,\sigma_M^2)^t$. One problem with Model (4.1) is that even without an intercept term α_0 the model is unidentifiable. To understand this non-identifiability we first substitute in (4.1) the logit link function with the probit link function, i.e.

$$\begin{aligned} Y_i|p_i &\sim \text{Bernoulli}(p_i) \text{ conditionally independent with} \\ p_i &= \mathbf{P}\{Y_i=1|\mathbf{x_i},\boldsymbol{\alpha},b_{j(i)},c_{m(i),k(i)}\}=\Phi(\mathbf{x_i^t}\boldsymbol{\alpha}+b_{j(i)}+c_{m(i),k(i)}) \end{aligned} \tag{4.2}$$

where $\Phi(\cdot)$ is the standard normal distribution function. This allows for the latent variable representation (compare to Albert and Chib 1993):

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

We have for $i=1,\dots,n$

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

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

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

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:

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

Using (4.3) it follows, that also the joint distribution of $\mathbf{Y}$ in Models (4.2) and (4.6) are equal. Therefore Model (4.6) is an equivalent re-parametrization of Model (4.2). But this representation (4.6) has one parameter less and is therefore identifiable. The above discussion helps us to understand the non-identifiability of logit model (4.1), since the behavior of both probit and logit link functions is quite similar and they differ only significantly in the tails. We use the same idea to construct an identifiable logit model. In particular we assume for $i = 1, \ldots, n$

$$\begin{aligned} &Y_i|p_i \sim \text{Bernoulli}(p_i) \text{ conditionally independent with} \\ &\log\left(\frac{p_i}{1-p_i}\right) = \begin{cases} \mathbf{x_i^t}\boldsymbol{\alpha}' + b'_{j(i)} & \text{if } m(i) = 1 \\ \frac{\mathbf{x_i^t}\boldsymbol{\alpha}' + b'_{j(i)}}{\sigma'_{m(i)}} & \text{if } m(i) = 2, \ldots, M \end{cases}, \end{aligned} \tag{4.7}$$

where $\boldsymbol{\alpha}', \mathbf{b}', \boldsymbol{\sigma}^{\mathbf{2}'} := (\sigma_1^{2'}, \ldots, \sigma_M^{2'})^t$ are defined as in (4.5). From (4.7) it follows that the likelihood of the response vector $\mathbf{Y}$ is proportional to

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

where $\boldsymbol{\sigma}' := (1, \sigma_2', \ldots, \sigma_M')^t,\ \sigma_m' := \sqrt{\sigma_m^{2'}},\ m = 2, \ldots, M$. We use the hyper-parameters $\phi' := \phi$ and $\tau^{2'} := \frac{\tau^2}{\sqrt{1+\sigma_1'^2}}$ and assume the following joint prior $[\boldsymbol{\alpha}',\ \mathbf{b}',\ \boldsymbol{\sigma}',\ \phi', \tau^{2'}] = [\boldsymbol{\alpha}'] \times [\mathbf{b}'|\phi', \tau^{2'}] \times [\phi'] \times [\tau^{2'}] \times [\boldsymbol{\sigma}']$. Large deviations in (4.5) from 1 for some σ_m', $m = 2, \ldots, M$, correspond to large values for some σ_m^2, $m = 1, \ldots, M$, in the primary Model (4.1), which implies insignificant regression and spatial effects in these clusters. Therefore one should use a prior for σ_m' which is relatively concentrated around 1. To avoid boundary problems a normal distribution $N(1, 4)$ truncated on the interval $[0.2, +\infty)$ is chosen as prior for σ_m', $m = 2, \ldots, M$. Even though σ_1' is fixed to 1, a value of $\sigma_m' \geq 1 (\leq 1)$ corresponds to $\sigma_m^2 \geq \sigma_1^2 (\sigma_m^2 \leq \sigma_1^2)$. This shows that our prior choice can support high and low variability of cluster m compared to cluster 1.

The parameters $\boldsymbol{\alpha}', \mathbf{b}'$ and $\boldsymbol{\sigma}'$ will be updated using individual MH steps. Since the full conditionals of $\tau^{2'}$ and ϕ' depend only on the spatial effects $\mathbf{b}'$, their MCMC updates have the same form as described in Table 3.2.

5 Simulation Studies

We conducted two simulation studies for spatial logistic regression one with group cluster and the other with individual cluster effects. The first study is based on logit model (3.2) with

$$\theta_i := \log\left(\frac{p_i}{1-p_i}\right) = x_{1i}\alpha_1 + x_{2i}\alpha_2 + b_{j(i)} + c_{m(i)}$$

for $i = 1, \ldots, n, j = 1, \ldots, J, m = 1, \ldots, M$. Adapted to our mobility study we simulated $n = 2100$ binary responses residing in $J = 70$ regions arranged on a 7×10 regular lattice and in $M = 5$ clusters so, that each cluster is represented in each region with 6 responses. 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 with $\alpha_1 = -1$ and $\alpha_2 = 0.05$. Spatial effects $\mathbf{b}$ are simulated from Model (2.3) with $\phi = 2$ giving significant spatial smoothing. We chose $\tau^2 = 0.64$ which gives a similar range of the observed spatial effects in the mobility data. A first order

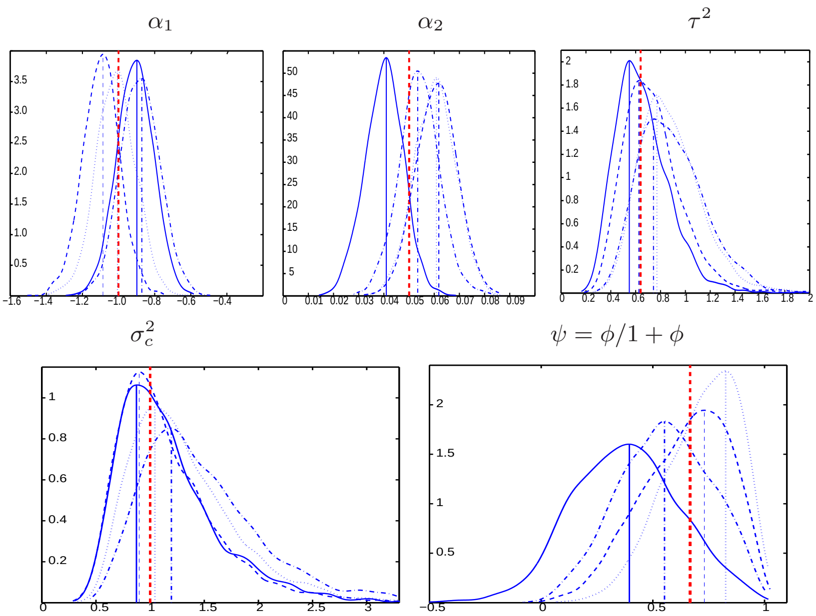

Figure 5.1: Estimated marginal posterior densities for $\alpha_1, \alpha_2, \tau^2, \psi, \sigma_c^2$ in Model (3.2) (solid for Data Set 1, dashed for Data Set 2, dash-dot for Data Set 3, dotted for Data Set 4)

neighborhood dependence defined by joint borders was selected. We simulated group cluster effects from $\mathbf{c} \sim N_5(0, \sigma_c^2)$ with $\sigma_c^2 = 1$. As priors we chose $\alpha_1 \sim N(0, 100^2)$, $\alpha_2 \sim N(0, 10^2)$ and $\tau^2 \propto 1$ reflecting a diffuse prior choice. For $\psi = \frac{\phi}{1+|\phi|}$, $\psi \in (-1, 1)$ the $J = 70$ regions may be too few to provide enough information for estimating ψ. Therefore we chose an informative prior density $[\psi] \sim \frac{1}{(1-|\psi|)^{1-a}}$ with $a = 1.25$. Similar, we took an informative prior for σ_c^2, namely $IG(5, 1/6)$ with mean 1.5, variance 3/4 and mode 1.

The MCMC algorithm of Section 3.2 was implemented in MATLAB and run for 20,000 iterations with every 10th iteration recorded. As "burn in" phase served 5 consecutive pilot runs with 300 iterations per each pilot run giving an acceptance rate of $30 - 60$ % for the MH step. The resulting trace plots (not shown) show that such a "burn in" is enough. Further empirical 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 four independent simulated data sets, where the vertical fat dashed lines correspond to the true parameter value. For simplicity we used kernel density plots for the recorded iterations, however estimates based on Gelfand and Smith (1990) could have been used as well. The mode of each density curve is marked by a thin vertical line. In all cases the true values are well inside 90% credible intervals (CI). Although the estimation of ψ is somewhat dispersed, posterior mode estimates of spatial and cluster effects (not shown) are quite precise.

The second simulation is based on the logit model (4.7) with mean structure:

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

We used the same spatial and fixed effect structure as in the first simulation, in particular we set $\alpha_1' = -1,\ \alpha_2' =\ 0.05,\ \tau^{2'} =\ 0.64,\ \phi' = 2$. As true values for the cluster parameters $\sigma_m',\ m = 2, \dots, M$, we take values $0.5, 1.25, 1.5$ and 2.5, respectively. According to Model (4.7) we set

$\sigma_1' = 1$. Prior choices for $\boldsymbol{\alpha}', \mathbf{b}'$ and $\tau^{2'}$ remain the same. For the prior of $\psi' = \frac{\phi'}{1+|\phi|'}$ we used $Uni(-1,1)$, since a similar prior choice as for the group cluster case causes slight underestimation of ψ'. For the cluster parameters σ_m', $m = 1, \ldots, M$, we used a $N(1,4)$ distribution truncated to $[0.2, +\infty]$ as prior. Figure 5.2 gives posterior density estimates of the cluster variance parameters $\boldsymbol{\sigma}'$ using the MCMC algorithm of Section 4.2 based on 20000 iterations with every 10th iteration recorded indicating a satisfactory behavior. The remaining parameters show a similar behavior as the corresponding parameters in logit model (4.7) (not shown).

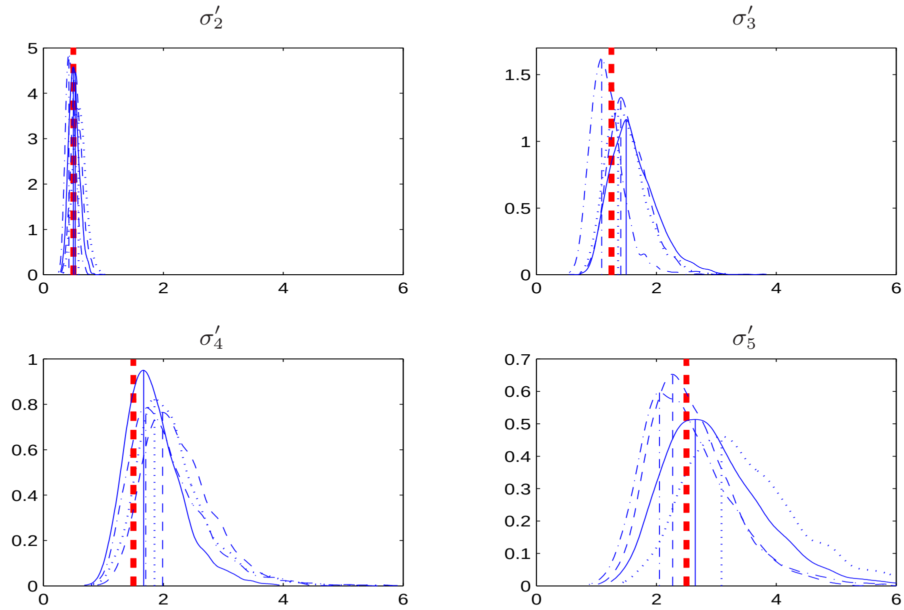

Figure 5.2: Estimated marginal posterior densities for $\boldsymbol{\sigma}'$ in Model (4.7) (solid for Data Set 1, dashed for Data Set 2, dash-dot for Data Set 3, dotted for Data Set 4)

Summarizing we see that all posterior estimates for the main parameters $\boldsymbol{\alpha}$, $\mathbf{b}$, $\mathbf{c}$ for Model (3.2) and $\boldsymbol{\alpha}$, $\mathbf{b}$, $\boldsymbol{\sigma}'$ for Model (4.7) lie quite closely around the corresponding true values. With regard to the spatial hyper-parameters τ^2 $(\tau^{2'})$ and ψ (ψ') for the group cluster Model (3.2) (individual cluster Model (4.7)) we conclude that the number of regions $J = 70$ is enough for τ^2 $(\tau^{2'})$, while this is not the case for ψ (ψ'). Further, a simulation study with a large number of regions showed good precision for estimation of ψ, as well as robustness of the posterior with respect to prior choice already for $J = 500$. In contrast to the individual cluster Model (4.7),the small number of clusters M in the group cluster Model (3.2) causes lack of information for estimating σ_c^2. The computational effort for the simulation results presented required about 7 hrs (14 hrs) per data set for the first (second) simulation setup on a SUN workstation with CPU 600 MHz and RAM 2.56 GB.

Covariable type	Variable	Levels	Number of Trips Using Individual Transport	Public Transport	Total	Most frequently obs. value
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 *poly(age,2)*)				**median** 42 years
HOUSE-HOLD 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: Covariates identified in logistic model selection without spatial and cluster effects

6 Application: Mobility Data

6.1 Data Description

We analyze a data set studying mobility behavior of private households in Munich. We want to identify areas of low/high utilization of public transport after adjusting for trip, individual and household related attributes. Since extra heterogeneity is to be expected we allow for cluster effects. The data was collected within the study “Mobility 97” (Zängler 2000). The participants are German-speaking persons not younger then 10 years, who live in a private household in the state of Bavaria. In order to take into consideration seasonal fluctuations in mobility behavior the survey was carried out in three waves in March, June and October of 1997 with different participants for each wave. 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 213 persons in 123 households living in 74 postal code areas of Munich. For each trip Y has value 1, if individual transport was used and value 0, if public transport was used. 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, must have a driver license and a car available in the household. In addition we allowed for interaction effects, which were identified by partial deviance tests. The following significant interactions were identified: 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 and DAY TYPE:AGE. We used this model as a starting model. We note that a seasonal effect measured by temperature is not significant. Trips which have been taken together by let us say k persons are treated as k individual trips with person specific covariates. The fact that these trips were conducted together is taken into account by the covariate WAY ALONE.

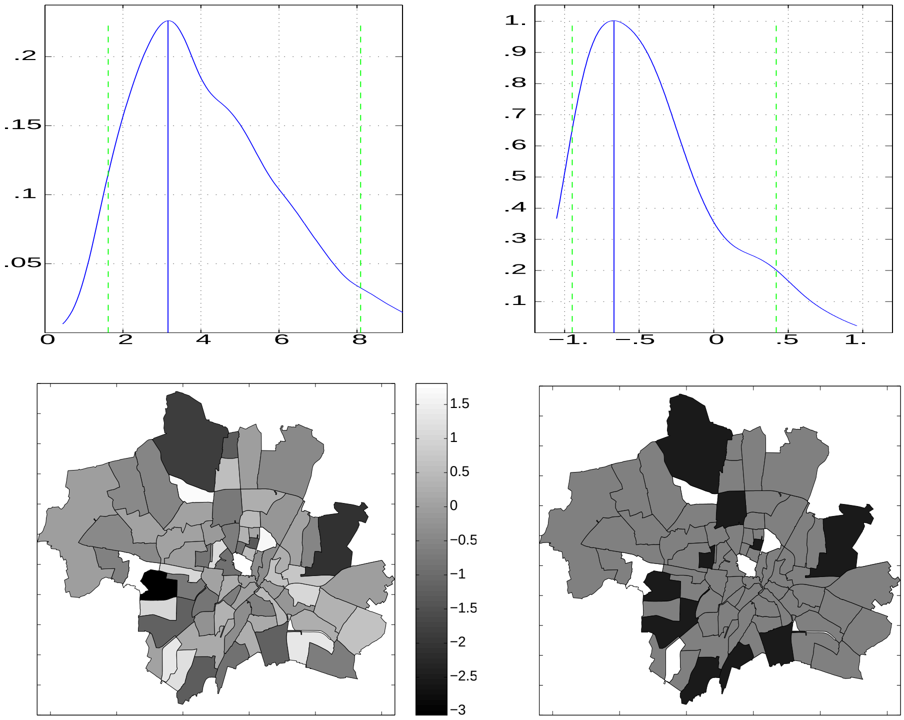

Figure 6.1: Results for Model 1: Top: Estimated posterior densities of spatial hyperparameters τ^2 and ψ, respectively (solid line = estimated posterior mode, dashed line = 90% CI). Bottom: Estimated posterior mean spatial effects $\hat{b}_j$, $j = 1, \ldots, 74$ and 90% CI (white=0 below 90% CI, black=0 above 90% CI and gray=0 $\in$ 90% CI)

6.2 Results

We present results for 8 different model specifications. Model 1 is a spatial logit regression model with no cluster effects, while Models 2 - 5 are spatial logit models with group cluster effects. Finally Models 6 - 8 are spatial logit models with individual cluster effects. After 10 consecutive pilot runs with 300 iterations each 25000 MCMC iterations were run and every 25^{th} iteration was recorded, giving acceptable low autocorrelations (not shown).

As a starting point for the choice of fixed effects we used the starting model discussed above with 36 regression parameters. The intercept effect is modeled within the spatial and cluster part. For $\alpha_1, \ldots, \alpha_{36}$ we chose independent normal priors with zero mean and standard deviation equal to 5. We consider an interaction as insignificant when the corresponding estimated 90%

CI contains the zero value for all interaction terms. If an interaction is found to be insignificant, then the corresponding terms were removed and the model re-estimated. Continuing with this procedure we arrive at a model where all interactions are significant.

For Model 1 we chose a uniform prior for $\psi = \frac{\phi}{1+|\phi|}$ on $(-1, 1)$ and for $[\tau^2] \propto 1$. The top row of Figure 6.1 shows estimated posterior densities for τ^2 and ψ. Since $\psi < 0$ positive spatial effects in an area can be surrounded by negative spatial effects and vice versa. This is seen in the bottom row of Figure 6.1.

In Models 2 - 5 we used group cluster specifications. In Model 2 group clusters are formed by the 74 postal codes. Since single trips are taken by individuals and households we ideally would like to allow for person or household specific effects. However we only have 123 households and 213 persons available over 74 postal code areas. Therefore serious confounding with the group random effects in Model 2 is to be expected. Further the results of Model 2 presented below show that the unstructured postal code random effects are not significant, therefore we choose not to allow for person or household specific random effects. However we are able to discover further structures from household and person specific behaviors such as the intensity a household or person undertakes trips. Therefore we form cluster groups by the number of trips taken by individuals or households. Since it is unclear how many cluster groups should be considered, we investigated several specifications. To avoid unbalanced cluster groups we chose the cut points given by empirical quantiles in such a way that the corresponding cluster groups consists of about equal number of trips. For example for Model 3 we used 5 clusters with 1st cluster group defined with ≥ 23 trips, the 2nd one with $16 - 22$ trips, the 3rd one with $12 - 15$ trips, the 4th one with $8 - 11$ trips and the 5th one with ≤ 7 trips per household. Model 4 uses 12 cluster groups formed by the number of trips a household has taken, while Model 5 uses 5 cluster groups formed by the number of trips a person has taken.

For σ_c^2 we choose the relatively uninformative prior $IG(3, 2)$, 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 hyper-parameters and their 90% CI are given in Table 6.2. In Model 2 we have as cluster groups the 74 postal codes. Therefore both structured ($b_j,\ j = 1, \ldots, 74$) and unstructured ($c_j,\ j = 1, \ldots, 74$) spatial effects are included in Model 2. Figure 6.2 presents 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 6.2. Both structured and unstructured effects are insignificant, while their sum is significant, and form a similar spatial pattern as in Model 1. Therefore it is not surprising that the posterior density of ψ, are also similar (see bottom row of Figure 6.2).

Figure 6.3 shows estimated posterior densities of the group cluster effects c_m for Model 3. A cluster effect is significant (marked with *), if $0 \notin 90\%$ CI. Cluster effects for households with large (small) numbers of trips are positive (negative). The bottom row of Figure 6.4 shows the estimated spatial effects.

Also in Models 4–5 only the higher cluster effects (i.e. with fewest numbers of trips) are significant. For brevity we omit the corresponding density plots. For Models 4–5 the 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. The posterior density of ψ also remains

Model	Number of Clusters	Parameter	Mode	Mean	Median	90% CI 5%	95%
	74 formed	ψ	−0.500	−0.271	−0.372	−0.857	0.646
2	by postal	τ^2	3.628	4.777	4.313	0.981	10.335
	codes	σ_c^2	0.554	0.836	0.678	0.315	1.912
	5 formed	ψ	−0.541	−0.422	−0.446	−0.930	0.149
3	by # of trips	τ^2	6.262	9.124	8.233	3.358	18.417
	per household	σ_c^2	0.802	1.270	1.076	0.486	2.797
	12 formed	ψ	−0.507	−0.516	−0.538	−0.954	0.031
4	by # of trips	τ^2	6.293	8.299	7.452	3.194	16.067
	per household	σ_c^2	0.880	1.272	1.122	0.589	2.398
	5 formed	ψ	−0.874	−0.543	−0.594	−0.956	0.058
5	by # of trips	τ^2	4.025	5.298	4.777	2.020	9.685
	per person	σ_c^2	0.526	0.753	0.646	0.324	1.585
6	3 formed	ψ	−0.468	−0.396	−0.418	−0.870	0.181
	by household	$\tau^{2'}$	4.861	6.854	5.931	2.553	14.196
	type	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	5 formed	ψ	−0.410	−0.413	−0.422	−0.865	0.075
	by # of trips	$\tau^{2'}$	10.769	17.101	14.799	6.002	36.512
	per household	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
8	5 formed	ψ	−0.476	−0.403	−0.439	−0.876	0.199
	by # of trips	$\tau^{2'}$	7.538	9.468	8.232	3.167	19.895
	per person	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

Table 6.2: Point and interval posterior estimates for the hyper-parameters in Models 2 - 5 (with group cluster effects) and Models 6 - 8 (with individual cluster effects)

similar (not shown for Models 3, 4 and 5). Table 6.2 gives posterior centrality estimates and 90% CI's for the hyper-parameters.

We consider now model specifications with individual cluster effects given in Table 6.2. As before, we chose a flat prior $[\tau^{2'}] \propto 1$ and take $[\psi'] \propto (1 - |\psi'|)^{0.5}$ to avoid numerical problems. In Models 6–8 we assume for $\sigma'_2, \ldots, \sigma'_M$ a normal $N(1,1)$ prior truncated to $(0.2, +\infty)$. The posterior centrality estimates and their 90% CI's of the hyper-parameters for Models 6–8 are given in Table 6.2. We see that cluster components of the higher clusters are significant, i.e. $1 \notin 90\%\ CI$. This shows that the heterogeneity within the group with the fewest numbers of trips per household (or per person) is the largest. Further, 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 hyper-parameter ψ is negative and about the same size.

The estimates for the fixed effects $\boldsymbol{\alpha}'$ in all 8 models are given in Table 6.3. Posterior mode estimates are marked with *, when the corresponding parameter is insignificant, i.e. $0 \in 90\%$ CI. If all terms of an interaction effect were insignificant, the model was reduced and re-estimated. Those interactions are marked with "*n.r.*", correspond to "not represented" in the model. In particular the significant interactions PERSONAL INCOME:NET CARD and DAY TIME:AGE from the starting logistic model disappear. Further, the interaction effect DISTANCE:AGE is only significant in a model without group or individual cluster effects. This shows that some of the extra heterogeneity is captured by this interaction.

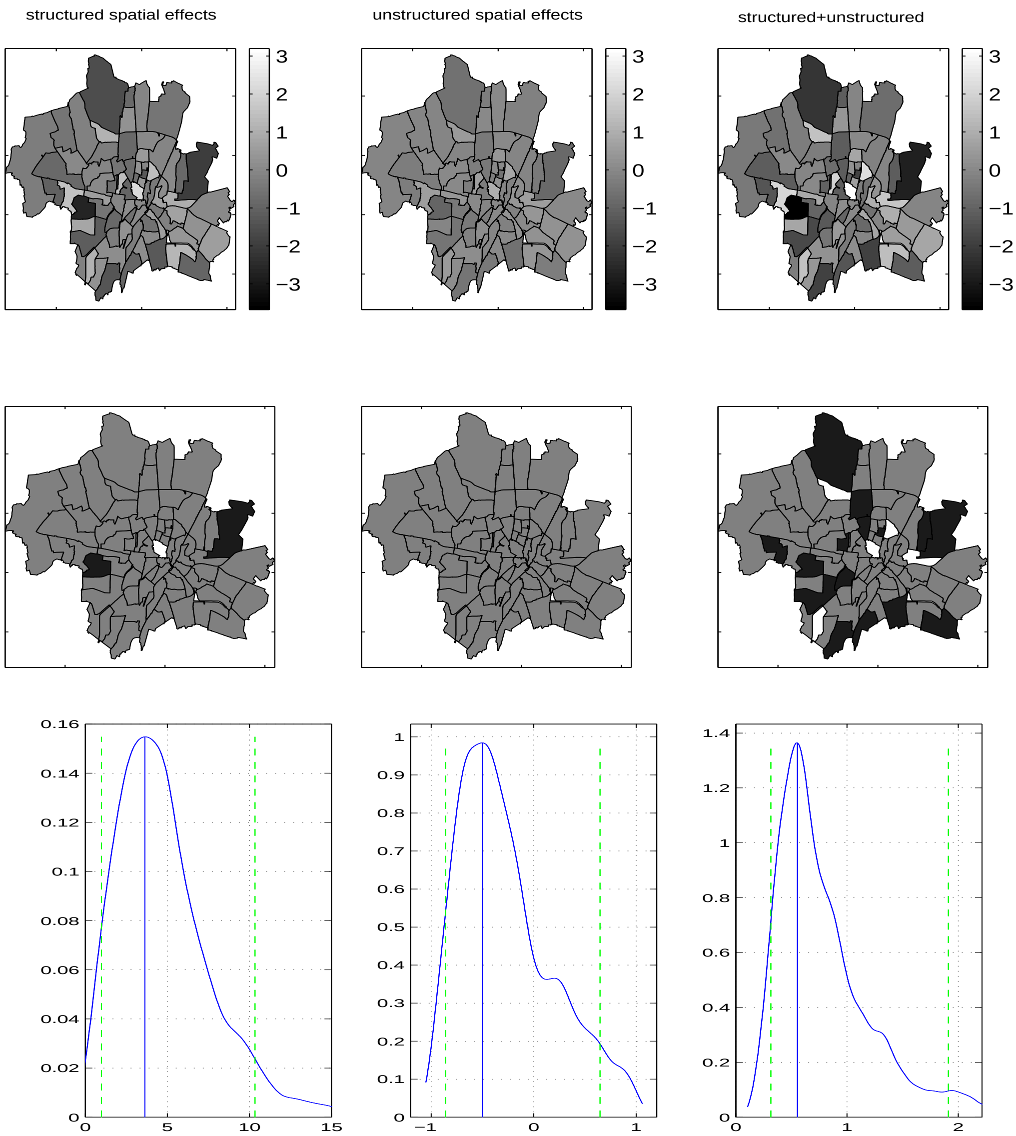

Figure 6.2: Results for Model 2: Top: estimated posterior mean spatial effects: structured $\hat{b}_j, j = 1 : 74$ (left), unstructured $\hat{c}_j, j = 1 : 74$ (middle) and their sum $\hat{b}_j and \hat{c}_j, j = 1 : 74$ (right). Middle: 90% CI's for structured effects, unstructured effects and their sum (white = 0 below 90% CI, black = 0 above 90% CI, gray = 0 $\in$ 90% CI). Bottom: estimated posterior densities of hyper-parameters τ^2, ψ and σ_c^2 (solid line = estimated posterior mode, dashed line = 90% CI)

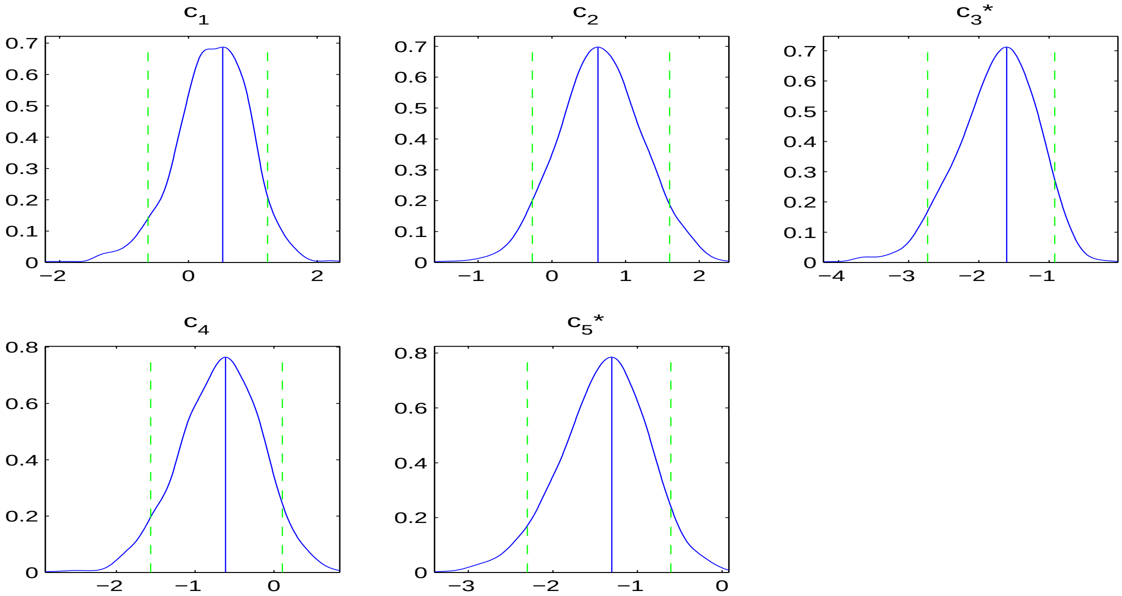

Figure 6.3: Estimated posterior densities of group cluster effects c_m, $m = 1, \ldots, 5$ in Model 3. (solid line = estimated posterior mode, dashed line = 90% CI)

	Model							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	spatial only	**spatial+group cluster**				**spatial+individual cluster**		
Main Effects								
PERSONAL INCOME								
MIDDLE	0.41*	0.48*	1.63	1.41	0.71*	1.06	1.62	0.64*
HIGH	0.25*	0.42*	1.27	1.14	0.12*	0.76	1.46	0.24*
USAGE								
SECOND.USER	0.38*	1.09*	1.27	1.16*	0.88*	1.11	1.23*	1.51
NOT.USER	−3.87	−6.41	−6.52	−6.52	−5.90	−6.38	−9.99	−7.44
NET CARD								
NO	2.07	2.67	3.03	3.32	2.72	2.78	3.90	3.11
SEX								
FEMALE	0.28*	0.16*	−0.19*	−0.47*	0.10*	0.30*	−0.48*	0.01*
AGE								
POLY.AGE.1	16.80	8.73	11.64	11.53	9.95	6.11	7.97	9.81
POLY.AGE.2	−13.07	−8.96	−9.03	−8.64	−9.67	−8.93	−9.69	−7.63
HOUSEHOLD								
SINGLE.PARENT	1.61	3.15	3.42	2.92	3.31	n. r.	4.24	3.65
NOT.SINGLE	0.70	0.68	0.25*	0.27*	0.90	n. r.	0.85*	0.96
DAY TYPE								
WEEKEND	1.44	2.21	2.46	2.52	2.11	2.25	3.32	2.78
DISTANCE								
MIDDLE	−0.96	−1.15	−1.06	−1.17	−1.05	−1.29	−1.90	−1.16
FAR	0.32*	0.81*	0.98*	0.83*	0.97*	1.21*	0.78*	0.85*
WAY ALONE								
NOT.ALONE	1.82	2.09	2.07	2.30	1.93	2.17	3.21	2.30
DAY TIME								
NIGHT	−0.58*	−1.02	−1.12	−1.29	−1.13	−1.19	−1.99	−1.30
Interaction Effects								
WAY ALONE:NET CARD								
NOT.ALONE:NO	−1.86	−2.39	−2.37	−2.76	−2.37	−1.54	−1.69	−2.53
USAGE:SEX								
SECOND.USER:FEMALE	−1.70	−2.13	−2.07	−1.81	−2.01	−2.30	−2.80	−2.50
NOT.USER:FEMALE	−0.20*	0.66*	0.58*	0.79*	0.40*	0.26*	1.39*	0.80*
WAY ALONE:USAGE								
NOT.ALONE:SECOND.USER	0.79	1.21	0.80	0.76*	1.22	1.09	1.20*	1.32
NOT.ALONE:NOT.USER	1.75	3.65	4.19	3.76	3.41	4.35	5.08	4.22
DISTANCE:USAGE								
MIDDLE:SECOND.USER	−0.68*	−1.03	−1.39	−0.97	−1.19	−1.31	−1.44*	−1.54
FAR:SECOND.USER	−1.02	−2.25	−2.12	−1.72	−2.22	−2.73	−2.41	−3.61
MIDDLE:NOT.USER	0.95	1.68	1.52	1.64	1.27	1.20	2.47	1.53
FAR:NOT.USER	−1.19	−1.19*	−1.55*	−2.01	−1.51*	−2.31	−2.68	−1.94
DAY TYPE:NET CARD								
WEEKEND:NO	n. r.	−0.91	−1.23	−1.23	−1.07	−0.82*	−1.51	−1.25
USAGE:DAY TIME								
SECOND.USER:NIGHT	1.32	5.01	5.22	6.63	5.71	5.07	6.17	5.67
NOT.USER:NIGHT	−0.06*	0.31*	0.45*	0.38*	0.26*	0.32*	0.72*	0.68*
SEX:DAY TIME								
FEMALE:NIGHT	1.70	2.88	3.36	3.55	3.49	3.02	2.94	3.30
DISTANCE:AGE								
MIDDLE:POLY.AGE.1	−12.93	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.
MIDDLE:POLY.AGE.2	−2.41*	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.

Table 6.3: Posterior mode estimates for main effect and interaction parameters (* = 90% CI does not include 0, n.r.= effect was not required in model, since model with effect has a 90% CI which includes 0)

	Model 1 spatial only	Model 2 spatial + group cluster	Model 3	Model 4	Model 5	Model 6 spatial + individual cluster	Model 7	Model 8
fixed effects	31	28	28	28	28	26	28	28
spatial effects	74	74	74	74	74	74	74	74
cluster effects	0	74	5	12	5	2	4	4
total number of parameters for D_w	105	176	107	114	107	102	106	106
$\mathbf{D_w}$	2.35	1.23	0.95	1.02	1.44	1.9	3.25	1.84
$\sum_{i=1}^{n}(\mu_i - y_i)^2$	110.49	106.54	96.05	94.91	102.86	108.30	103.77	107.56
$\sum_{i=1}^{n}\sigma_i^2$	129.75	111.96	104.19	102.64	109.84	114.21	111.50	114.55
PMCC	240.24	218.50	200.23	197.55	212.70	222.51	215.27	222.11
BS	.0866	.0840	.0768	.0761	.0812	.0851	.0821	.0850

Table 6.4: Model fit comparison using D_w, PMCC and BS

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). It is developed for exponential family models and based on the deviance. Even though binary logit models belong to this class, Collett (2002) has shown that the residual deviance in binary regression should not be used for model assessment, while the partial deviance is valid for nested model comparison. Further, Figure 1 of Spiegelhalter et al. (2002) shows that DIC does not perform satisfactory for binary responses. Since our binary responses 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 DIC. Meaningful DIC values of our models can be determined as long as the binary regression data can be grouped to binomial regression data with sufficiently large number of trials.

To facilitate model comparison we follow two approaches. In the first one we focus on the spatial fit, while in the second one we focus on the overall fit. The first focus uses D_w the sum of weighted squared residuals over all postal codes of Munich defined by

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

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

$$p_j^{\text{estim}} := \frac{1}{n_j * R} \sum_{i:\, j(i)=j} \sum_{r=1}^{R} \frac{\exp(\eta_{ir})}{1+\exp(\eta_{ir})}, \tag{6.2}$$

where

$$\eta_{ir} := \begin{cases} \mathbf{x_i^t}\boldsymbol{\alpha}_r + b_{j(i),r} & \text{for Model 1} \\ \mathbf{x_i^t}\boldsymbol{\alpha}_r + b_{j(i),r} + c_{m(i),r} & \text{for Models 2-5} \\ \frac{\mathbf{x_i^t}\boldsymbol{\alpha}'_r + b'_{j(i),r}}{\sigma'_{m(i),r}} & \text{for Models 6-8 .} \end{cases}$$

Here $\boldsymbol{\alpha}_r, b_{j,r}, c_{m,r}$ and $\sigma'_{m,r}$ are the corresponding MCMC values in the rth recorded iteration.

In Table 6.4 we present value D_w for all 8 models and the number of parameters required in calculating 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 ψ, σ^2 and σ_c^2 in group cluster models as hyper-parameters 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. For the focus on assessing the spatial fit, the corresponding 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 Spiegelhalter et al. (2002). According to Table 6.4 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.

To complement our analysis of spatial fit we consider now also the predictive model choice criterion (PMCC) of Gelfand and Ghosh (1998) and the Brier score BS (Brier 1950) as proper scoring rule (Gneiting and Raftery 2007). The PMCC is defined as

$$PMCC = \sum_{i=1}^{n} (\mu_i - y_i)^2 + \sum_{i=1}^{n} \sigma_i^2$$

where $\mu_i := \frac{1}{R}\sum_{r=1}^{R} p_{ir}$ and $\sigma_i^2 := \frac{1}{R}\sum_{r=1}^{R} p_{ir}(1 - p_{ir})$ are MCMC based estimates of the mean and variance of the posterior predictive distribution for the ith observation. In our data set $n = 1375$ and $p_{ir} = \frac{\exp(\eta_{ir})}{1+\exp(\eta_{ir})}$. The second term is considered as a penalty term which will tend to be large both for poor and over-fitted models. Here the BS score is given by

$$BS = \frac{1}{nR} \sum_{r=1}^{R} \sum_{i=1}^{n} (p_{ir} - y_i)^2.$$

Both PMCC and BS are not optimal. First PMCC is not a proper scoring rule, second BS is a coarse measure resulting from the binary nature of the data, which utilizes only the mean of the distribution. In addition we chose not to validate the models using out-of-sample predictions since the data are sparse and a validation data set would cover only a small percentage of the postal code areas investigated. PMCC and BS are given in Table 6.4 and again show that Models 3 and 4 are the preferred models. This substantiates that Model 3 is the preferred overall model.

For Model 3 we present a map with estimated spatial probabilities over postal codes of Munich (Figure 6.4, top right map), which coincides quite well with the map showing the empirical spatial probabilities (Figure 6.4, top left map). This indicates that Model 3 has a reasonably good fit of the data with respect to the spatial resolution.

6.4 Model Interpretation

After model fitting and model selection one is interested in what can be learned about the travel mode decisions based on Model 3. First we estimate individual transport probabilities when one or combinations of two covariates change. The remaining covariates are set to their "most usual values", corresponding to the mode for categorical covariates and median values for quantitative covariates (Table 6.1). Since Model 3 includes spatial effects we have to specify a postal code for which we estimate these probabilities. We chose 81377, since this postal code area has a large observed number of trips and the smallest 90% CI for its spatial effect. Since each cluster

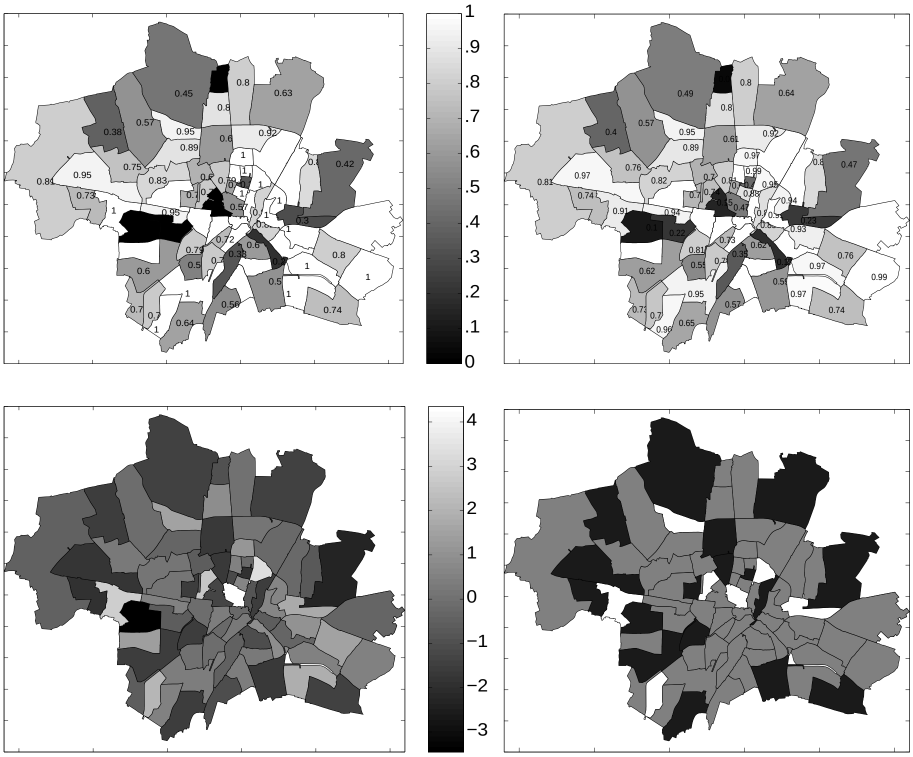

Figure 6.4: Top right: Observed probabilities of individual transport use by postal codes in Munich, Germany; Top left: Posterior mean probability estimates of individual transport use by postal codes in Munich, Germany for Model 3; Bottom left: Estimated posterior mean spatial effects $\hat{b}_j,\ j = 1, \ldots, 74$ in Model 3; Bottom right: 90 % CI (white = 0 below 90% CI, black = 0 above 90% CI, gray = 0 $\in$ 90% CI)

group contains a similar number of individual trips, for our investigations we chose the last, i.e. the 5th cluster group corresponding to households with ≤ 7 trips, which has the smallest 90% CI for its cluster effect c_5. For "the most usual" trip associated with postal code 81377 and 5th cluster, the estimated posterior mean probability for taking individual transport is equal to 0.7.

Figure 6.5 gives the estimated posterior mean probability with 90% CI's for choosing individual transport as age changes in 81377 and trips associated with the 5th cluster when the remaining covariates are set to their "most usual value". It is not 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 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. 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 for using individual transport. For brevity we interpret only 2 of the 7 interaction plots. From top left panel of Figure 6.6 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 in general a net card can only be used by a single person. Further a net card

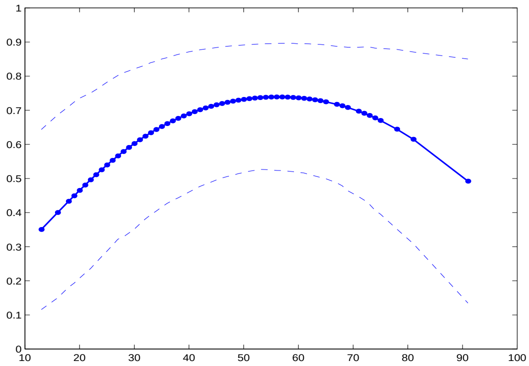

Figure 6.5: Estimated posterior mean individual transport probabilities in 81377 and 5th cluster group for different AGE, while other covariates are set as in Table 6.1 (dotted lines correspond to 90% CI's)

holder might use it for commuting to work, which is a solo activity. In contrast users without a net card take individual transport options much more often regardless if the trip is taken alone or not. The right panel in the second row shows an interesting behaviorial difference between females and males. During the day there is a little difference. However during 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, while males might prefer a car free option at night. This shows that some expected behaviorial patterns can be captured when interactions are allowed in the model. The remaining panels of Figure 6.6 are interpreted in detail in Section 6.4 of Prokopenko (2004).

We continue now with the interpretation of spatial effects. There are 24 postal codes whose 90% CI's do not include zero and therefore are significant. 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 6.5 confirms our assumption in general. The left column shows the numbers of postal code areas, which have U- or S-stops inside. The right column contains the numbers

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

Table 6.5: 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.

of postal code areas without stops. The estimated odds ratio of Table 6.5 is $\frac{2 \cdot 6}{11 \cdot 5} \approx 0.22$, which is below 1, however a 90% confidence interval is $[0.044, 1.091]$. This indicates 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 6.5). These areas should therefore be of special interest to

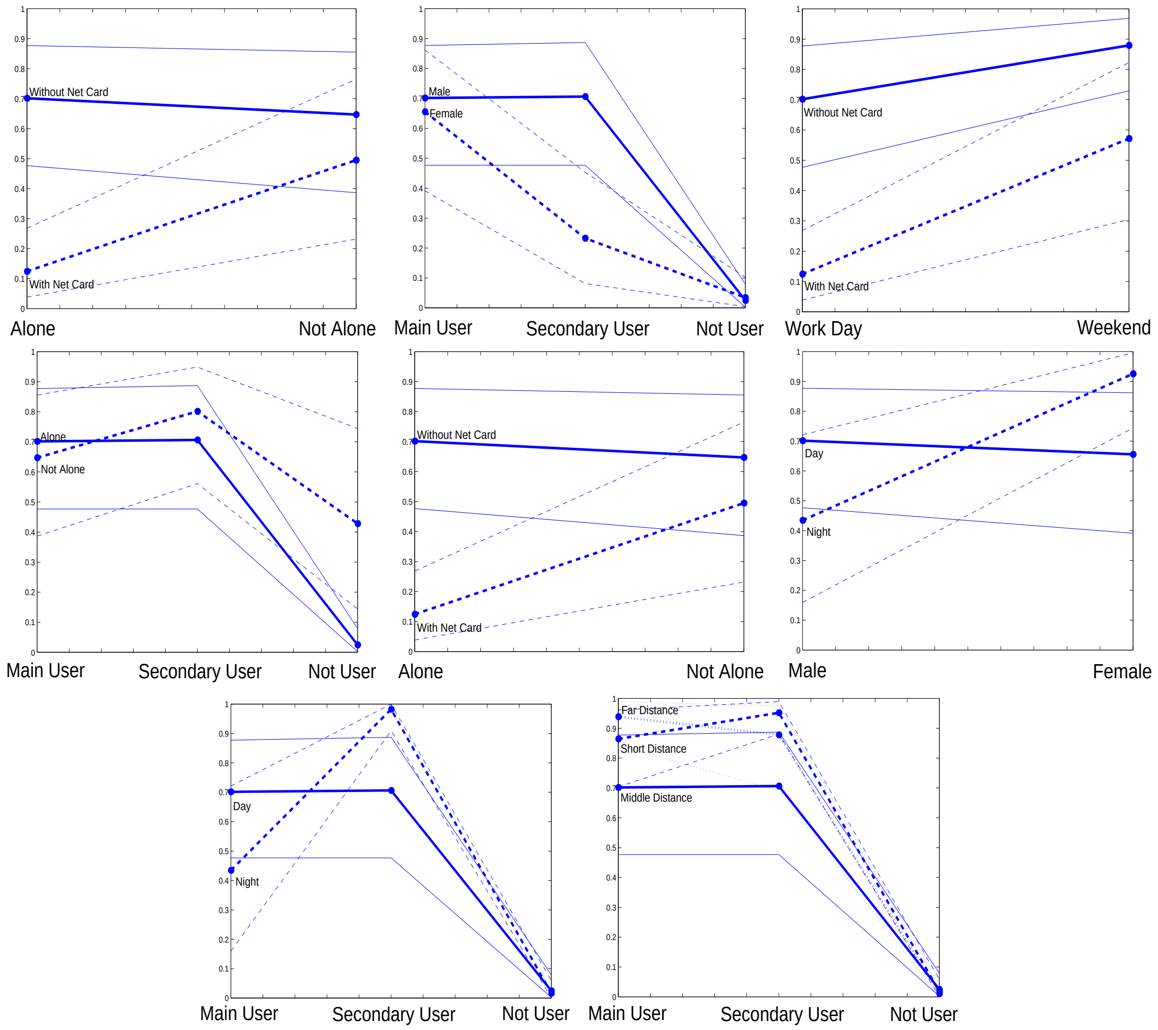

Figure 6.6: Estimated posterior mean individual transport probabilities in 81377 and 5th cluster group for different covariate combinations which form the interaction, while other covariates are set to the “most usual value”. Dotted lines correspond to 90% CI’s.

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, household specific effects and the structure of the public transportation network. We noted that the estimate of the spatial dependency parameter $\hat{\psi} \approx -0.5$ is negative. This can be possibly 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 (Figure 6.3). This implies that households with high mobility needs use a car more often than households with low mobility needs.

7 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. This model possesses a proper joint distribution and allows for a fast update of the spatial dependence parameter. Additionally, 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). The application to the mobility data shows that such a model extension is useful, since ψ is significantly different than unity, which corresponds to the intrinsic CAR model.

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 yields an unidentifiable model. It is shown how the model can be re-parameterized to overcome non-identifiability. Parameter estimation is facilitated by an MCMC approach. Separate MCMC algorithms have been developed for the two hierarchical model classes considered: logistic regression with spatial and group cluster effects and logistic regression with spatial and individual cluster effects. Probit formulations could have used as well and have been investigated in Prokopenko (2004). There latent variables are used for probit models with individual cluster effects requiring only a single MH update. This is faster because of better mixing behavior than a corresponding MCMC algorithm based on the logit formulation. However logit formulations are easier to interpret and therefore more often used in practice. A different approach to logit models is given in Holmes and Held (2006). All MCMC algorithms presented in this paper are validated through simulation. The usefulness of these models has been demonstrated by the application to a mobility study. 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 in addition to PMCC and Brier score. 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, however their calibration is difficult in such complex settings (see Hjort and Steinbakk 2006).

The problem of how to include interaction effects between cluster and spatial effects has been considered in Prokopenko (2004). However the data set at hand is too sparse to support such model extensions. The mobility study also included information on trips conducted by foot and bicycle which have been ignored so far. A multinomial logit (MNL) analysis without spatial and cluster effect of this data has been performed by Ehrlich (2002). Therefore we plan to extend our analysis to MNL models with spatial and cluster effects. For point location data a MNL model with spatial effects based on spatial distances has been considered by Mohammadian and Kanaroglou (2003). However many discrete choice modelers have objected to the restrictions implied by a MNL model. In particular the MNL model assumes that the random utilities are independent identically distributed and that the responsiveness to attributes of alternatives across individuals after controlling for individual characteristics is homogenous. To relax these two restrictions the generalized extreme value (GEV) class of models and the mixed multinomial logit (MMNL) class have been proposed (see for example Bhat 2002 and Bhat 2006). Bhat and Guo (2004) consider a mixed spatially correlated logit model based on a GEV structure to accommodate correlations between spatial units of a location point referenced data. They use the Halton simulation method (see Train 2003) to simulate the corresponding likelihood for parameter estimation. It would be interesting to provide alternative Bayesian estimates for these discrete choice models for point location data. In addition one can develop models for spatially aggregated data allowing for a spatial CAR formulation. The addition of cluster effects would provide an alternative to the MMNL model class.

Acknowledgements

Our sincere thanks goes to Dipl.-Ing. Christoph Huß, BMW Group, Munich and Dr. Thomas Zängler, Lehrstuhl für Wirtschaftslehre des Haushalts, TU München for introducing us to the topic of mobility research and providing us with this interesting data set. Further we like to thank the Editor, the Associate Editor and the referees for the helpful suggestions on how to improve earlier versions of the manuscript.

References

Albert, J. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *J. Am. Statist. Ass. 88*, 669–679.

Anselin, L. (1988). *Spatial econometrics: Methods and Models*. Boston: Kluwer Academic Publishers.

Anselin, L. and R. J. G. M. Florax (1995). *New Directions in Spatial Econometrics*. Berlin: Springer.

Anselin, L., R. J. G. M. Florax, and S. Rey (2004). *Advances in Spatial Econometrics*. Berlin: Springer.

Banerjee, S., B. Carlin, and A. Gelfand (2004). *Hierarchical Modeling and Analyses for Spatial Data*. New York: Chapman & Hall/CRC.

Ben-Akiva, M. and S. R. Lerman (1985). *Discrete Choice Analysis: Theory and Applications to Travel Demand*. Cambridge, MA: MIT Press.

Bennett, J. E., A. Racine-Poon, and J. C. 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. and P. J. Green (1993). Spatial statistics and Bayesian computation (with discussion). *J. R. Statist. Soc. B,55*, 25–37.

Besag, J., P. J. Green, D. Higdon, and K. Mengersen (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science 10*, 3–66.

Bhat, C. R. (2002). Recent methodological advances relevant to activity and travel behavior. In *Perpetual Motion: Travel Behavoir Research Opportunities and Application Challenges (ed. by H.S. Mahmassani)*, pp. 381–414. Elsevier Science (Oxford).

Bhat, C. R. (2006). Flexible model structures for discrete choice analysis. In *Handbook of transport modelling (ed. by D. A. Hensher and K. J. Button)*. Elsevier (Amsterdam).

Bhat, C. R. and J. Guo (2004). A mixed spatially correlated logit model: formulation and application to residential choice modeling. *Transportation Research B,38*, 147–168.

Brier, G. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review 78*, 1–3.

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.

Cressie, N. (1993). *Spatial Statistics*. New York: John Wiley & Sons.

Ehrlich, K. (2002). *Regressionsverfahren mit nominaler Zielvariable, Diploma thesis*. Munich University of Technology: (http://www-m4.ma.tum.de/Diplarb/).

Fahrmeir, L. and T. Kneib (2007). Propriety of posteriors in structured additive regression models: Theory and empirical evidence. Technical Report Discussion Paper 510, SFB 386 Statistical Analysis of Discrete Structures (http://www.stat.uni-muenchen.de/sfb386/).

Fahrmeir, L., T. Kneib, and S. Lang (2004). Penalized structured additive regression: A Bayesian perspective. *Statistica Sinica 14*, 731–761.

Fahrmeir, L. and S. Lang (2001). Bayesian semiparametric regression analysis of multicategorical time-space data. *Annals of the Institute of Statistical Mathematics 53*, 10–30.

Fleming, M. M. (2004). Techniques for estimating spatially dependent discrete choice models. In *Advances in Spatial Econometrics (ed. by L. Anselin, R. J. G. M. Florax and S. J. Rey)*, pp. 145–168. Springer (Berlin).

Gelfand, A. and S. Ghosh (1998). A minimum posterior predictive loss approach. *Biometrika 85*, 1–11.

Gelfand, A. E. and F. M. Smith (1990). Sampling-based approaches to calculating marginal densities. *JASA 85*, 398–409.

Gelman, A., X. Meng, and H. Stern (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica 6*, 733–807.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. New York: Chapman & Hall.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction and estimation. *JASA 102*, 359–378.

Hjort, N. L.and Dahl, F. A. and G. Steinbakk (2006). Post-processing posterior predictive p-values. *JASA 101*, 1157–1174.

Holmes, C. and L. Held (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis 1*, 145–168.

Huffer, F. W. and H. Wu (1998). Markov Chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics 54*, 509–524.

Jin, X., B. Carlin, and S. Banerjee (2005). Generalized hierarchical multivariate CAR models for area data. *Biometrics 61*, 950–961.

Knorr-Held, L. and H. Rue (2002). On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics 29*, 597–614.

McFadden, D. (1974). The measurement of urban travel demand. *J. Public Economics 3*, 303–328.

McFadden, D. (1984). Econometric analysis of qualitative response models. In *Handbook of Econometrics Vol. 2 (edited by Z. Griliches and M. Intriligator)*. North Holland (Amsterdam).

McFadden, D. (2001). Disaggregate behavioral travle demand' rum side: A 30 years retrospective. In *Travel behavior Research: The leading edge (ed. by D.A. Hensher)*, pp. 17–64. Elsevier (Amsterdam).

McGillivray, R. G. (1972). Binary choice of urban transport mode in the San Francisco bay region. *Econometrica 40,5*, 827–848.

McMillen, D. P. (1992). Probit with spatial autocorrelation. *J. Regional Science 32, 3*, 335–348.

Mohammadian, A. and P. Kanaroglou (2003). Application of spatial multinomial logit model to transportation planning. In *Proceedings of the 10th International Conference on Travel Behaviour Research, Aug. 2003, Switzerland*. CD-ROM.

Pettitt, A. N., I. S. Weir, and A. G. 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.

Pinkse, J. and M. E. Slade (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Econometrics 85*, 125– 154.

Prokopenko, S. (2004). *Hierarchical binary spatial regression models with cluster effects, Ph.D. thesis*. Munich University of Technology: (http://mediatum2.ub.tum.de/).

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

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. V. D. Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B 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.

Sun, D., R. K. Tsutakawa, and Z. He (2001). Propriety of posteriors with improper priors in hierarchical linear mixed models. *Statistica Sinica 11*, 77–95.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press.

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference 121*, 311–324.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika 41*, 434–449.

Zängler, T. W. (2000). *Mikroanalyse des Mobilitätsverhaltens in Alltag und Freizeit*. Berlin: Springer Verlag.