Bayesian Spatial Modeling: Propriety and Applications to Small Area Estimation with Focus on the German Census 2011

Thesis

submitted in partial fulfillment of the requirements for the degree

Dr. rer. pol.

to the

Faculty IV, Economic and Social Statistics Department

at the University of Trier

submitted by

Diplom-Mathematiker Martin Vogt

Zellstraße 1

54290 Trier

born 14.07.1981 in Düsseldorf

Supervisor: Prof. Dr. Ralf Münnich
Co-supervisor: Prof. Partha Lahiri, PhD

Trier, 12.04.2010

Acknowledgment

I would like to express my deep and sincere gratitude to my supervisor, Professor Dr. Ralf Münnich, University of Trier. His invaluable support, encouraging and personal guidance have provided a good basis for the present thesis.

I am also deeply grateful to Professor Dr. Partha Lahiri, The Joint Program in Survey Methodology at the University of Maryland, for his detailed and constructive comments, and for his important support throughout the work.

I wish to express my thanks to the Studienstiftung des deutschen Volkes for financing the thesis, two research stays and the attendance of a conference.

My thanks are also to the Landesgraduiertenförderung Rheinland-Pfalz for financing the first six month of the dissertation.

I warmly thank Pablo Burgard, Tobias Enderle, Jan-Philipp Kolb, and Stefan Zins for helpful discussions about datasets and R programming.

I owe my loving thanks to my girlfriend Simone Bast. For her understanding, whenever I could not stop thinking about statistics, or stayed abroad for research.

My special gratitude is due to my brothers and parents for their loving support.

Contents

List of Figures VII

List of Tables IX

List of Abbreviations X

List of Symbols XI

1 Introduction 1

2 Basics and Fundamental Concepts 5

2.1 Fundamental Concepts of Small Area Estimation 5

2.1.1 The Basic Idea of Small Area Estimation 5

2.1.2 Standard Small Area Models 8

2.2 The Posterior Distribution and Improper Priors in Bayesian Analysis . . . 13

2.2.1 The Posterior Distribution 15

2.2.2 Improper Priors 17

2.3 The Impact of Impropriety on Markov Chain Monte Carlo Methods 19

2.3.1 The Effect of Impropriety on the Gibbs Sampler 19

2.3.2 Markov Chain Monte Carlo in R and WinBUGS 25

2.4 Spatial Modeling in Small Area Estimation 29

2.4.1 Spatial Data 29

2.4.2 Spatial Modeling: Conditional Autoregressive Model 30

2.4.3 Specifying the Geographic Weights Matrix 33

3 Propriety of the Posterior Distribution for Spatial Hierarchical Linear Mixed Models **37**

3.1 Introduction . . . 37

3.2 The Hobert and Casella Propriety Theorem . . . 38

3.3 The Spatial Adaption . . . 39

3.4 The Gibbs Conditionals . . . 43

3.5 Application . . . 45

3.6 Conclusion . . . 48

4 An Existence Theorem of the Posterior Distribution for a Class of Spatial Bernoulli Mixed Models **50**

4.1 Introduction . . . 50

4.2 The Natarajan and McCulloch Theorem . . . 52

4.3 The Spatial Existence Theorem . . . 53

4.4 Alternative Proof of the Existence Theorem . . . 59

4.5 Application . . . 59

4.6 Conclusion . . . 63

5 The Spatial Fay-Herriot Model: Propriety and Prediction **65**

5.1 Introduction . . . 65

5.2 Theory . . . 66

5.2.1 Spatial Hierarchical Extension of the Fay-Herriot Model . . . 66

5.2.2 The Predicted Mean of One Unsampled Area . . . 67

5.2.3 Propriety of the Posterior Distribution . . . 71

5.3 Application . . . 76

5.3.1 Setup . . . 76

5.3.2 Results . . . 77

5.4 Conclusion . . . 83

6 Simulation Study **84**

6.1 Introduction and Aim 84

6.2 Dataset Description 85

6.3 Simulation Setup 87

6.3.1 Models and Propriety 87

6.3.2 Setups 88

6.3.3 Measures 90

6.4 Results 91

6.4.1 The Standard Models 91

6.4.2 The Influence of the Sample Size 101

6.4.3 The Effect of Different Neighborhood Structures 105

6.4.4 The Influence of the Prior Distribution 110

6.4.5 The Impact of the Spatial Correlation Parameter 115

6.4.6 Dummy Variable for East Germany 118

6.5 Conclusion 123

7 Summary and Outlook **125**

A Gibbs Sampler for a Proper and an Improper Model **128**

B WinBUGS Models **129**

References **135**

List of Figures

1 Unemployment Rate 2007 in Germany at the Federal State Level; Data Source: Bundesagentur für Arbeit (2008). 2

2 Gamma Prior Distribution for $(\sigma_u^2)^{-1}$ with Parameters 0.5 and 0.0005. . . . 14

3 Trace Plot of the Simulated Values of x (Chain 1 : Red, Chain 2 : Blue) of the Gibbs Sampler for the Improper Model. 23

4 Trace Plot of the Simulated Values of x (Chain 1 : Red, Chain 2 : Blue) of the Gibbs Sampler for the Proper Model. 24

5 Different Neighborhood Structures for Federal States in Germany. 35

6 Nearest Neighbor Structure of 3 Areas in a Row. 35

7 Unemployment Rate 2007 in the Federal States in Germany and Deviance of the RRMSE for the Normal Models. 47

8 RRMSE for the Spatial and Non-spatial Normal Model. 48

9 Unemployment Rate 2007, RRMSE, and Deviance of the RRMSE for the Spatial and Non-spatial Bernoulli Model. 62

10 Deviance of the RRMSE for the Federal States, Sorted by Number of Neighbors (Ascending Order). 64

11 Predicted Values of the Spatial and Non-spatial FH Model Compared to the Official Estimates from 1993 : 4 and No Covariates. 80

12 Predicted Values of the Spatial and Non-spatial FH Model Compared to the Official Estimates from 1993 : Covariates $x3$, $x4$. 81

13 Predicted Values of the Spatial and Non-spatial FH Model Compared to the Official Estimates from 1993 : Covariates $x1$, $x2$. 82

14 Squared Deviance of the Spatial and Non-spatial FH Model for 4 (Upper Plots) and No (Lower Plots) Covariates. 83

15 Professional Training or University/College Degree? 86

16 Boxplots of the RRMSE for the Standard Models. 91

17 RRMSE for the 16 Federal States of the Spatial Against the Non-spatial BHF and Bern Model. 92

18 RRMSE for the 16 Federal States of the Spatial Against the Non-spatial FH Model. 93

19 Deviance of the RRMSE for the Standard BHF and Bernoulli Models. . . . 97

20 Deviance of the RRMSE for the Standard Fay-Herriot Model. 97

21 Boxplots of the Relative Bias for the 6 Standard Models. 98

22 Relative Bias for the Spatial and Non-spatial BHF Models on the Map. . . 99

23 Relative Bias for the Spatial and Non-spatial Bern Models on the Map. . . 100

24 Relative Bias for the Spatial and Non-spatial FH Models on the Map. . . . 100

25 Boxplots of the Relative Dispersion for the 6 Standard Models. 101

26 Boxplots of the RRMSE for the Spatial and Non-spatial BHF Models with Varying Sampling Sizes. 102

27 Boxplots of the RRMSE for the Spatial and Non-spatial Bernoulli Models with Varying Sampling Sizes. 102

28 Boxplots of the RRMSE for the Spatial and Non-spatial FH Models with Varying Sampling Sizes. 103

29 Boxplots of the Relative Bias for the Spatial and Non-spatial Models with Varying Sampling Sizes. 104

30 Boxplots of the Relative Dispersion for the Spatial and Non-spatial Models with Varying Sampling Sizes. 104

31 Boxplots of the RRMSE for the Standard and City Structure Models. . . . 105

32 Deviance of the RRMSE for the Nearest Neighbor and City Structure BHF and Bern Models. 109

33 Deviance of the RRMSE for the Nearest Neighbor and City Structure FH Models. 109

34 Relative Bias of the BHF, Bern, and FH Models for the City Structure. . . 110

35 Relative Dispersion of the BHF, Bern, and FH Models for the City Structure. 110

36 Boxplots of the RRMSE of the BHF Models for Different Prior Distributions. 111

37 Boxplots of the RRMSE of the Bern Models for Different Prior Distributions. 111

38 Boxplots of the Relative Bias for Different Prior Distributions. 112

39 Boxplots of the Relative Dispersion for Different Prior Distributions. 112

40 RRMSE for the BHF and Spatial BHF Models with $a = -0.9$ and Varying b. 113

41 RBias and RDisp for the BHF and Spatial BHF Models with $a = -0.9$ and Varying b. 113

42 RRMSE for the BHF and Spatial BHF Models with $b = -0.9$ and Varying a. 114

43 RBias and RDisp for the BHF and Spatial BHF Models with $b = -0.9$ and Varying a. 114

44 RRMSE for BHF Models with Varying Spatial Correlation Parameter. . . 115

45 RRMSE for Bern Models with Varying Spatial Correlation Parameter. . . . 115

46 Maximum of the RRMSE for BHF and Bernoulli Models with Varying Spatial Correlation Parameter. 116

47 Relative Bias of the BHF and Bern Models with Varying Spatial Correlation Parameter. 117

48 RDisp for BHF and Bern Models with Varying Spatial Correlation Parameter. 117

49 Boxplot of the RRMSE for the BHF and Bern Model with Dummy Variable East and Spatial Versions. 118

50 Boxplot of the RRMSE for the FH Model with Dummy Variable East and Spatial Versions. 119

51 Deviance of the RRMSE for the BHF Model with Dummy Variable East and Spatial Versions. 120

52 Deviance of the RRMSE for the Bern Model with Dummy Variable East and Spatial Versions. 120

53 Deviance of the RRMSE for the FH Model with Dummy Variable East and Spatial Versions. 121

54 RBias for the Models with Dummy Variable East and the Spatial Versions. 122

55 RDisp for the Models with Dummy Variable East and the Spatial Versions. 122

56 RBias for the Spatial and Non-spatial BHF Model with Dummy Variable on the map. 123

List of Tables

1 Unemployment Rate for the 16 German Federal States in Percent, as well as the Direct and Shrinkage Estimator; Data Source: Bundesagentur für Arbeit (2008). . . . 7

2 Number of Iterations and Values of the Scale Reduction Factor for the Improper Model. . . . 22

3 Number of Iterations and Values of the Scale Reduction Factor for the Proper Model. . . . 23

4 Unemployment Rate for the 16 German Federal States, as well as Bias, RRMSE, and Difference of the RRMSE for the BHF Models; Data Source: Bundesagentur für Arbeit (2008). . . . 46

5 Unemployment Rate for the 16 German Federal States, as well as Bias, RRMSE, and Difference of the RRMSE for the Bernoulli Models; Data Source: Bundesagentur für Arbeit (2008). . . . 60

6 RRMSE and Bias for the Spatial and Non-spatial Normal and Bernoulli models. . . . 63

7 SAIPE Simulation Results for 1993. . . . 78

8 SAIPE Simulation Results for 1989. . . . 79

9 Size of the Transformed Mikrocensus Dataset per German Federal State. . . . 86

10 Maximum of the RRMSE of the BHF, Bernoulli, and FH Standard Models. 92

11 Professional Training or University/College Degree, as well as RBias, RRMSE, and Difference of the RRMSE for the BHF Models; Data Source: German Mikrocensus (2002). . . . 94

12 Professional Training or University/College Degree, as well as RBias, RRMSE, and Difference of the RRMSE for the Bern Models; Data Source: German Mikrocensus (2002). . . . 95

13 Professional Training or University/College Degree, as well as Bias, RRMSE, and Difference of the RRMSE for the FH Models; Data Source: German Mikrocensus (2002). . . . 96

14 Mean and Maximum of the RRMSE for the BHF, Bernoulli, and FH Models under Varying Sample Sizes. . . . 103

15 Mean and Maximum of the RRMSE for the Spatial BHF, Bernoulli, and FH Models Under Nearest Neighbor (NN) and City Structure (City). . . . 106

16 RRMSE for the Spatial BHF and Bernoulli Models Under Nearest Neighbor (NN) and City Structure (City). 107

17 RRMSE for the Spatial FH Models Under Nearest Neighbor (NN) and City Structure (City) 108

List of Abbreviations

Bern	Bernoulli
BHF	Battese-Harter-Fuller
CAR	Conditional autoregressive
CPS	Current Population Survey
Exp	Exponential distribution
FH	Fay-Herriot
GREG	Generalized regression estimator
ind	Independent distributed
i.i.d.	Independent and identically-distributed
IG	Inverse gamma distribution
IRS	Internal Revenue Service
MCMC	Markov chain Monte Carlo
MC	Mikrocensus
RBias	Relative Bias
RDisp	Relative Dispersion
RRMSE	Relative Root Mean-Squared Error
S	Sampled
SAR	Simultaneous autoregressive
SAIPE	Small Area Income and Poverty Estimates
SSE	Sum of squared errors
UMACS	Universal Markov chain sampler (R package)
US	Unsampled

List of Symbols

$\mathbf{A_i}$	Set of neighboring areas of area i
β	Regression coefficient
$\mathbf{D}$	Index set defined in Definition 2.19
E	Mean value
ϵ	Sampling error term
$\mathbf{f}(\cdot)$	Density
$\mathbf{f_{X\|Y=y}}(\cdot)$	Conditional distribution of X given $Y = y$
$\mathbf{f_{X,Y}}(\cdot)$	Joint density of X and Y
$\mathbf{k}$	Number of areas
$\mathbf{K}$	Constant
$\mathbf{M}$	$M = \text{diag}(\sigma^2_{u,i})$ in a spatial model
$\mathbf{n}$, $\mathbf{n_i}$	Number of observations, number of observations in area i
$(\mathbf{\Omega}, \mathbf{S}, \mathbf{P})$	Abstract probability space
$\mathbf{p}$	Spatial correlation parameter
$\oplus$	Direct sum
$\varphi_\mathbf{1}$ and $\varphi_\mathbf{2}$	Polyhedral cones of Theorems 4.1 and 4.3
$\pi(\cdot)$	Density for hyperparameter
$\mathbf{\Pi_1}$ and $\mathbf{\Pi_2}$	Conditions for the propriety in Theorems 4.1 and 4.3
$\mathbf{Q}$	Neighborhood matrix in a spatial model
$\mathbf{\tilde{Q}}$	Row standardized neighborhood matrix in a spatial model
$\theta^\mathbf{o}$	Overall sample mean in Example 2.1
$\mathbb{R}$	Real numbers
$\mathbf{\hat{R}}$	Scale reduction factor

$\mathbf{S^2}$	Squared deviance of the estimator to the true value
$\sigma^2_{\epsilon,\mathbf{i}}$	Variance of the error term of area i
$\sigma^2_{\mathbf{u},\mathbf{i}}$	Variance of the random effect of area i
$\mathbf{\Sigma_\epsilon}$	Matrix of variance of the error term
$\mathbf{\Sigma_u}$	Matrix of variance of the random term
$\mathbf{u}$	Vector of the random effects
$\mathbf{W}$	Weighting matrix in the CAR model
$\mathbf{X}$	Matrix of the covariates
$\mathbf{X}^*$, $\mathbf{x}^*_\mathbf{i}$	Defined below Formula (4.3)
$\mathbf{Y}$	Dependent variable (vector)
$\mathbf{Z}$	Known matrix for the random effects
$\mathbf{Z}^*$, $\mathbf{z}^*_\mathbf{i}$	Defined below Formula (4.3)
$\mathbf{Z}(\mathbf{s})$	Random quantity defined in Definition 2.19

1 Introduction

For the first time, the German Census 2011 will be conducted via a new method the register based census. In contrast to a traditional census, where all inhabitants are surveyed, the German government will mainly attempt to count individuals using population registers of administrative authorities, such as the municipalities and the Federal Employment Agency. Census data that cannot be collected from the registers, such as information on education, training, and occupation, will be collected by an interview-based sample survey. Moreover, the new method reduces citizens' obligations to provide information and helps reduce costs significantly (see Website: Census, 2010a).

The census provides information as a basis for a number of political and economic decisions. For example, the revenue equalization between the Länder and between the Länder and its municipalities is calculated on the basis of population figures (see Website: Census, 2010b). Further, the data supports future leaders in answering questions such as: "How many schools and kindergartens does a city need?". The survey also has the task to examine the quality of the data from the registers (see Website: Census, 2010d).

The use of sample surveys is limited if results with a detailed regional or subject-matter breakdown have to be prepared. Classical estimation methods are sometimes criticized, since estimation is often problematic for small samples (see Website: Census, 2010c). Fortunately, model based small area estimators (cf. Rao, 2003 and Jiang and Lahiri, 2006) serve as an alternative. These methods help to increase the information, and hence the effective sample size. For example, we can combine information from similar structured areas or by including auxiliary information out of the administrative registers, in essence borrowing strength from neighboring areas. (cf. Jiang and Lahiri, 2006, pp. 3).[1]

In the German Census 2011 it is possible to embed areas on a map in a geographical context. This may offer additional information, such as neighborhood relations or spatial interactions. Figure 1 shows the unemployment rate in Germany at the federal state

[1]Further information about the German Census 2011 may be found among others in Magg et al. (2006), Münnich et al. (2007) and on the official website www.zensus2011.de.

level in 2007. The map provides a visual image of spatial interaction, showing that the unemployment rate increases from west to east and from south to north. Standard small area models, like Fay-Herriot (cf. Fay and Herriot, 1979) or Battese-Harter-Fuller (cf. Battese et al., 1988), do not account for such interactions explicitly. The aim of our work is to extend the classical models by integrating the spatial information explicitly into the model. In addition, the possible gain in efficiency will be analyzed.

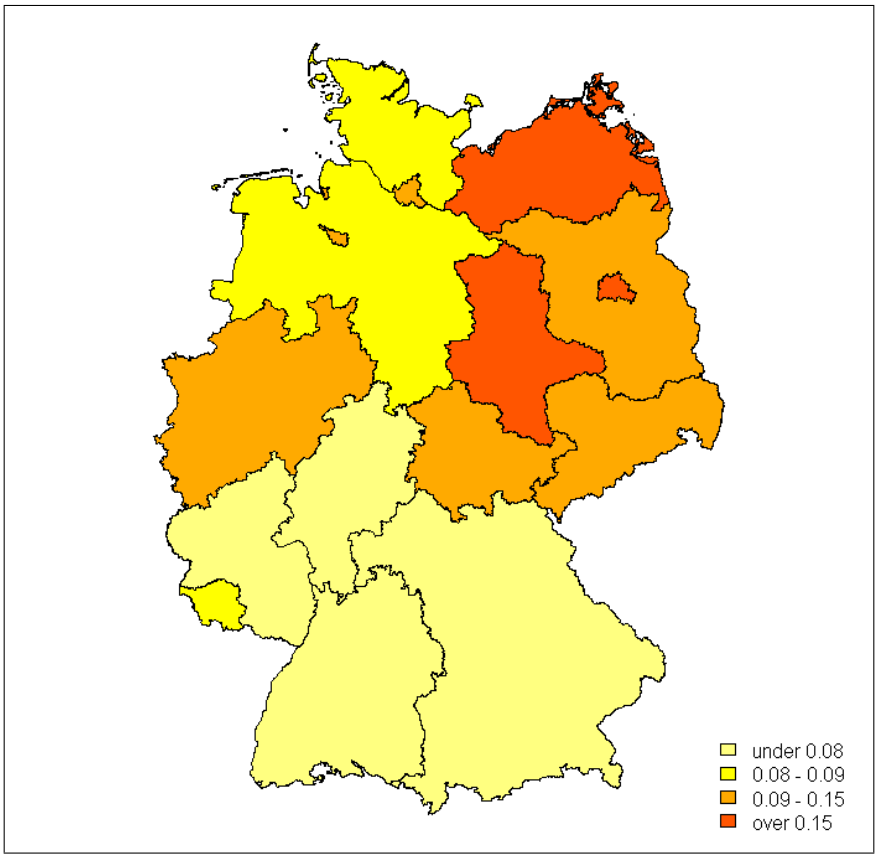

Figure 1: Unemployment Rate 2007 in Germany at the Federal State Level; Data Source: Bundesagentur für Arbeit (2008).

We consider three model classes, the Gaussian general linear mixed model, the Bernoulli mixed model, and the Fay-Herriot model, which all allow the modeling of many problems. For all of these classes we show how to include spatial information via a Bayesian method, where the complete inference is based on the posterior distribution. Bayesian methods often assume improper prior for the hyperparameters, which could result in an improper posterior distribution. This can cause misleading results (cf. Hobert and Casella,

1996). Therefore, in the present work the propriety of the posterior distribution under certain assumptions for the three model classes is shown. The existence theorems build the theoretical basis for applying the proposed models.

Possible gains and limitations of the derived models, with special focus on applications to the German Census 2011 are tested by means of a simulation study. Since German census data is not available, a data set with similar structure and variables is chosen: the public use file of the German Mikrocensus (see German Mikrocensus, 2002).

The organization of the remaining chapters follows accordingly:

In Chapter 2, a review is done of technical background in small area estimation, Bayesian modeling, Markov chain Monte Carlo, and spatial modeling. Standard small area models (Fay-Herriot and Battese-Harter-Fuller) are introduced (Section 2.1) and we show how to extend them to hierarchical small area models in Section 2.2. In addition, improper prior distributions and possible resulting impropriety of the posterior distribution are discussed. Because the posterior distributions of hierarchical models are often intractable, Markov chain Monte Carlo Methods (MCMC) are often employed. These methods are introduced in Section 2.3. Furthermore, standard small area models and their hierarchical extensions assume independent random effects. This assumption is changed to allow for spatial correlations in Section 2.4 and we explain how to model those correlations.

In Chapter 3, the propriety of the posterior distribution is proved for the spatial general linear mixed model with homoscedastic sampling variances, which includes the spatial hierarchical extension of the Battese-Harter-Fuller model. Moreover, the Gibbs conditionals, necessary for the implementation of the Gibbs sampler, are derived. Benefits of this method are shown via an application to unemployment data at the federal state level in Germany.

In Chapter 4, normality is not assumed and the propriety is proved for the posterior of the spatial Bernoulli mixed model. Then the model is applied to the unemployment data and compared to the according non-spatial version and the normal model of Chapter 3.

The model class of Chapter 3 covers a wide variety of area and unit-level spatial small area models. However, one popular area-level model is not included the Fay-Herriot model. Finally in Chapter 5, the Fay-Herriot model is extended to a spatial hierarchical Fay-Herriot model and the propriety of the posterior distribution is proved. Afterward, the model is applied to Small Area Income and Poverty Estimates (SAIPE) of the U.S. Census Bureau. In this setting one area is assumed to be unsampled, and the area mean of this area is predicted via a spatial and non-spatial model.

The model classes of the previous sections allow modeling many variables of the German Census 2011. In Chapter 6, gains and limitations of area and unit-level models out of these classes are tested via simulation study. Various parameter combinations have been adopted, allowing investigation of a number of model configurations.

The results are summarized in Chapter 7.

Appendix A illustrates the computer code for a situation where the posterior distribution is improper and the Gibbs sampler leads to misleading results. Appendix B includes the WinBUGS models used in this work.

2 Basics and Fundamental Concepts

In this chapter, fundamentals of small area estimation, Bayesian modeling, Markov chain Monte Carlo methods, and spatial modeling will be explained.

2.1 Fundamental Concepts of Small Area Estimation

In the following the term small area estimation is defined and the underlying idea is explained by means of an example using unemployment data. Afterward, two basic small area models are introduced: the area-level Fay-Herriot model (cf. Fay and Herriot, 1979) and the Battese-Harter-Fuller unit-level model (cf. Battese et al., 1988).

2.1.1 The Basic Idea of Small Area Estimation

Central to the concept of small area estimation is the definition of small areas. Rao, 2003, p. xxi defines a **small area** as any subpopulation for which direct estimates of adequate precision cannot be produced. Similarly, Jiang and Lahiri, 2006, p. 1 note that the term small area typically refers to a population for which reliable statistics of interest cannot be produced due to certain limitations of the available data. Thus, an area is considered to be small if reliable statistics of adequate precision cannot be produced. However, there does not exist a cut off for deciding whether or not an area is too small (cf. Münnich and Schmidt, 2002). Small area estimation refers to techniques with the intension to overcome the problem of the small sample size. This is done by combining information from similar structured areas and thus to extend the effective sample size. In the following example a simple shrinkage small area estimator (cf. James and Stein, 1960) is derived.

Example 2.1: The second column of Table 1 on page 7 shows the unemployment rate of the 16 German federal states from 2007. In this example this data is seen as the underlying gold standard which shall be estimated. To do this for every state a sample (simple random sampling) of size 20 is drawn out of a Bernoulli distribution with probability equal to the corresponding gold standard. The sample mean for every state is given in column 3.

A direct and a small area estimator will be derived. The estimators will then be compared by the sum of the quadratic difference between the corresponding estimator and the gold standard value, θ_i, where the quadratic difference is defined to be:

$$S^2 = \sum_{i=1}^{16} (\text{Estimator}_i - \theta_i)^2. \tag{2.1}$$

Since in this setting the samples are independent of each other and no covariates are available the direct estimator is equal to the sample mean given in column 3 of Table 1. In this case $S^2 = 461.46$.

Next, an alternative method is presented. Using the central limit theorem, assuming that the approximation is valid, the sum of unemployed (sample size 20) in the i^{th} area, Z_i, is approximately normally distributed according to:

$$Z_i \sim N(20\theta_i, 20\theta_i(1-\theta_i)),$$

where θ_i is the gold standard unemployment rate. The unemployment rate for a sample size of 20 in the i^{th} area $Y_i = Z_i/20$ is then also normally distributed:

$$Y_i \sim N\left(\theta_i, \frac{\theta_i(1-\theta_i)}{20}\right),$$

with mean θ_i and variance $\dfrac{\theta_i(1-\theta_i)}{20} = \sigma^2_{\epsilon,i}$. Therefore, the observed sample y_i (column 3 of Table 1) is a realization out of a normal distribution. If the sample size is small, the direct estimator will be unreliable. One idea to stabilize the estimator might be to take the mean of all sample values and combine it with the sample value of one state. The direct estimator moves toward the overall sample mean of all federal states, $\theta^o = \frac{1}{16}\sum_{i=1}^{16} y_i$. The larger the variance of Y_i the less reliable is the direct estimator, and thus, the weight of the overall mean should increase. Since the variance of Y_i is unknown an estimator $\hat{\sigma}^2_{\epsilon,i}$ will be used: $\hat{\sigma}^2_{\epsilon,i} = \dfrac{y_i(1-y_i)}{20}$. Combining the above idea with the variance estimator $\hat{\sigma}^2_{\epsilon,i}$ yields a new estimator:

$$\hat{\theta}_i = \theta^o + \left(1 - \frac{\hat{\sigma}^2_{\epsilon,i}}{\hat{\sigma}^2_u + \hat{\sigma}^2_{\epsilon,i}}\right)(y_i - \theta^o), \qquad i = 1, \cdots, 16, \tag{2.2}$$

where $\hat{\sigma}^2_u = \dfrac{\theta^o(1-\theta^o)}{20}$ is an estimator for the variance of the overall mean. Formula (2.2) may be interpreted as follows: The variance $\hat{\sigma}^2_u$ represents the uncertainty of the overall

State	Gold standard	Sample mean	Shrinkage estimator
Berlin	15.5	15	13.1
Bremen	12.7	20	13.9
Brandenburg	14.9	10	10.8
Baden-Württ.	4.9	10	10.8
Bavaria	5.3	15	13.1
Hamburg	9.2	10	10.8
Hesse	7.6	5	5.9
Mecklenburg-Vorp.	16.5	15	13.1
Lower Saxony	8.9	10	10.8
North Rhine Westph.	9.5	15	13.1
Rhineland-Palatinate	6.5	10	10.8
Saxony-Anhalt	16	15	13.1
Saxony	14.7	20	13.9
Schleswig-Holstein	8.4	15	13.1
Saarland	8.4	0	0
Thuringia	13.2	5	5.9

Table 1: Unemployment Rate for the 16 German Federal States in Percent, as well as the Direct and Shrinkage Estimator; Data Source: Bundesagentur für Arbeit (2008).

mean. If $\hat{\sigma}_u^2$ is large, then the overall mean is not reliable and thus $\hat{\theta}_i$ is close to the direct estimator y_i. If the sample variance $\hat{\sigma}_{\epsilon,i}$ is large, then more weight is given to the overall mean. The estimated values for the unemployment data are given in column 4 of Table 1. The squared deviance (2.1) for this estimator is $S^2 = 325.28$. Thus, S^2 is considerably larger when using the direct estimate compared to the estimator in (2.2).

The estimator in (2.2) is a simple form of a small area estimator, also known as shrinkage estimator (cf. James and Stein, 1960). In Vogt (2007), the shrinkage estimator is explained and derived in more detail.

In Example 2.1 a direct estimator using information of one federal state and a shrinkage estimator which combines information of all states were presented. This technique helped to reduce the sum of the squared deviance. In other words, because the sample size was too small to obtain reliable direct estimates strength was borrowed (Ghosh and Rao, 1994, p. 55) from other states. In the example, the sum of the squared deviance for the small area estimator is smaller than that for the direct estimator. On the contrary, the squared deviance might not be smaller for each of the states. Row 1 of Table 1 shows the data for Berlin with a relatively high unemployment rate. In this case the direct estimator is closer to the gold standard value than the small area estimator. This is due to the fact, that the overall mean is not representative for this state. Therefore, in the next section, extensions of the above derived small area estimator are presented, making use of covariates. The connection to the shrinkage estimator is visible in Example 2.12.

2.1.2 Standard Small Area Models

In this section the two basic small area models, the Fay and Herriot (1979) and the Battese et al. (1988) model, are introduced using the German Census 2011 and unemployment data at the federal state level in Germany. Both model types are currently analyzed for a possible application within the German census (cf. Münnich et al., 2007). Extensions of these models play a vital role throughout the work.

Example 2.2: We briefly introduce the Fay-Herriot model. A detailed derivation can be found in Vogt (2007). However before doing so, as explained in Chapter 1, the German Census 2011 will be conducted using a new method. Rather than a total enumeration, a sample is drawn. Unfortunately, the sample size might be too small to obtain reliable results with a detailed regional or subject-matter breakdown with classical estimation methods (cf. Website: Census, 2010c). A way around this problem may be to impose model assumptions.

To explain what is meant by model assumptions, imagine the following situation: Assume that the unemployment rate at the federal state level in Germany is of interest (Figure 1 on page 2).[2] We assume that Y_i, the proportion of people unemployed in state i, is normally distributed with unknown mean θ_i and variance component $\sigma^2_{\epsilon,i}$. The parameter of interest θ_i represents the true proportion of unemployed. Thus, our model becomes:

$$Y_i \overset{\text{ind}}{\sim} N(\theta_i, \sigma^2_{\epsilon,i}), \qquad i = 1, \cdots, k.$$

The sample is a realization of this distribution. If the sample size is too small to obtain directly reliable estimates of θ_i, the idea is to employ prior information on the mean θ_i. In the German Census 2011 prior information out of the administrative registers may be used, by means of covariates (see Website: Census, 2010c). To include prior information, a second model assumption is made, by specifying a distribution for θ_i as well, where:

$$\theta_i \overset{\text{ind}}{\sim} N(X_i\beta, \sigma^2_u), \qquad i = 1, \cdots, k.$$

This means that θ_i is normally distributed around a regression coefficient, representing information out of the registers. Thus, the complete model is:

$$\begin{aligned} Y_i &\overset{\text{ind}}{\sim} N(\theta_i, \sigma^2_{\epsilon,i}) \\ \theta_i &\overset{\text{ind}}{\sim} N(X_i\beta, \sigma^2_u), \qquad i = 1, \cdots, k, \end{aligned} \tag{2.3}$$

which can be transformed into:

$$\begin{aligned} Y_i &\overset{\text{ind}}{\sim} N(\theta_i, \sigma^2_{\epsilon,i}) \\ \theta_i &= X_i\beta + u_i \\ u_i &\overset{\text{i.i.d.}}{\sim} N(0, \sigma^2_u), \qquad i = 1, \cdots, k, \end{aligned} \tag{2.4}$$

where ind stands for independently distributed and i.i.d. for independently and identically distributed.

[2] In the German Census 2011 the sample size at the federal state level will be large enough. Problems may arise on smaller areas, like small municipalities with less than 10,000 inhabitants (cf. Website: Census, 2010c).

The model may be further transformed into (cf. Jiang and Lahiri, 2006, p. 6):

$$\begin{aligned} Y_i &= X_i\beta + u_i + \epsilon_i \\ u_i &\overset{\text{i.i.d.}}{\sim} N(0, \sigma_u^2) \\ \epsilon_i &\overset{\text{ind}}{\sim} N(0, \sigma_{\epsilon,i}^2), \quad i = 1, \cdots, k. \end{aligned} \tag{2.5}$$

Definition 2.3: The models (2.3), (2.4), and (2.5) are defined as **Fay-Herriot (FH) model** (cf. Fay and Herriot, 1979).

In the specification (2.5) the Fay-Herriot model can be viewed as a mixed regression model. The factor is called random effect and u incorporates additional variation between the areas not accounted for by the regression component. In Example 2.12 it is shown how to obtain estimates under this model.

In the next example the Battese-Harter-Fuller model will be introduced.

Example 2.4: In Example 2.2, for every small area one sample value Y_i is used. This means that data enters the modeling process in an aggregated way, namely at the federal state (area) level. Therefore, this type of model is called an area-level model. In contrast, the Battese-Harter-Fuller model is a unit-level model, meaning that the data is at the individual level. For example in area i, n_i individuals may be sampled. The model may then be defined as follows:

Definition 2.5: The **Battese-Harter-Fuller (BHF) model** is defined as (cf. Battese et al., 1988):

$$\begin{aligned} Y_{ij} &= X_{ij}\beta + u_i + \epsilon_{ij} \\ u_i &\overset{\text{i.i.d.}}{\sim} N(0, \sigma_u^2) \\ \epsilon_{ij} &\overset{\text{i.i.d.}}{\sim} N(0, \sigma_\epsilon^2), \quad i = 1, \cdots, k,\ j = 1, \cdots, n_i, \end{aligned} \tag{2.6}$$

where we denote the area by index i the individual by index j.

Like the Fay-Herriot model, the Battese-Harter-Fuller model consists of three components, namely a regression term $X_{ij}\beta$, a random effect u_i and errors ϵ_{ij}. The Battese-Harter-Fuller

model is a particular case of a general model, known as general linear mixed model. This model is quite flexible and allows for unit and area-level specification. It is defined below.

Definition 2.6: The **general linear mixed model** is defined as (cf. Hobert and Casella, 1996):

$$\begin{aligned} Y &= X\beta + Zu + \epsilon \\ \epsilon &\sim N(0, \Sigma_\epsilon) \\ u &\sim N(0, \Sigma_u), \end{aligned} \tag{2.7}$$

where Y is $n \times 1$, X is $n \times p$, β is $p \times 1$, ϵ is $n \times 1$. Further, $\Sigma_\epsilon = I_n \sigma_\epsilon^2$, Z is $n \times q$, $u = (u_1, \cdots, u_r)$ is $q \times 1$, where u_i is $q_i \times 1$ and $\sum_{i=1}^r q_i = q$. Finally, $\Sigma_u = \oplus_{i=1}^r I_{q_i} \sigma_{u,i}^2$, where the direct sum $\oplus$ of two matrices A, B is defined as $A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$.

Notes 2.7:

1. Jiang and Lahiri, 2006, pp. 8, present a specification of model (2.7), where the Fay-Herriot model (2.5) is included. In this work (Chapter 3), a theorem of Hobert and Casella (1996) is extended and their specification of model (2.7) is followed.

2. In Chapter 3 the existence of the posterior distribution for a spatial adaption of model (2.7) is proved. In the spatial model, only one random effect is considered, and thus, the dimension of u is equal to the number of areas k.

The theory and applications of mixed models is explained in Demidenko (2004) and Faraway (2006) in more detail.

Example 2.8: The Battese-Harter-Fuller model (2.6) may be written in the form (2.7):

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{kn_k} \end{pmatrix} = \begin{bmatrix} 1 & x_{111} & \cdots & x_{p11} \\ 1 & x_{112} & \cdots & x_{p12} \\ \vdots & \vdots & \vdots & \\ 1 & x_{1kn_k} & \cdots & x_{pkn_k} \end{bmatrix} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \oplus_{i=1}^k I_{n_i} \cdot \begin{pmatrix} u_1 \\ \vdots \\ u_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

In this section, we have introduced the ideas behind small area estimation as well as the standard small area models. A comprehensive introduction can be found in Rao (2003). Other books explaining the concepts and ideas of small area estimation are written by Mukhopadhyay (1998) (Small Area Estimation in Survey Sampling) and Longford (2005) (Missing Data and Small-Area Estimation). Some review papers involving small area estimation include Rao (1986), Rao (1999), Rao (2001), Ghosh and Rao (1994), Marker (1999), Pfeffermann (2002), Lahiri and Meza (2002), and Jiang and Lahiri (2006). Finally, the Eurarea (2004) project report describes various small area estimation methods.

2.2 The Posterior Distribution and Improper Priors in Bayesian Analysis

The idea of the basic small area models of Section 2.1.2 is to specify a distribution for the data, as well as for the unknown variables of interest θ_i. The distribution on θ_i represents the idea that there exists extra information about the parameter. Other parameters are included in the Fay-Herriot model (2.3) due to its hierarchical structure such as the regression coefficient β and the variance component σ_u^2. If prior information on these parameters is available we can include this in the model by specifying distributions on the unknown parameters. In a Bayesian framework these beliefs can be represented by a prior distribution, which is placed on the unknown parameter. For the Fay-Herriot model (2.3) such priors are given in the next example.

Example 2.9: The Fay-Herriot model (2.3) may be extended to a hierarchical Fay-Herriot model, by assuming prior distributions for β and σ_u^2.

Following Hobert and Casella (1996) this might be:

$$\begin{aligned} \pi_\beta(\beta) &\propto 1 \\ \pi_{\sigma_u^2}(\sigma_u^2 \mid b) &\propto (\sigma_u^2)^{-(b+1)}, \end{aligned} \tag{2.8}$$

where b is a constant. Another option for σ_u^2 is an inverse gamma distribution (cf. Sun et al., 1999):

$$1/\sigma_u^2 \sim \Gamma(0.5, 0.0005). \tag{2.9}$$

Figure 2 shows the prior distribution (2.9). This distribution takes positive values and states the belief that the random effects standard deviation is centered around 0.05 with a 1% prior probability of being smaller than 0.01 or larger than 2.5 (cf. GeoBUGS User Manual, 2004).

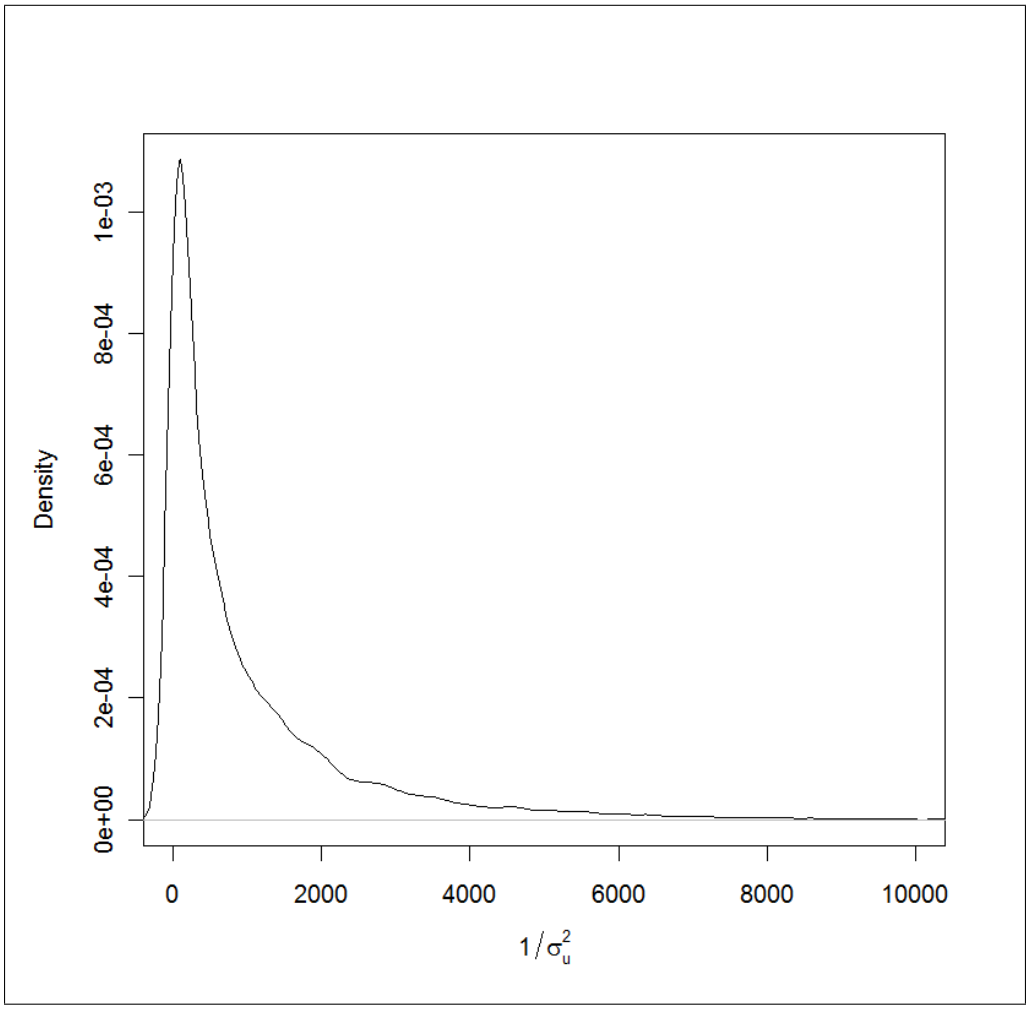

Figure 2: Gamma Prior Distribution for $(\sigma_u^2)^{-1}$ with Parameters 0.5 and 0.0005.

It is not easy to select appropriate prior distributions (cf. Kass and Wassermann, 1996). Throughout the present work standard priors distributions are taken and the question of how to choose prior distributions is not considered.

The representation of information, as well as the uncertainty about parameters as probabilities is central to Bayesian inference, and this concept is explained in more detail in various texts. For instance, introductions in Bayesian methodology and computation include Ghosh et al. (2006) and Albert (2007). Classical Bayesian books to consider are Box and Tiao (1973) and Berger (1985). Congdon (2003), Congdon (2005), and Congdon (2007) deal with various topics in Bayesian statistics, like Bayesian modeling and Bayesian categorical data analysis. Carlin and Louis (2000) cover Bayes and Em-

pirical Bayes methods for data analysis. Bernardo and Smith (1994) explain Bayesian theory from a mathematical perspective. Finally, the books Gelman et al. (1995) and Gelman and Hill (2007) are about Bayesian data analysis and multilevel models.

In the next section, we show how to utilize Bayesian methods for inference, via the posterior distribution.

2.2.1 The Posterior Distribution

If all the uncertainty is included in the model it is possible to combine prior and sample information and thus, perform an update via the posterior distribution. The complete inference is then based on this distribution. This concept is introduced more formally in Witting, 1985, pp. 128.

Definition 2.10: Let X and Y be real-valued random variables defined on an abstract probability space (Ω, S, P) with joint density:

$$f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}.$$

The **posterior density (conditional density function)** of the **posterior distribution** of X given $Y = y$ is defined by:

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{\int_{\mathbb{R}} f_{X,Y}(\xi, y)d\xi}, \quad x \in \mathbb{R}$$

for all y as long as the denominator is defined.

Theorem 2.11: Let X and Y be random variables as those in Definition 2.10, and furthermore, let $\tilde{f}$ denote the probability function of X. The conditional density function of the distribution of X given $Y = y$ is given by:

$$f_{X|Y=y}(x) = \frac{f_{Y|X=x}(y)\tilde{f}(x)}{\int\limits_{-\infty}^{\infty} f_{Y|X=x}(y)\tilde{f}(x)dx}. \tag{2.10}$$

The proof can be found in Witting, 1985, pp. 128.

The posterior density in the form (2.10) states the principle that updated knowledge combines prior knowledge with the data at hand:

$$\text{posterior density} = \frac{\text{likelihood} \times \text{prior density}}{(\int \text{likelihood} \times \text{prior density})},$$

where the denominator is a fixed normalizing factor which ensures that the posterior probabilities sum to 1 (cf. Congdon, 2007, p. 3). We can then show:

$$\text{posterior density} \propto \text{likelihood} \times \text{prior density}.$$

Example 2.12: The posterior distribution of θ_i of the standard Fay-Herriot model (2.3) is given by (see for example Vogt, 2007, p. 52):

$$\theta_i|Y_i \sim N\left(X_i\beta + \left(1 - \frac{\sigma^2_{\epsilon,i}}{\sigma^2_u + \sigma^2_{\epsilon,i}}\right)(Y_i - X_i\beta), \frac{\sigma^2_{\epsilon,i}\sigma^2_u}{\sigma^2_{\epsilon,i} + \sigma^2_u}\right), \quad i = 1, \cdots, k.$$

The posterior mean is a weighted average of the prior mean $X_i\beta$ and the data Y_i. Note that the posterior mean is an extension of the shrinkage estimator derived in Example 2.1.

The following hierarchical version of the general linear mixed model (2.7) is given by Hobert and Casella, 1996, p. 1463:

$$\begin{aligned} y \mid u, \sigma^2_\epsilon, \beta &\sim N_n(X\beta + Zu, I\sigma^2_\epsilon) \\ \pi_\beta(\beta) &\propto 1 \\ u \mid \sigma^2_{u,1}, \cdots, \sigma^2_{u,r} &\sim N_q(0, \Sigma_u) \\ \pi_\epsilon(\sigma^2_\epsilon \mid b) &\propto (\sigma^2_\epsilon)^{-(b+1)} \\ \pi_{\sigma^2_{u,i}}(\sigma^2_{u,i} \mid a_i) &\propto (\sigma^2_{u,i})^{-(a_i+1)}, \quad i = 1, \cdots, 16, \end{aligned} \tag{2.11}$$

where y is $n \times 1$, β is a $p \times 1$ vector of fixed effects, $u = (u_1', u_2', \cdots, u_r')'$ is a $q \times 1$ vector, u_i is $q_i \times 1, \Sigma_u = \oplus_{i=1}^r I_{q_i}\sigma^2_{u,i}$, $\sum_{i=1}^r q_i = q$, X and Z are known design matrices whose dimensions are $n \times p$ and $n \times q$. Also, a_i and b are constants.
In this model the posterior density is given by (cf. Hobert and Casella, 1996, p. 1463):

$$\begin{aligned} f(\sigma^2_{u,1}, \cdots, \sigma^2_{u,r}, \sigma^2_\epsilon, u, \beta \mid y) \propto\ & f(y \mid u, \sigma^2_\epsilon, \beta) f(u \mid \sigma^2_{u,1}, \cdots, \sigma^2_{u,r}) \cdot \\ & \cdot \pi_\beta(\beta) \pi_{\sigma^2_\epsilon}(\sigma^2_\epsilon \mid b) \prod_{i=1}^r \pi_{\sigma^2_{u,i}}(\sigma^2_{u,i} \mid a_i) \ . \end{aligned} \tag{2.12}$$

It is often difficult, or even not possible, to integrate posterior densities analytically. For the simple Fay-Herriot model (2.3) the posterior distribution is given in Example 2.12. Unfortunately, the posterior density (2.12) for model (2.11) is not analytically tractable and may even not be proper (cf. Hobert and Casella, 1996). The propriety will be further discussed in the next section.

2.2.2 Improper Priors

In the previous section the general linear mixed model (2.7) has been extended to a hierarchical version (2.11) by specifying prior distributions on all parameters. If there is neat information about a parameter the prior distribution may be quite specific. Consider for example the unemployment data of Example 2.2 and the Fay-Herriot model (2.3). Assume that for a specific covariate, for example education, there might be some knowledge about the connection to unemployment, and thus, for the regression component β perhaps a normal distribution with a small variance might be specified. The small variance represents strong confidence in the prior information. In many other settings prior information on parameters are quite vague. When prior information is unknown improper priors are chosen (cf. Gelman et al., 1995, p. 52 or Ghosh et al., 2006, p. 40), which are defined below.

Definition 2.13: A density $f(\theta)$ is called **improper**, if $f(\cdot)$ is non-negative for all $\theta \in \Theta \neq \emptyset$ and:

$$\int_{\Theta} f(\theta) d(\theta) = \infty.$$

One example is the improper prior (2.8) used in the hierarchical extension of the Fay-Herriot model, where $\pi_\beta(\beta) \propto 1$.

If an improper prior is taken, the posterior density of the corresponding model might also not integrate to a finite number and thus might fail to be a proper probability distribution. Since all the inference is based on the posterior distribution, this is problematic in Bayesian inference. Another possible danger is that frequently Markov chain Monte Carlo methods like the Gibbs sampler (presented in the next Section 2.3) are used to obtain estimates using

the posterior distribution. These methods may not detect impropriety (Datta and Smith, 2003, p. 176) and lead to wrong conclusions, as will be shown in Example 2.16. Therefore, it is vital to check the propriety of the posterior distribution in advance. In Chapters 3, 4, and 5 we introduce existence theorems regarding the propriety of the posterior distribution for extensions of three model classes widely used in small area estimation. The model classes considered are the general linear mixed model, the Bernoulli model, and the hierarchical Fay-Herriot model.

Propriety for similar model classes is discussed by Ibrahim and Laud (1991), who proved a sufficient condition for the existence of the posterior for the general linear model. Ghosh et al. (1998) considered generalized linear models for small area estimation and showed the propriety for spatial and non-spatial models under this framework. The theorem is generalized in Ghosh et al. (1999). Datta and Smith (2003) showed the propriety for (non-spatial) small area models, like the Fay-Herriot model under a bounded prior.

In this section the idea of Bayesian modeling has been explained, where all the variation is included in the model by specifying prior distributions on all unknown parameters. The inference is then based on an update of the sample and the prior information, yielding the posterior distribution. However, the derivation of posterior densities (2.10) is not easy and involves (multi-dimensional) integration. Therefore, in most applications the density is not derived, but a sample out of the distribution is drawn via Markov chain Monte Carlo methods. These methods are introduced in the next section.

2.3 The Impact of Impropriety on Markov Chain Monte Carlo Methods

When using hierarchical Bayes methods (Section 2.2) there is a need to calculate the posterior distribution (2.10), and thus, to compute multi-dimensional integrals. One method used to compute posterior quantities is exact numerical integration. However, the numerical integration is not applicable in high dimensions, making MCMC procedures a natural choice for approximating posterior densities. The idea behind these methods is to construct a Markov chain which eventually converges to the posterior distribution. Instead of calculating the posterior distribution directly, a sample from the posterior is obtained. Then the mean, variance, and other statistics of the posterior distribution may be estimated.

The most popular MCMC procedure is the Gibbs sampler, which will be introduced in the next section. In addition, it is shown that the Gibbs sampler may lead to misleading results if the posterior distribution is improper. We also introduce software for implementing MCMC methods.

2.3.1 The Effect of Impropriety on the Gibbs Sampler

The Gibbs sampler was derived by Geman and Geman (1984). An explanation is given by Casella and George (1992). The Gibbs sampler is a technique for generating random variables from a distribution with density, say $f(x)$. But rather than to compute or approximate the distribution directly; the Gibbs sampler generates the sample by generating the conditional distributions of the model of interest.
In a two random variable case, say (X, Y) the Gibbs sampler constructs a sample of $f_X(x)$ by sampling instead from the conditional densities (Gibbs conditionals) $f_{X|Y=y}(x)$ and $f_{Y|X=x}(y)$, that are more likely to be known in statistical models (cf. Casella and George, 1992, p. 168). Then the basic scheme of the Gibbs sampler is as follows:

1. Choose an arbitrary starting point, say (x_0, y_0).

2. Suppose we have generated some arbitrary point in the chain (x_i, y_i),

$$\text{Generate } x_{i+1} \sim f_{X_{i+1}|Y=y_i}(x)$$
$$\text{Generate } y_{i+1} \sim f_{Y_{i+1}|X=x_{i+1}}(y)$$

3. Now that (x_{i+1}, y_{i+1}) has been generated update and repeat the cycle.

It can be shown that the Gibbs sampler converges under certain regularity conditions to a stationary distribution (cf. Gelfand and Smith, 1990). Thus, after some finite number of steps the generated values from the Gibbs sampler represent values from the stationary distribution. But why do the stationary and the posterior distribution coincide? The following theorem shall clarify the ties between the posterior distribution and the Gibbs conditionals (cf. Cressie, 1993, pp. 412 and Besag, 1974, pp. 195). The same theorem is frequently used to prove an important result in spatial modeling (Theorem 2.22 in Section 2.4).

Theorem 2.14: *Let $X_1, \cdots, X_k$ be a finite collection of random variables with joint probability mass function $f(\cdot)$ whose support satisfies the positivity condition. That is, if $f(x_i) > 0$ for each i, then $f(x) > 0$. That means if $x_1, \cdots, x_k$ can individually occur at sites $1, \cdots, k$, then they can occur together. Let $x = (x_1, \cdots, x_k)$ and $y = (y_1, \cdots, y_k)$ be two realizations of $X_1, \cdots, X_k$. Then:*

$$\frac{f(x)}{f(y)} = \prod_{i=1}^{k} \frac{f_{X_i|X_{-i}}(x_i|x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_k)}{f_{X_i|X_{-i}}(y_i|x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_k)}. \tag{2.13}$$

The proof may be drawn from (Cressie, 1993, pp. 412) and (Besag, 1974, pp. 195).

Theorem 2.14 shows that under certain assumptions the posterior density is essentially proportional to the Gibbs conditionals. But this does not mean that the posterior distribution exists every time the Gibbs conditionals exist, as the following example shows (cf. Robert and Casella, 2010, pp. 232).

Example 2.15: Let the Gibbs conditionals be given by:

$$X_1|X_2 = x_2 \sim \text{Exp}(x_2)$$
$$X_2|X_1 = x_1 \sim \text{Exp}(x_1),$$

where Exp denotes the Exponential distribution. Using Theorem 2.14 it follows that:

$$\begin{aligned} f(x_1, x_2) &\propto \frac{f_{X_1|X_2}(x_1|y_2) \cdot f_{X_2|X_1}(x_2|x_1)}{f_{X_1|X_2}(y_1|y_2) \cdot f_{X_2|X_1}(y_2|x_1)} \\ &\propto \exp(-x_1 x_2). \end{aligned}$$

Since:

$$\int \int \exp(-x_1 x_2) dx_1 dx_2 = \infty,$$

there exists no joint density with the above Gibbs conditionals.

In Example 2.15 it can be seen, that the existence of Gibbs conditionals does not ensure the propriety of the joint distribution. This is a potentially perilous situation, since the Gibbs conditionals allow implementing the Gibbs sampler, but there is no joint distribution to which the Gibbs sampler may converge. What happens in this case? The dangerous effects are illustrated in the next example, where an example from Casella and George (1992) is implemented in R and analyzed. The code is given in Appendix A.

Example 2.16: Consider the conditional densities given by:

$$\begin{aligned} f_{X|Y=y}(x) &\propto y \exp(-yx), 0 < x < \infty \\ f_{Y|X=x}(y) &\propto x \exp(-xy), 0 < y < \infty. \end{aligned} \tag{2.14}$$

For these densities, Casella and George, 1992, p. 171 showed that the marginal density for which the Gibbs sampler should converge is given by:

$$f(x) = \frac{1}{x}.$$

However, $f(x)$ is not a proper density. This may have serious implications for the convergence of the Gibbs sampler as will be shown. Convergence is typically assessed by the scale reduction factor ($\hat{R}$) developed by Gelman and Rubin (1992). To accomplish this different Markov chains from overdispersed starting points are run parallel and the within-chain and between-chain variances are compared. Convergence is diagnosed if the output from all chains is indistinguishable (independent of the initial values). The scale reduction factor points to 1 in this case. Table 2 shows the maximum value of the scale reduction factor for varying number of iterations. Some of the values are near 1 and point

to convergence. This is a dangerous situation. Despite the fact that $\hat{R}$ indicates convergence, the Gibbs sampler did not converge. One indication of this is that starting with iteration number 6000 and contributing on, all the $\hat{R}$ values increase. From Figure 3, we can see that we are clearly not reaching the stationary distribution of the chain. In the case of convergence, we should see that the two chains are eventually mixing.

If the conditionals (2.15) are restricted to the interval $(0, K)$, where K is a positive constant, the marginal distribution is proper. Table 3 and Figure 4 give the values of the scale reduction factor and the trace plot for the proper model ($K = 20$). In this case both measures indicate the model has not failed to converge.

Note the difference that trace plots are able to detect, which is not detected by $\hat{R}$. This is dangerous in simulation studies, where typically many repetitions are made. In this case, it is difficult to check convergence by means of plots and $\hat{R}$ is much more convenient since it allows to check for convergence automatically. In addition, Robert and Casella, 2010, p. 233 mention that graphical monitoring may sometimes lead to misleading results.

500	1000	1500	2000	2500	3000	3500	4000	4500	5000
1.04	1.28	1.16	**1.06**	1.13	1.31	1.22	**1.07**	1.17	1.28
5500	6000	6500	7000	7500	8000	8500	9000	9500	10000
1.00	1.21	1.76	2.13	2.46	2.81	3.08	3.41	3.87	3.85

Table 2: Number of Iterations and Values of the Scale Reduction Factor for the Improper Model.

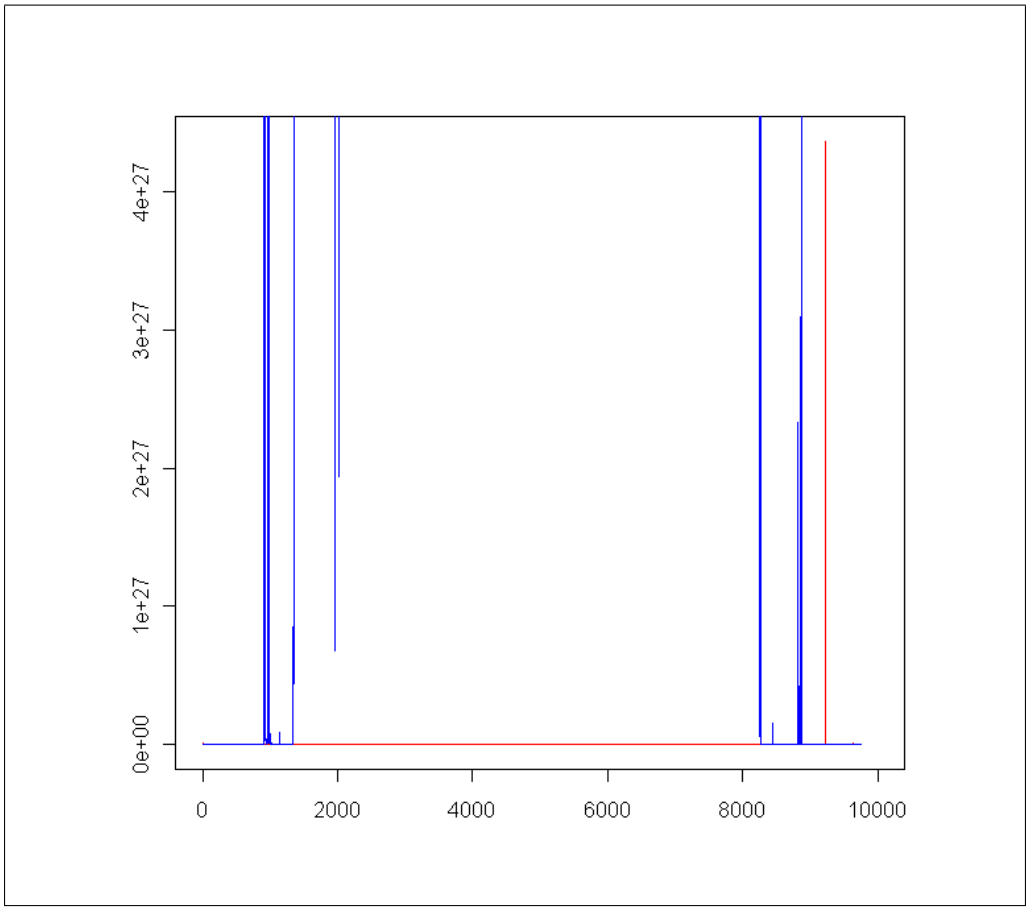

Figure 3: Trace Plot of the Simulated Values of x (Chain 1 : Red, Chain 2 : Blue) of the Gibbs Sampler for the Improper Model.

500	1000	1500	2000	2500	3000	3500	4000	4500	5000
1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5500	6000	6500	7000	7500	8000	8500	9000	9500	10000
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 3: Number of Iterations and Values of the Scale Reduction Factor for the Proper Model.

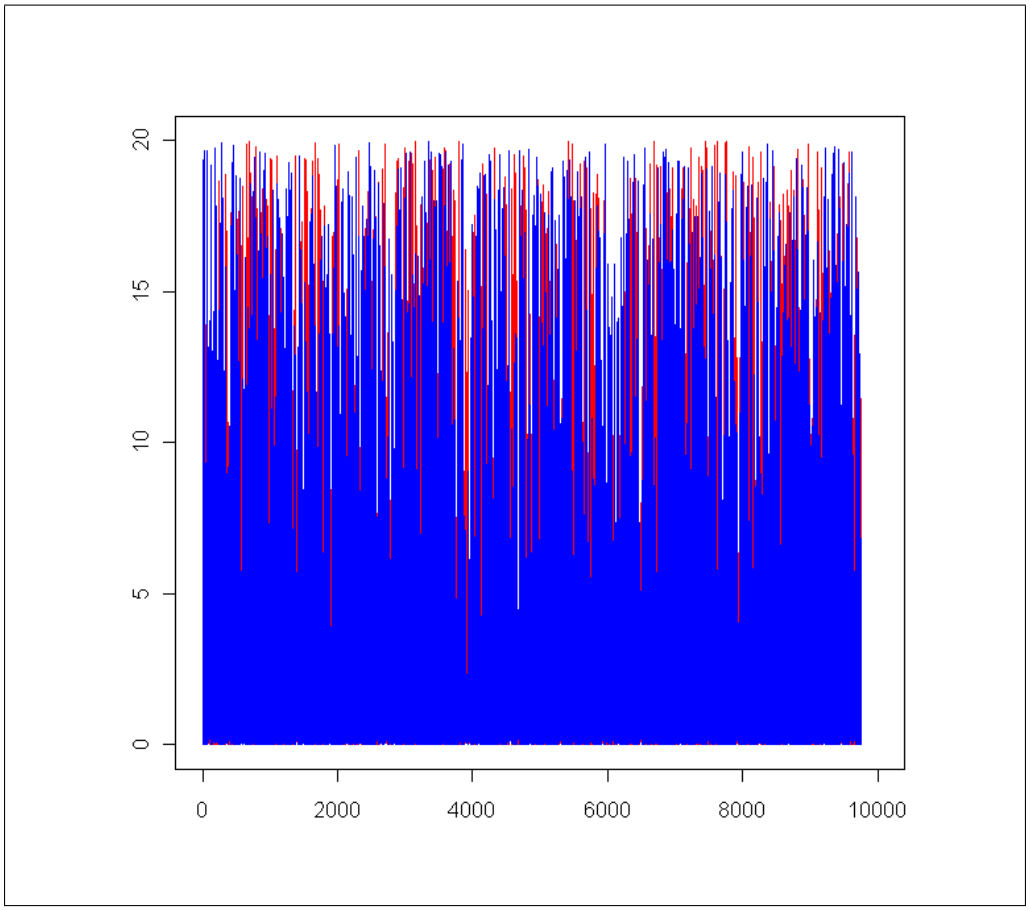

Figure 4: Trace Plot of the Simulated Values of x (Chain 1 : Red, Chain 2 : Blue) of the Gibbs Sampler for the Proper Model.

The situation of Example 2.16 may arise when the Gibbs sampler is applied to a model with improper priors (see Section 2.2.2). Then, as in Example 2.15, the Gibbs conditionals may be of standard forms, but the posterior distribution for which the Gibbs conditionals correspond may be improper. Example 2.16 showed that this is a dangerous situation because the Gibbs sampler may lead to seemingly reasonable inferences about an improper posterior distribution (Hobert and Casella, 1996, p. 1462). Robert and Casella, 2010, p. 233 mention that the only way to make sure the Gibbs sampler is valid is to check that the joint density has a finite integral. Thus, it is absolutely vital to apply only models with proper posterior distributions. Therefore, in Chapters 3, 4, and 5 theorems are proved, which ensure the existence of the posterior distribution for different models of interest.

Besides the Gibbs sampler there are numerous other MCMC methods, like the Metropolis-Hastings algorithm. Since the Gibbs sampler is the only explicitly used MCMC method in this work, the reader is referred to the literature for other methods (for example Fishman, 2006 or Gilks et al., 1996).

2.3.2 Markov Chain Monte Carlo in R and WinBUGS

In the last section, the Gibbs sampler and general MCMC methods were introduced. Several R packages to implement these methods in R are described in the CRAN Task View: Bayesian Inference (cran.r-project.org/web/ views/Bayesian.html). For example the R package UMACS (Universal Markov chain sampler) allows implementation of some MCMC methods, such as the Gibbs sampler. However, the problem is that the Gibbs conditionals are needed, which are often hard to derive. Therefore, in this section, we introduce another software package, called WinBUGS. WinBUGS allows for an easier implementation of Bayesian models (Ntzoufras, 2009 covers Bayesian modeling using WinBUGS) at the cost of flexibility.

The following example provides the Gibbs conditionals for the Fay-Herriot model. In addition the corresponding Gibbs sampler is implemented in UMACS.

Example 2.17: The Gibbs conditionals for the Fay-Herriot model (2.4) with known homoscedastic variances σ_ϵ^2 and σ_u^2 are given by (cf. Hobert and Casella, 1996, p. 1464):

$$\begin{aligned} f(u|\sigma_\epsilon^2, y, \sigma_u^2, \beta) &= N\left((I_k + \frac{\sigma_\epsilon^2}{\sigma_u^2} I_k)^{-1}(y - X\beta), \sigma_\epsilon^2 (I_k + \frac{\sigma_\epsilon^2}{\sigma_u^2} I_k)^{-1}\right) \\ f(\beta|\sigma_\epsilon^2, y, \sigma_u^2, u) &= N\left((X'X)^{-1}X'(y-u), \sigma_\epsilon^2 (X'X)^{-1}\right). \end{aligned} \tag{2.15}$$

Those functions may be used to implement the Gibbs sampler for example via UMACS. The UMACS code consists of initial values of u, β, and θ, update functions containing the Gibbs conditionals (2.15), and the sampler. The sampler contains the needed variables and data. One cycle of the Gibbs sampler is completed when all parameters are updated once. Each update step ends with random numbers drawn out of the corresponding Gibbs conditional.

As soon as convergence is reached, we can sample from the posterior distribution. For model (2.15) the update functions are:

```
1
  u.update <- function () {
      theta.hat <- solve(diag(k)+sigmau/sigmae*diag(k))%*%y-(X%*%t(beta))
      V.theta <- (sigmau*solve(diag(k)+sigmau/sigmae*diag(k)))
      rmnorm(1, theta.hat, V.theta)
6 }

  beta.update<- function () {
    u<-as.vector(u)
    V.beta<-sigmau*solve(t(X)%*%(X))
11  beta.hat <-1/sigmau*V.beta%*%t(X)%*%(y-u)
    rmnorm(1,beta.hat,(V.beta))
  }

  theta.update<-function() {
16  u<-as.vector(u)
    theta<-X%*%t(beta)+u
    theta<-as.vector(theta)
    theta  # There are no Gibbs conditionals for theta. theta is X%*%t(beta)+u,
           # but formally a Gibbs update has to be defined.
21 }
```

Listing 1: UMACS Update Functions for the Fay-Herriot Model with Known Homoscedastic Variances

UMACS allows implementation of various MCMC methods, and thus, specially tailored algorithms may be programmed. The drawback of this flexibility is that it may not be easy to obtain the update functions. In Section 3.4 the needed Gibbs conditionals for an implementation, for example in UMACS, of a spatial extension of model (2.7) are derived.

An alternative to programming in R is WinBUGS. Like UMACS, WinBUGS uses MCMC methods to calculate samples out of the posterior distribution. But statisticians just have to specify the model, there is no need to worry about Gibbs conditionals or to program the MCMC methods explicitly. The following example provides the WinBUGS code for the Fay-Herriot model.

Example 2.18: The WinBUGS model specification for the Fay-Herriot model (2.4) with known homoscedastic variances σ_ϵ^2 and σ_u^2 is as follows:

```
model {

  for (i in 1:k) {
    Y[i]~dnorm(theta[i],sigmae)
    theta[i]<-alpha+beta*X[i]+u[i]
  }

  alpha~dflat()
  beta~dflat()

  for (i in 1:k){
    u[i]~dnorm(0,sigmau)
  }
}
```

Listing 2: WinBUGS Specification of the Fay-Herriot Model with Known Homoscedastic Variances

The code reflects the notation of the Fay-Herriot model (2.4). Thus, the main advantage of WinBUGS is that models can be directly written down and WinBUGS does all the work. This is especially useful if different models are tested against each other. However, the user still has to be cautious about issues like convergence, the number of chains, the length of burn-in, or the propriety of the posterior distribution. In Chapters 3, 4, and 5 the last of these issues, propriety of the posterior, is tackled by providing theorems for different model classes, which ensure the propriety.

The results of the applications in the thesis are obtained via WinBUGS controlled by R (package R2WinBUGS). The R code and the WinBUGS models are provided in the appendix.

Until now the idea of small area estimation as well as standard models have been introduced and the models were extended to hierarchical models by specifying prior distributions on all parameters. Example 2.16 underlined that one has to be careful with certain priors

(improper priors) since this may lead to misleading results. This topic will be further discussed in Chapters 3, 4, and 5. Before doing so, in the next section, the independence assumption of the random effects of the standard and hierarchical models will be dropped and spatial correlations will be included in the models.

2.4 Spatial Modeling in Small Area Estimation

In Example 2.2 we explain how to use the Fay-Herriot model to estimate the unemployment rates for the federal states in Germany. One assumption of model (2.11) is independence of the random effects u_i. Thus, the variation in addition to the regression component between the areas is assumed to be independent. This assumption is not always appropriate, since people tend to cluster. Clustering can occur geographically such as in towns, villages or socially. Spatial interactions may be included into a model by means of spatial modeling. Introductions can be found in the landmark work by Cressie (1993) and the Practical Handbook of Spatial Statistics by Griffith (1996), in Haining (2003), and Ripley (2004). The book by Banerjee et al. (2004) combines hierarchical modeling and analysis for spatial data and Rao (2003) includes a section on spatial modeling in small area estimation. Bivand et al. (2008) explain how to use R for spatial modeling.

2.4.1 Spatial Data

Standard statistical data consists of a data vector Y, and possibly auxiliary variables X. So questions like "Why?", "How?" and "When?" can be tackled. But sometimes it is interesting to know "Where?". Therefore spatial data consists of X, Y, and the location. Cressie, 1993, p. 8 defines spatial data in a very general form.

Definition 2.19: Let $s \in \mathbb{R}^d$ be a generic data location in d-dimensional Euclidean space and suppose that the potential datum $Z(s)$ is a random quantity. Now let s vary over index set $D \subset \mathbb{R}^d$ so as to generate the random process:

$$\{Z(s) : s \in D\} \ . \tag{2.16}$$

Then a realization of (2.16), denoted by $\{z(s) : s \in D\}$, is called **spatial data**.

Where can spatial data be observed? Probably Bivand et al., 2008, p. 1 puts it best by saying "spatial data are everywhere". Examples of spatial data include the weather forecast, route planners, or plain maps showing the temperature or unemployment rate in a certain region.

Since there are many kinds of spatial data, we define three types of spatial data below (cf. Cressie, 1993, pp. 8-13 or Banerjee et al., 2004, p. 2).

Definition 2.20:

1. Spatial data is called **geostatistical** or **point-referenced data** if D is a fixed subset of $\mathbb{R}^d$ that contains a d-dimensional rectangle of positive volume and Y(s) is a random vector at a location $s \in \mathbb{R}$, where s varies continuously over D.
   This means that the data can be theoretically measured everywhere (continuously). One example of this occurrence is temperature in a certain region.

2. The data is referred to as **point pattern data** if the location, D, is random.
   This data type arises if the location is the important variable to analyze, such as the location of a certain tree type.

3. Spatial data is called **lattice** or **areal data** if D is a fixed subset partitioned into a finite number of areal units with well-defined boundaries.
   Unemployment data on federal state, county, or municipality level in Germany is an example of areal data. The data arising from the German Census 2011 will be partitioned into well-defined areas, such as municipalities or cities. Thus, the data is areal. Because the focus of our work is using small area estimation techniques in the context of the German Census 2011, we will only consider areal data.

2.4.2 Spatial Modeling: Conditional Autoregressive Model

In the last section three types of spatial data were introduced. Shortly, we will show how to include spatial information in a model, like the Fay-Herriot (2.4) or hierarchical linear mixed model (2.11). There are several options to do this. In our work the focus will be on the conditional autoregressive (CAR) spatial model, which allows us to incorporate spatial information into a model of interest.

Consider the hierarchical linear mixed model (2.11) and assume that the data of interest is areal (Definition 2.20) with an observed spatial trend, such as the unemployment rate

example in Figure 1. The random effects u_i in the models (2.4) and (2.11) shall handle the differences between the areas which are not captured by the standard regression model. In both models the random effects u_i are assumed to be independent and thus the covariance matrix Σ_u in the following distribution:

$$u \mid \sigma^2_{u,1}, \cdots, \sigma^2_{u,k} \sim N_k(0, \Sigma_u) \tag{2.17}$$

is diagonal. One way to incorporate spatial effects is done by allowing for dependence between the random effects. This may be achieved via the conditional autoregressive model (cf. Rao, 2003, pp. 412 or Banerjee et al., 2004, p. 79).

Definition 2.21: Let $u = (u_1, \cdots, u_k)$ be a k dimensional vector of random variables. Then, the **conditional autoregressive (CAR)** spatial model assumes that the conditional distribution of u_i given $\{u_l : l \neq i\}$ is given by:

$$u_i | \{u_l : l \neq i\} \sim N\left(p \sum_{l \in A_i} Q_{il} u_l, \sigma^2_{u,i}\right), \tag{2.18}$$

where A_i denotes a set of neighboring areas of the i^{th} area, $\{Q_{i,l}\}$ are known constants satisfying $Q_{il} = Q_{li}$ and $p, \sigma^2_{u,i}$ is the unknown parameter vector.

In the CAR model the random effect u_i of area i depends on the effect of the other areas. For better theoretical handling of Formula (2.18), it will be shown that the joint distribution of $u = (u_1, \cdots, u_k)$ is given by:

$$u \sim N_k(0, (I - pQ)^{-1} M), \tag{2.19}$$

where $Q = (Q_{i,l})$ is a $k \times k$ matrix with $Q_{i,l} = 0$ whenever $l \notin A_i$ (including $Q_{i,i} = 0$) and $M = \text{diag}(\sigma^2_{u,i})$. Assumption (2.19) is similar to the independence structure (2.17) with a non-diagonal covariance matrix Σ_u. Using Theorem 2.14 the following theorem may be proved (cf. Cressie, 1993, pp. 412).

Theorem 2.22: *The conditional autoregressive specification (2.18) is given by:*

$$u_i | \{u_l : l \neq i\} \sim N\left(p \sum_{l \in A_i} Q_{i,l} u_l, \sigma^2_{u,i}\right),$$

which implies that:

$$u \sim N(0, (I - pQ)^{-1}M),$$

provided that $(I - pQ)$ *is invertible and* $(I - pQ)^{-1}$ *is symmetric and positive-definite.*

The proof is given in Cressie, 1993, pp. 412.

Banerjee et al., 2004, p. 79 ensure that the covariance matrix in Theorem 2.22 is symmetric and positive-definite by choosing:

$$u_i|\{u_l : l \neq i\} \sim N\left(p\sum_{l\in A_i}\frac{Q_{i,l}}{\sum_{j=1}^{k}Q_{i,j}}u_l, \frac{\sigma_u^2}{\sum_{j=1}^{k}Q_{i,j}}\right). \tag{2.20}$$

The CAR structure in the form (2.20) will be used throughout the work.

Remark 2.23:

1. The CAR structures (2.18) and (2.20) are convenient for the implementation of the Gibbs sampler, since the conditional specification reflects the form of the Gibbs conditionals (cf. Banerjee et al., 2004, p. 86).

2. In applications, the neighborhood matrix Q is symmetric if the neighbors of an area have the area as a neighbor.

3. Another popular spatial model is the **simultaneous autoregressive (SAR)** model, defined as (cf. Banerjee et al., 2004, p. 84) the following:

   $$u \sim N\left(0, \sigma_u^2\left[(I - pQ)(I - pQ)'\right]^{-1}\right).$$

   In this work, the focus will be on the CAR structure, since the conditional specification is more convenient in the Bayesian framework (cf. Banerjee et al., 2004, p. 86).

Theorem 2.22 demonstrates the fact that the conditional autoregressive model imposes a dependence structure on the random effects. We need to decide how to choose this structure in a concrete setting, since the results depend on the chosen neighborhood set A_i in equation (2.20). This topic is discussed next.

2.4.3 Specifying the Geographic Weights Matrix

In Sections 2.4.1 and 2.4.2, we introduced spatial data and the conditional autoregressive structure. The CAR approach (2.18) and (2.20) allows us to incorporate neighborhood information into the model. However, the CAR structure depends on the neighborhood matrix Q. In this matrix the correlation structure between the areas is defined.

In general the spatial neighborhood (weight) matrix is defined as follows (cf. Bavaud, 1998, p. 154).

Definition 2.24: Let $S = \{1, \cdots, k\}$ be a set of places. A **spatial weight matrix** is a $k \times k$ matrix Q of components Q_{ji} satisfying:

1. $Q_{ji} \geq 0, \quad i = 1, \cdots, k, \; j = 1, \cdots, k$
2. $\sum_{i=1}^{k} Q_{ji} = 1, \quad j = 1, \cdots, k.$

In addition, frequently Q_{ii} is assumed to be 0 for $i = 1, \cdots, k$, indicating that an area cannot be its neighbor.

We still must resolve how to specify the set of neighbors, and the literature on this is quite sparse. In many texts on spatial statistics the need to specify the geographic weights matrix is noted and some ideas are mentioned briefly.

1. Cressie, 1993, pp. 384 mentions the following:
   (a) Call any area j, which is within a certain distance from area i a neighbor of that area.
   (b) Define areas as neighbors who share a common boundary.
   (c) Consider different distance metrics: At times a site may be close as the crow flies to another site, but a spatial analysis based on this distance may be strictly for our feathered friends. The closeness of two areas might be a function of both distance and, say, percent urbanization.

2. Banerjee et al., 2004, p. 4 notes using an irregular lattice that physical adjacency is the most obvious (but not the only) way to define a region's neighbor.

3. Hurn et al., 2003, p. 90 mentions that in image analysis problems, typical structures are nearest neighbor with either 4, 8 or 12 neighbors (see Figure 5 for a nearest neighbor setting).

4. In Moura and Migon, 2002, p. 6, they consider Bayesian spatial models for small area estimation of proportions and define two areas to be neighbors if they are contiguous.

5. Haining, 2003, pp. 80, notes that rather than defining linkages between objects in purely geometrical or spatial terms, ancillary data may be used.

6. Bivand et al., 2008, pp. 239 show how to construct contiguity, graph and distance based, higher order, and grid neighbors with R.

Example 2.25: Figure 5 shows different neighborhood structures at the federal state level in Germany. In each plot, non-neighbor states are plotted in yellow, neighbor states in orange, and the actual state in red. The first three plots show the nearest neighbor structure for states Thuringia, Rhineland-Palatinate, and Berlin. Thuringia and Berlin are surrounded by neighbors. Rhineland-Palatinate is situated at the border to Belgium, France, and Luxembourg and thus is not completely surrounded by federal states. Even though Berlin is recognized as a state, it is contained within the state Brandenburg as shown in southwest quadrant of Figure 5. Thus, even though a nearest neighbor structure is assumed for every state the actual structure differs from state to state.

The nearest neighbor structure is taken for example in the Application 4.5. It can be also observed there (Figure 10) that an increase in the number of neighbors leads to an improvement of the quality of the estimates. The last plot in Figure 5 shows a different neighborhood structure. Neighbors of the city state Berlin are the following German city states: Bremen, Hamburg, and the small state Saarland. This structure is tested in the simulation study of Section 6.4.3.

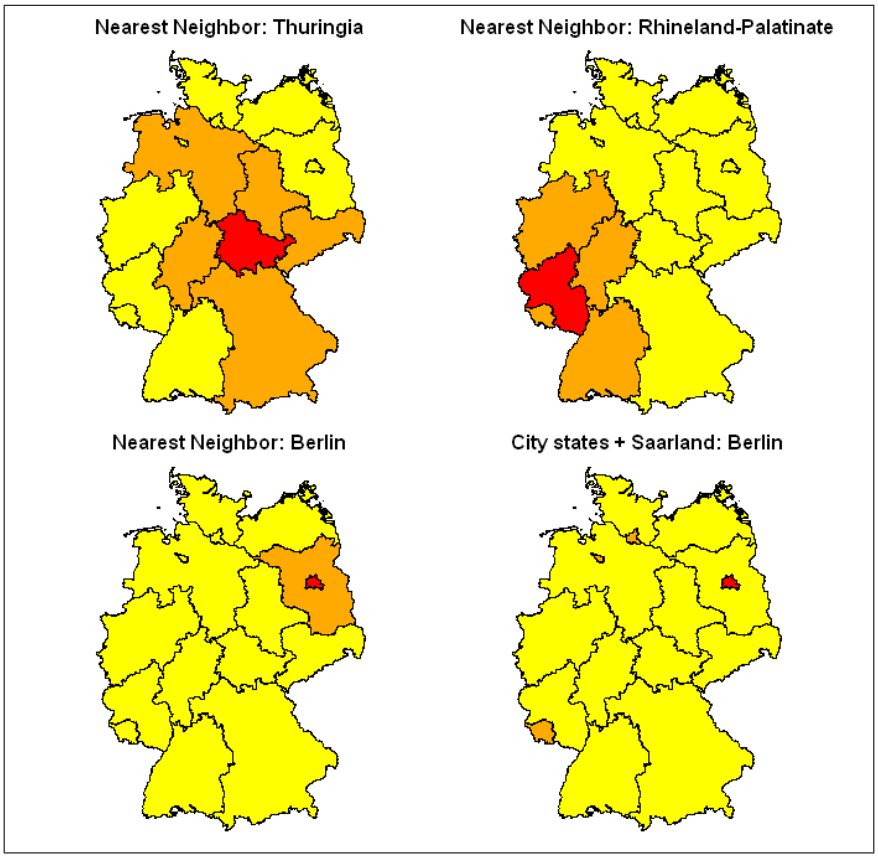

Figure 5: Different Neighborhood Structures for Federal States in Germany.

Example 2.26: For a situation with 3 areas ordered according to Figure 6 the nearest neigbor matrix Q is given by:

$$Q = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

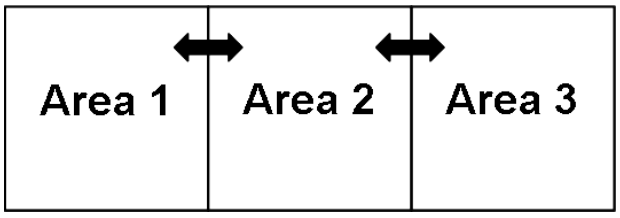

Figure 6: Nearest Neighbor Structure of 3 Areas in a Row.

Although there are many possibilities in choosing the neighborhood structure, only Griffith, 1996 points out some rules of thumb how to choose the neighborhood matrix.

1. It is better to posit some reasonable geographic weights matrix specification than to assume independence.

2. It is best to use a surface partitioning that falls somewhere between a regular square and a regular hexagonal tessellation.

3. Relative large numbers of areal units should be employed.

4. In general, it is better to employ a somewhat under-specified than a somewhat over-specified geographic weights matrix, as long as dependence is assumed.

In the applications of this work the nearest neighbor approach is employed and it is not discussed how to choose the neighborhood structure. The purpose of this section has been to show that many different neighborhood structures are possible. Further research on how to choose the neighborhood in special problems is ongoing and extends the goal of this work.

3 Propriety of the Posterior Distribution for Spatial Hierarchical Linear Mixed Models

In Chapter 2, the concept of Bayesian analysis was introduced, where prior distributions are imposed on all unknown parameters. If only weak prior information is available, frequently improper priors are chosen, which may result in the posterior distributions possibly being improper. Example 2.16 illustrated that this is a dangerous situation. Thus, it is important to check the propriety of the posterior distribution before using a model. Hobert and Casella (1996) proved the propriety of the posterior distribution for the linear mixed model with power priors on the variance components (2.11). Sun et al. (1999) and Sun et al. (2001) extended the Hobert and Casella (1996) theorem to allow for spatial correlations with inverse gamma priors on the variance components and also considered generalized linear mixed models. In this chapter, a simpler proof is presented for a widely used model class with power priors on the variance components. This class includes many models which may be applied to the German Census 2011, for example the spatial extension of the Battese-Harter-Fuller model.

3.1 Introduction

In Bayesian mixed modeling, the distribution of the variance components is generally specified. If this information is not available, non-informative prior distributions are applied, as described in Section 2.2.2. This, however, may lead to integrability problems of the posterior distribution, if the non-informative prior distributions are improper (cf. Datta and Ghosh, 1995 for a discussion of non-informative priors).

Hobert and Casella (1996) proved a theorem that ensures the propriety of the posterior distribution (2.12) for the general linear mixed model (2.11). However, in Hobert and Casella (1996), the random effects are assumed to be independent. In many applications the independence assumption seems to be unreasonable, since spatial correlations are observed. This occurs for example in the case of the unemployment data of Figure 1 on page

2. Spatial interactions are also likely to be present in the context of the German Census 2011. Especially, when sample sizes are small, like in small area estimation, any additional information may be extremely valuable and should be included into the model. Spatial information may be integrated into the model via the popular conditional autoregressive (CAR) approach as described in Section 2.4.2.

In the following section, we review the Hobert and Casella (1996) result regarding propriety of the general linear mixed model. Then the model is extended to allow for spatial correlations, and the propriety of the posterior for the spatial extension is shown. Moreover, the Gibbs conditionals, necessary for an implementation of the Gibbs sampler are derived. Finally, an application to the unemployment data shows potential gains of a spatial model over a non-spatial one.

3.2 The Hobert and Casella Propriety Theorem

Hobert and Casella (1996) considered the following general linear mixed model (2.11):

$$\begin{aligned}
y \mid u, \sigma_\epsilon^2, \beta &\sim N_n(X\beta + Zu, I\sigma_\epsilon^2) \\
\pi_\beta(\beta) &\propto 1 \\
u \mid \sigma_{u,1}^2, \cdots, \sigma_{u,r}^2 &\sim N_q(0, \Sigma_u) \\
\pi_\epsilon(\sigma_\epsilon^2 \mid b) &\propto (\sigma_\epsilon^2)^{-(b+1)} \\
\pi_{\sigma_{u,i}^2}(\sigma_{u,i}^2 \mid a_i) &\propto (\sigma_{u,i}^2)^{-(a_i+1)}, \quad i = 1, \cdots, 16,
\end{aligned}$$

where y is $n \times 1$, β is a $p \times 1$ vector of fixed effects, $u = (u_1', u_2', \cdots, u_r')'$ is a $q \times 1$ vector, u_i is $q_i \times 1$, $\Sigma_u = \oplus_{i=1}^r I_{q_i} \sigma_{u,i}^2$, $\sum_{i=1}^r q_i = q$, X and Z are known design matrices whose dimensions are $n \times p$ and $n \times q$.

They proved the following theorem with necessary and sufficient conditions for the propriety of the posterior distribution.

Theorem 3.1: *Let* $t = rank(P_X Z) = rank(Z' P_X Z) \leq q$ *and define* $P_X = (I - X(X'X)^{-1}X')$. *There are two cases:*

1. *If $t = q$ or if $r = 1$ then the following conditions (a), (b), and (c) are necessary and sufficient for the propriety of the posterior distribution of model (2.11):*

   (a) $a_i < 0$

   (b) $q_i > q - t - 2a_i$

   (c) $n + 2\sum a_i + 2b - p > 0$.

2. *If $t < q$ and $r > 1$ then the foregoing conditions (a), (b), and (c) are sufficient for the propriety of the posterior distribution of model (2.11) while necessary conditions result when (b) is replaced with (b') $q_i > -2a_i$.*

In the next section, this model will be adapted to allow for spatial correlation.

3.3 The Spatial Adaption

In the general linear mixed model (2.11) the random effects are assumed to be independent. This assumption is changed, when allowing for spatial correlation, via the following CAR structure (2.20):

$$u \mid \sigma_u^2, p \sim N_q(0, \sigma_u^2(I - p\tilde{Q})^{-1}W),$$

where $W = \text{diag}(1/\sum_{j=1}^{k} Q_{i,j})$ and $\tilde{Q} = \{Q_{i,l}/\sum_{j=1}^{k} Q_{i,j}\}_{il}$. Note that in contrast to model (2.11) just one random variable u will be included here. Therefore, the dimension of u (q) is equal to the number of areas k. The case of different random variables with variances $\sigma_{u,i}^2$ is discussed in the Notes 3.3. The spatial model is then given by:

$$\begin{aligned} y \mid u, \sigma_\epsilon^2, \beta &\sim N_n(X\beta + Zu, I\sigma_\epsilon^2) \\ \pi_\beta(\beta) &\propto 1 \\ u \mid \sigma_u^2, p &\sim N_k(0, \sigma_u^2(I - p\tilde{Q})^{-1}W) \\ \pi_\epsilon(\sigma_\epsilon^2 \mid b) &\propto (\sigma_\epsilon^2)^{-(b+1)} \\ \pi_{\sigma_u^2}(\sigma_u^2 \mid a) &\propto (\sigma_u^2)^{-(a+1)}. \end{aligned} \tag{3.1}$$

In the spirit of Hobert and Casella (1996), the following existence theorem is stated.

Theorem 3.2: *Let* $t = rank(P_X Z) = rank(Z' P_X Z) \leq k$, *where* $P_X = (I - X(X'X)^{-1}X')$. *The following conditions (a), (b), and (c) are necessary and sufficient for the propriety of the posterior distribution of the above model:*

(a) $a < 0$

(b) $t > -2a$

(c) $n + 2a + 2b - p > 0$,

Note that $r = 1$ and therefore just the first case of Theorem 3.1 will be considered.

Proof*:* Since the covariance matrix is symmetric and positive definite, there exists an orthogonal transformation O such that:

$$O'(I - p\tilde{Q})^{-1}WO = \text{diag}(\lambda_i),$$

where $\lambda_i, i = 1, \cdots, q$ are the eigenvalues of the covariance matrix. Therefore:

$$(I - p\tilde{Q})^{-1}W = ODDO' =: AA'$$

with $D = \text{diag}(\sqrt{\lambda_i})$ and $A = OD$. It then follows that:

$$A^{-1}u \sim N\left(0, A^{-1}\sigma_u^2(I - p\tilde{Q})^{-1}W(A^{-1})'\right) = N(0, \sigma_u^2 I).$$

Defining $u^* = A^{-1}u$ and $Z^* = ZA$, yields:

$$y \mid u^*, \sigma_\epsilon^2, \beta \sim N_n(X\beta + Z^*u^*, I\sigma_\epsilon^2)$$

$$\pi_\beta(\beta) \propto 1$$

$$u^* \mid \sigma_u^2, p \sim N_k(0, \sigma_u^2 I)$$

$$\pi_\epsilon(\sigma_\epsilon^2 \mid b) \propto (\sigma_\epsilon^2)^{-(b+1)}$$

$$\pi_{\sigma_u^2}(\sigma_u^2 \mid a) \propto (\sigma_u^2)^{-(a+1)},$$

and thus, the original theorem with $r = 1$ may be applied. □

Notes 3.3:

1. (a) The standardization via orthogonal transformation is not possible when using the original model with different variance components $\sigma^2_{u,i}$ on the random effects. The covariance matrix is $(I-pQ)^{-1}M$ in this case, where $M = \text{diag}(\sigma^2_{u,i})$. Then the orthogonal transformation of the matrix $(I-pQ)^{-1}$ leads to:

$$(I-pQ)^{-1} = ODDO' =: AA'$$

with $D = \text{diag}(\sqrt{\lambda_i})$ and $A = OD$. It follows that:

$$A^{-1}u \sim N\left(0, A^{-1}(I-pQ)^{-1}M(A^{-1})'\right).$$

Unfortunately, generally:

$$(A^{-1})'M \neq M(A^{-1})'.$$

(b) The transformation:

$$(I-pQ)^{-1}M = ODDO' =: AA'$$

leads to:

$$A^{-1}u \sim N\left(0, A^{-1}(I-pQ)^{-1}M(A^{-1})'\right) = N(0, I),$$

but leaving a dependency of $\sigma^2_{u,i}$ in the orthogonal matrices O and is thus not valid.

2. The model with different variance components is not proper. Calculating the full conditional density of $\sigma^2_{u,i}$ yields:

$$\begin{aligned} & f(\sigma^2_{u,i} \mid \sigma^2_{u,1}, \cdots, \sigma^2_{u,i-1}, \sigma^2_{u,i+1}, \cdots, \sigma^2_{u,k}, y, u, \sigma^2_\epsilon, \beta) \\ = & \frac{f(y|u,\beta,\sigma^2_\epsilon) f(u|\sigma^2_{u,1}, \cdots, \sigma^2_{u,k}) \prod_{i=1}^k \pi_i(\sigma^2_{u,i}|a_i)\pi(\beta)\pi_\epsilon(\sigma^2_\epsilon \mid b)}{\int f(y|u,\beta,\sigma^2_\epsilon) f(u|\sigma^2_{u,1}, \cdots, \sigma^2_{u,k}) \prod_{i=1}^k \pi_i(\sigma^2_{u,i}|a_i)\pi(\beta)\pi_\epsilon(\sigma^2_\epsilon \mid b) d\sigma^2_{u,i}} \\ \propto & \frac{(\det(\Sigma_u))^{-\frac{1}{2}} \exp(-\frac{1}{2}u'\Sigma_u^{-1}u)(\sigma^2_{u,i})^{-(a_i+1)}}{\int (\det(\Sigma_u))^{-\frac{1}{2}} \exp(-\frac{1}{2}u'\Sigma_u^{-1}u)(\sigma^2_{u,i})^{-(a_i+1)} d\sigma^2_{u,i}}, \end{aligned}$$

where the fraction in the above formulas is reduced by the terms independent of $\sigma^2_{u,i}$. Note that:

$$\det(\Sigma_u) = \det((I - pQ)^{-1}M) = \det(M)\det((I - pQ)^{-1}) = \prod_{i=1}^{k} \sigma^2_{u,i} \det((I - pQ)^{-1}),$$

which is proportional to $\sigma^2_{u,i}$. Therefore:

$$\begin{aligned} & f(\sigma^2_{u,i} \mid \sigma^2_{u,1}, \cdots, \sigma^2_{u,i-1}, \sigma^2_{u,i+1}, \cdots, \sigma^2_{u,k}, y, u, \sigma^2_\epsilon, \beta) \\ \propto \; & (\sigma^2_{u,i})^{-\frac{1}{2}} \exp(-\frac{1}{2} u' \Sigma_u^{-1} u)(\sigma^2_{u,i})^{-(a_i+1)} \\ = \; & (\sigma^2_{u,i})^{-(\frac{3}{2}+a_i)} \exp(-\frac{1}{2} u' \Sigma_u^{-1} u), \end{aligned}$$

where $\Sigma_u^{-1} = M^{-1}(I - pQ)$ is a symmetric matrix and $M^{-1} = \text{diag}(1/\sigma^2_{u,i})$. For the exponential function, it follows that:

$$\begin{aligned} \exp(-\frac{1}{2} u' \Sigma_u^{-1} u) &= \exp(-\frac{1}{2} u' M^{-1}(I - pQ)u) \\ &\propto \exp(-\frac{1}{2} u_i \sum_{j=1}^{n} (I - pQ)_{ij} u_j 1/\sigma^2_{u,i}). \end{aligned}$$

Hence:

$$\begin{aligned} & f(\sigma^2_{u,i} \mid \sigma^2_{u,1}, \cdots, \sigma^2_{u,i-1}, \sigma^2_{u,i+1}, \cdots, \sigma^2_{u,k}, y, u, \sigma^2_\epsilon, \beta) \\ \propto \; & (\sigma^2_{u,i})^{-(\frac{3}{2}+a_i)} \exp(-\frac{1}{2} u_i \sum_{j=1}^{n} (I - pQ)_{ij} u_j 1/\sigma^2_{u,i}) \end{aligned}$$

and:

$$\begin{aligned} f(\sigma^2_{u,1} \mid \; & \sigma^2_{u,i}, \cdots, \sigma^2_{u,i-1}, \sigma^2_{u,i+1}, \cdots, \sigma^2_{u,k}, y, u, \sigma^2_\epsilon, \beta) = \\ = \; & IG\left(a_i + \frac{1}{2}, \frac{2}{u_i \sum_{j=1}^{n} (I - pQ)_{ij} u_j}\right), \end{aligned}$$

where IG stands for inverse gamma and $X \sim IG(r, s)$ if $f_X(t) \propto t^{-r-1} \exp(-1/st)$ for positive t. The inverse gamma distribution is defined only when both parameters are positive. Since u_i is normally distributed this may not be the case here. One improper conditional implies an improper posterior (Hobert and Casella, 1996, p. 1464). Therefore, a direct extension of the Hobert and Casella theorem is not possible.

3. Hobert and Casella (1996) applied a similar transformation to an animal breeding model using a different correlation structure: $u_i \mid \sigma^2_{u,i} \sim N_{q_i}(0, G_i\sigma^2_{u,i})$, where G_i are known positive definite matrices.

3.4 The Gibbs Conditionals

In the last section the propriety of the spatial hierarchical general linear mixed model (3.1) was proved under certain conditions. The propriety of the posterior is vital for a meaningful analysis of researchers. The model type (3.1) is typically implemented via MCMC methods, such as the Gibbs sampler, introduced in Section 2.3. To implement the Gibbs sampler, the Gibbs conditionals have to be known. These conditionals are derived below.

Hobert and Casella (1996) derive the following conditionals for the hierarchical general linear model (2.11):

$$f(\sigma^2_{u,i} \mid \sigma^2_{u,i}, \cdots, \sigma^2_{u,i-1}, \sigma^2_{u,i+1}, \cdots, \sigma^2_{u,r}, y, u, \sigma^2_\epsilon, \beta) = IG\left(a_i + \frac{q_i}{2}, \frac{2}{u_i' u_i}\right),$$

$$f(\sigma^2_\epsilon \mid \sigma^2_{u,1}, \cdots, \sigma^2_{u,r}, y, u, \beta) = IG\left(b + \frac{n}{2}, 2\{(y - (X\beta + Zu))'(y - (X\beta + Zu))\}^{-1}\right),$$

$$f(u \mid \sigma^2_{u,1}, \cdots, \sigma^2_{u,r}, y, \sigma^2_\epsilon, \beta) = N_k\left((Z'Z + \sigma^2_\epsilon \Sigma_u^{-1})^{-1} Z'(y - X\beta), \sigma^2_\epsilon (Z'Z + \sigma^2_\epsilon \Sigma_u^{-1})^{-1}\right),$$

$$f(\beta \mid \sigma^2_{u,1}, \cdots, \sigma^2_{u,r}, y, \sigma^2_\epsilon, u) = N_p\left((X'X)^{-1}X'(y - Zu), \sigma^2_\epsilon (X'X)^{-1}\right).$$

These Gibbs conditionals are implemented in an adapted version via UMACS in Section 2.3.2.

Using the above work, the full Gibbs conditionals for the spatial model (3.1) of the unknown parameters of interest σ^2_ϵ, u, and β are given by:

$$f(\sigma^2_\epsilon \mid \sigma^2_u, y, u, \beta) = IG\left(b + \frac{n}{2}, 2\{(y - (X\beta + Zu))' \times (y - (X\beta + Zu))\}^{-1}\right)$$

$$f(u \mid \sigma^2_u, \sigma^2_\epsilon, \beta) = N_k\left((Z'Z + \sigma^2_\epsilon \Sigma_u^{-1})^{-1} Z'(y - X\beta), \sigma^2_\epsilon (Z'Z + \sigma^2_\epsilon \Sigma_u^{-1})^{-1}\right)$$

$$f(\beta \mid \sigma^2_u, y, \sigma^2_\epsilon, u) = N_p\left((X'X)^{-1}X'(y - Zu), \sigma^2_\epsilon (X'X)^{-1}\right),$$

where $\Sigma_u = \sigma^2_u (I - pQ)^{-1} W$. But the conditional density of σ^2_u still needs to be calculated.

Corollary 3.4: The full conditional density of σ_u^2 of model (3.1) is given by:

$$f(\sigma_u^2 \mid y, u, \sigma_\epsilon^2, \beta) = IG\left(a + \frac{k}{2}, \frac{2}{u'(W^{-1} - p\tilde{Q})u}\right).$$

Proof*:* The conditional density of σ_u^2 given all other parameters may be written as:

$$\begin{aligned} & f(\sigma_u^2 \mid y, u, \sigma_\epsilon^2, \beta) \\ = & \frac{f(y|u, \beta, \sigma_\epsilon^2) f(u|\sigma^2) \pi(\sigma_u^2|a) \pi(\beta) \pi_\epsilon(\sigma_\epsilon^2 \mid b)}{\int f(y|u, \beta, \sigma_\epsilon^2) f(u|\sigma_u^2) \pi(\sigma_u^2|a) \pi(\beta) \pi_\epsilon(\sigma_\epsilon^2 \mid b) d\sigma_u^2} \\ \propto & \frac{(\det(\Sigma_u))^{-\frac{1}{2}} \exp(-\frac{1}{2} u' \Sigma_u^{-1} u)(\sigma_u^2)^{-(a+1)}}{\int (\det(\Sigma_u))^{-\frac{1}{2}} \exp(-\frac{1}{2} u' \Sigma_u^{-1} u)(\sigma_u^2)^{-(a+1)} d\sigma_u^2}, \end{aligned}$$

where the fraction in the above formulas is reduced by the terms independent of σ_u^2. Note that:

$$\det(\Sigma_u) = \det(\sigma_u^2 W(I - p\tilde{Q})^{-1}) = (\sigma_u^2)^k \det(W(I - p\tilde{Q})^{-1}),$$

which is proportional to $(\sigma_u^2)^k$. Therefore:

$$\begin{aligned} & f(\sigma_u^2 \mid y, u, \sigma_\epsilon^2, \beta) \\ \propto & (\sigma_u^2)^{-\frac{k}{2}} \exp(-\frac{1}{2} u' \Sigma_u^{-1} u)(\sigma_u^2)^{-(a+1)} \\ = & (\sigma_u^2)^{-(\frac{k}{2}+a+1)} \exp(-\frac{1}{2} u' \Sigma_u^{-1} u), \end{aligned}$$

where $\Sigma_u^{-1} = W^{-1}(I - p\tilde{Q})/(\sigma_u^2)$ and:

$$\exp(-\frac{1}{2} u' \Sigma_u^{-1} u) = \exp\left(-\frac{1}{2} u' \frac{1}{\sigma_u^2} W^{-1}(I - p\tilde{Q})u\right).$$

Furthermore:

$$f(\sigma_u^2 \mid y, u, \sigma_\epsilon^2, \beta) = IG\left(a + \frac{k}{2}, \frac{2}{u'(W^{-1} - p\tilde{Q})u}\right),$$

where $(W^{-1} - p\tilde{Q})$ is positive definite. This is due to the fact that every covariance matrix is positive semi definite and hence every eigenvalue is greater than or equal to 0. Since $(W^{-1} - p\tilde{Q})$ is invertible, every eigenvalue is greater than 0. □

We can now show that the posterior distribution of the spatial adaption of model (2.11) is proper under certain conditions. In addition, the Gibbs conditionals, necessary for

the implementation of the Gibbs sampler can be derived. Thus, the theory allows an implementation of the spatial model. But, just because it is possible to implement a model this does not mean that it is actually worthy to do so. Therefore, in the next section an application is presented which demonstrates positive effects of spatial modeling.

3.5 Application

In this section an application to unemployment data in Germany shows potential advantages of the spatial model (3.1) over the non-spatial version (2.11). The models are compared via a simulation study using real data. The data set (see Table 4 and Figure 7) consists of 16 values representing the unemployment rates from 2007 for the 16 federal states in Germany. This application treats these values as true and no covariates are included. For the simulation setup, 1000 samples of size $16 \cdot 20$ are drawn (simple random sampling) out of a Bernoulli distribution with probability equal to the true value for each area. Each sample represents unit-level information on unemployment. The simulation is done in R and WinBUGS (package R2WinBUGS). The WinBUGS models are included in Appendix B.

In this application a spatial hierarchical extension of the Battese-Harter-Fuller model (2.6) is chosen, contained in model (3.1):

$$\begin{aligned} Y_{ij} &= \beta + u_i + \epsilon_{ij} \\ u \mid \sigma_u^2, p &\sim N_k(0, \sigma_u^2 (I - p\tilde{Q})^{-1} W) \\ \epsilon_{ij} &\overset{\text{i.i.d.}}{\sim} N(0, \sigma_\epsilon^2), \quad i = 1, \cdots, k, j = 1, \cdots, n_i \\ \pi_\beta(\beta) &\propto 1 \\ \pi_\epsilon(\sigma_\epsilon^2 \mid b) &\propto (\sigma_\epsilon^2)^{-(b+1)} \\ \pi_{\sigma_u^2}(\sigma_u^2 \mid a) &\propto (\sigma_u^2)^{-(a+1)}. \end{aligned} \tag{3.2}$$

The non-spatial model is equal to model (3.2) except that:

$$u|\sigma_u^2 \sim N_k(0, \sigma_u^2 I).$$

		Bias of results		RRMSE of results		
Federal state	Data	Non-sp.	Spatial	Non-sp.	Spatial	Diff.
Berlin	15.5	−0.0280	−0.0173	0.3235	0.3359	−0.0124
Bremen	12.7	−0.0118	−0.0114	0.3068	0.3430	−0.0362
Brandenburg	14.9	−0.0259	−0.0225	0.3229	0.2551	0.0678
Baden-Württ.	4.9	0.0328	0.0322	0.8402	0.8109	0.0293
Bavaria	5.3	0.0301	0.0378	0.7491	0.8180	−0.0689
Hamburg	9.2	0.0077	0.0081	0.3763	0.3616	0.0147
Hesse	7.6	0.0181	0.0172	0.4396	0.3501	0.0895
Mecklenburg-Vorp.	16.5	−0.0349	−0.0389	0.3402	0.3155	0.0247
Lower Saxony	8.9	0.0092	0.0181	0.3685	0.3045	0.0640
North Rhine Westph.	9.5	0.0075	0.0021	0.3382	0.2710	0.0672
Rhineland-Palatinate	6.5	0.0248	0.0210	0.5727	0.4848	0.0879
Saxony-Anhalt	16	−0.0292	−0.0333	0.3285	0.2920	0.0365
Saxony	14.7	−0.0231	−0.0248	0.3190	0.2742	0.0448
Schleswig-Holstein	8.4	0.0125	0.0178	0.3833	0.3732	0.0101
Saarland	8.4	0.0109	0.0027	0.3830	0.4178	−0.0348
Thuringia	13.2	−0.0121	−0.0186	0.3290	0.2585	0.0705

Table 4: Unemployment Rate for the 16 German Federal States, as well as Bias, RRMSE, and Difference of the RRMSE for the BHF Models; Data Source: Bundesagentur für Arbeit (2008).

In this application $a = b = -0.05$ and $p = 0.95$. Other choices of the parameters are discussed in Sections 6.4.4 and 6.4.5. Since $\text{rank}(P_X Z) = 1$, Theorem 3.2 ensures the propriety. Note that such models may also be applied to the German Census 2011. The models are compared via the following measures:

- Relative Root Mean-Squared Error (RRMSE) for each area i :

$$\text{RRMSE}_{i,\text{model}} = \sqrt{\frac{1}{1000} \cdot \sum_{l=1}^{1000} \left(\frac{\hat{\theta}_{l,i,\text{model}} - \theta_i}{\theta_i} \right)^2},$$

where the model is either a spatial or a non-spatial model.

- Bias, where:

$$\text{Bias}_i = \frac{1}{1000}\sum_{l=1}^{1000}\hat{\theta}_{l,i,\text{model}} - \theta_i.$$

- Average Gain, where the Gain of sample l is defined by:

$$\text{Gain}_l = \frac{\sum_{i=1}^{16}(\hat{\theta}_{l,i,\text{non-spatial}} - \theta_i)^2}{\sum_{i=1}^{16}(\hat{\theta}_{l,i,\text{spatial}} - \theta_i)^2}.$$

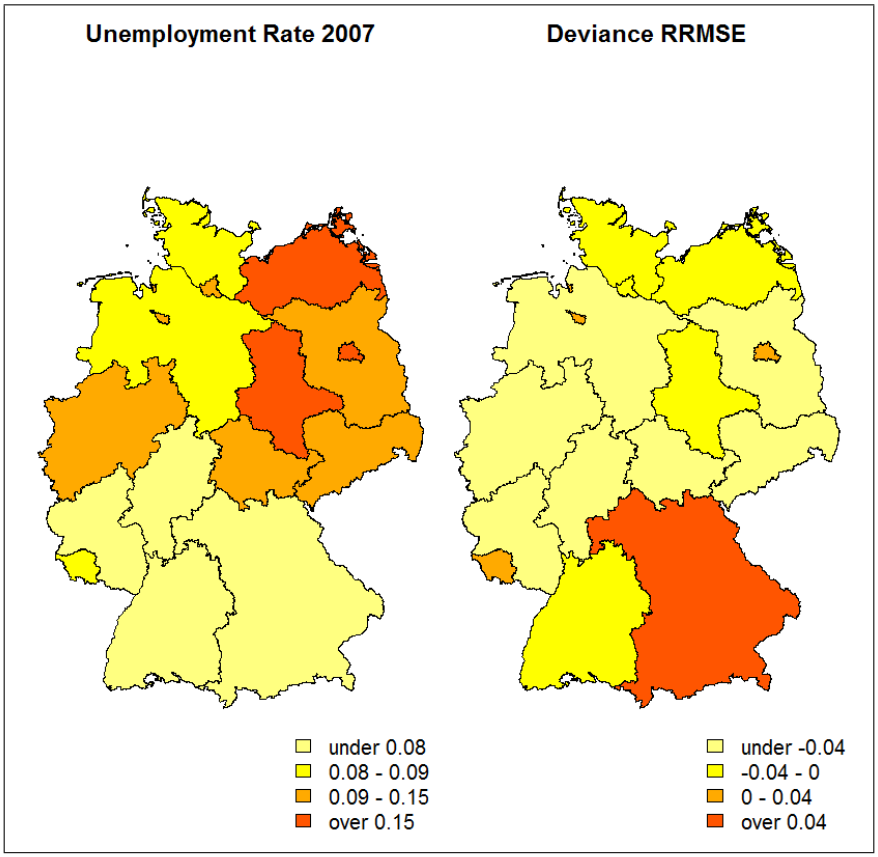

Figure 7: Unemployment Rate 2007 in the Federal States in Germany and Deviance of the RRMSE for the Normal Models.

In the simulation study an average Gain of 1.19 was obtained, indicating an average gain for the spatial model of 19%. In 82.4% of the samples there was a gain greater than 1. Furthermore, the deviance of the RRMSE was found to be:

$$\sum_{i=1}^{16}(\text{RRMSE}_{i,\text{non-spatial}} - \text{RRMSE}_{i,\text{spatial}}) = 0.45,$$

illustrating the better overall performance of the spatial model. For some regions the RRMSE of the non-spatial model is smaller. This will be further discussed in Section 4.5. The above measures indicate positive effects of spatial modeling. The sum of the absolute values of the bias is 0.319 for the non-spatial model and 0.324 for the spatial. Thus, the bias is slightly smaller for the non-spatial model. This may be due to the fact, that in the spatial model information from other areas is taken into account more explicitly, which may increase the bias if the areas do not follow the model. Figures 7 and 8 show and compare the RRMSE on the map for the spatial and non-spatial model. These plots reflect the better performance of the spatial model as noted before.

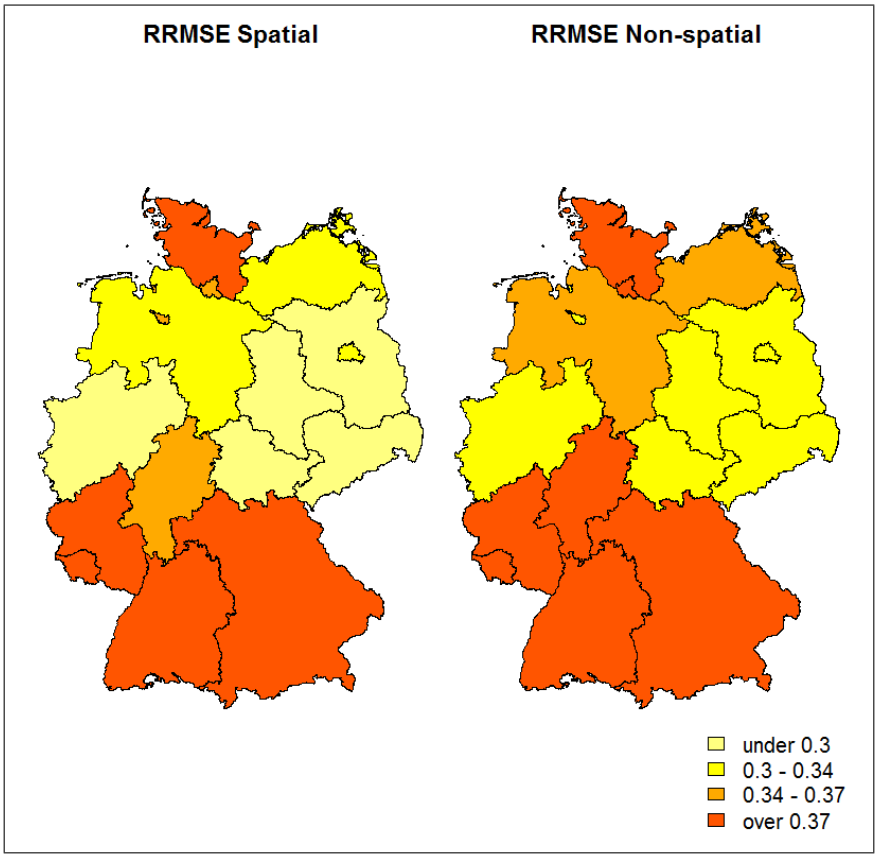

Figure 8: RRMSE for the Spatial and Non-spatial Normal Model.

3.6 Conclusion

In this chapter the propriety of the posterior distribution for the spatial general linear mixed model has been proved. This model class includes a wide range of area as well

as unit-level models, like the spatial hierarchical extension of the Battese-Harter-Fuller model. Since posterior propriety is vital for an application of Bayesian models, this section provided the theoretical basis for models of use to data from the German Census 2011. In addition, the derived Gibbs conditionals allow implementation of the Gibbs sampler, for example via the R package UMACS. An application to unemployment data showed possible gains of spatial modeling.

4 An Existence Theorem of the Posterior Distribution for a Class of Spatial Bernoulli Mixed Models

In Chapter 3 the propriety of the posterior distribution was proved for a Gaussian hierarchical model. In many applications, like the German Census 2011, frequently variables are categorical and normality may not be satisfied. Natarajan and McCulloch (1995) considered a class of Bernoulli mixed models with a power prior on the variance component and showed the propriety for the posterior of the variance component. Vogt and Münnich (2009) extended the Natarajan and McCulloch (1995) model to allow for spatially correlated random effects, proving the propriety of the posterior distribution and demonstrating potential benefits of spatial modeling by means of an application using unemployment data in Germany. This work is presented here. In addition to Vogt and Münnich (2009), an alternative proof of propriety is given, where knowledge of the Natarajan and McCulloch (1995) theorem is not necessary, and the application to unemployment data is explained in more detail. Finally, the results are compared to the findings of Section 3.5.

4.1 Introduction

A typical small area situation is where y_{ij} is binary, and the parameters of interest are the small area proportions $\overline{Y}_{i\cdot} = \sum_j y_{ij}/n_i$ (cf. Rao, 2003, p. 91). This situation arises, for example, in the case of unemployment data, where 0 represents employed and 1 unemployed. Something of interest might be the proportion of unemployed in a federal state in Germany (see Figure 1). In this situation a logistic regression model with random area-specific effects to estimate $\overline{Y}_{i\cdot}$ might be appropriate, where:

$$\begin{aligned} y_{ij} &\overset{\text{ind}}{\sim} \text{Bernoulli}(\theta_{ij}) \\ \text{logit}(\theta_{ij}) &= \log\left(\frac{\theta_{ij}}{1-\theta_{ij}}\right) = x_{ij}^T\beta + u_i, \quad i = 1, \cdots, k,\ j = 1, \cdots, n_i, \end{aligned}$$

where $u = (u_1, \cdots, u_k)^T \sim N(0, \sigma_u^2 I)$ and the x_{ij} are unit-specific covariates. An alternative formulation of this model in vector form, where y_1 corresponds to $y_{1,1}$ and y_n to y_{k,n_k},

is as follows:

$$\begin{aligned} y_i &\sim \text{Bernoulli}\left(h(x_i^T\beta + z_i u)\right) \\ u &\sim N(0, \sigma_u^2 I), \qquad i = 1, \cdots, n. \end{aligned} \tag{4.1}$$

Here, z_i is the first row of the matrix Z which consists of the blocks of length n_i of $1's$ and:

$$h(\cdot) = \frac{\exp(\cdot)}{1 + \exp(\cdot)}.$$

Denote the variance $\sigma_u^2 I$ of u as Σ_u. This more general notation allows for a non-diagonal spatial structure which will be considered later. In a hierarchical version of this model there is a need to specify priors on the hyperparameters. If little is known, a diffuse prior might be formulated, but care must be taken, since as shown in Section 2.2.2 this can lead to an improper posterior distribution. Since mixed effect models with Bernoulli responses are typically implemented via Markov chain Monte Carlo (MCMC) methods, the results of the MCMC iteration may be seriously misleading in cases where the posterior distribution is improper (see Example 2.16). Natarajan and McCulloch (1995) proved a theorem ensuring the propriety of the posterior distribution of the variance component for a class of mixed models with Bernoulli responses. This, however, has serious implications on the use of MCMC methods in Bayesian random effect modeling.

In Natarajan and McCulloch (1995) the random effects are assumed to be independent. In many applications as in economics, social, or geographical sciences, spatial correlations are observed frequently such that the independence assumption seems to be violated. Spatial interactions may occur for example when modeling unemployment data (see Figure 1 on page 2), or in the context of the German Census 2011. Especially when sample sizes are small, any additional information should be included into the model in order to stabilize or improve estimates. Spatial information may be integrated into the model by allowing for dependencies between the random effects. This can be done via the popular conditional autoregressive (CAR) approach described in Section 2.4.2.

The next section reviews the model and theorem previously mentioned of Natarajan and McCulloch (1995). Then, spatial correlations are added to our model and the propriety of the posterior distribution of the variance component for the new model is

verified. Finally, an application (Section 4.5) to unemployment data in Germany reveals potential benefits of the spatial method.

4.2 The Natarajan and McCulloch Theorem

Let $y_1, \cdots, y_n$ be a set of n correlated binary observations. A flexible class of models is assumed by linking the mean to fixed and random effects via a function $h(\cdot)$. Assuming a prior distribution on the variance component σ_u^2, the Natarajan and McCulloch (1995) model is:

$$\begin{aligned} y_i|u &\sim \text{Bernoulli}\left(h(x_i\beta + z_i u)\right) \\ u|\sigma_u^2 &\sim N_k(0, \sigma_u^2 I) \\ \pi_{\sigma_u^2}(\sigma_u^2|a) &\propto \frac{1}{(\sigma_u^2)^{a+1}} \ . \end{aligned} \tag{4.2}$$

The pre-specified constant a characterizes the prior distribution of σ_u^2, and u is a vector of length k of independent random effects. One interpretation of this model in terms of small area estimation is that a binary sample of size n is drawn out of k areas. The random effects vector u accounts for differences between the areas not covered by the regression term $x_i\beta$. In this model the random effects are independent, indicating that apart from the spatial correlation explained by $x_i\beta$, the areas being modeled are treated as independent.

To implement this model via MCMC methods, it is vital that the posterior distribution is proper. Natarajan and McCulloch (1995) proved the existence of the posterior distribution of the variance component σ_u^2 given by:

$$f(\sigma_u^2|y_1, \cdots, y_n) = \frac{\int L(\beta, u|y_1, \cdots, y_n) f(u|\sigma_u^2) f(\sigma_u^2|a) du}{\int\int L(\beta, u|y_1, \cdots, y_n) f(u|\sigma_u^2) f(\sigma_u^2|a) du d\sigma_u^2} \tag{4.3}$$

for model (4.2), where:

$$L(\beta, u|y_1, \cdots, y_n) = \prod_{i=1}^{n} \{1 - h(x_i^*\beta + z_i^* u)\}$$

and $x_i^* = -x_i$ if $y_i = 1$, and $x_i^* = x_i$ if $y_i = 0$. Further, X^* is a $n \times p$ matrix with rows x_i^* and Z^* a $n \times k$ matrix with the rows z_i^*, where z_i^* is defined similarly to x_i^*.

The existence theorem (Natarajan and McCulloch, 1995, p. 641) is as follows.

Theorem 4.1: *For the model (4.2)*

1. *The posterior distribution of* σ_u^2 *exists only when* Π_1 *is satisfied and* $-\frac{k}{2} < a < 0$.
2. *When* $h(\cdot)$ *is the logit or probit function, the posterior distribution of* σ_u^2 *exists if* Π_2 *is satisfied and* $-\frac{1}{2} < a < 0$,

where Π_1 : *dimension*$(\varphi_1) < k$ *and* Π_2 : *dimension*$(\varphi_2) < k$ *and* φ_1 *and* φ_2 *are polyhedral cones given by*

$$\varphi_1 = \{\alpha : Z^*\alpha \leq 0\}, \varphi_2 = \{\alpha : (X^*\beta + Z^*\alpha) \leq 0\}.$$

The proof is given in Natarajan and McCulloch, 1995, pp. 642.

Notes 4.2:

1. The conditions Π_1 and Π_2 are satisfied, i.e. the cones are less than full-dimensional, if for at least one level of the random effect there is a success and a failure (cf. Natarajan and McCulloch, 1995, pp. 641).
2. Chen et al. (2002) extended the Natarajan and McCulloch (1995) model to allow for more than one independent random effect and gave necessary and sufficient conditions for the propriety of generalized linear mixed models. In this chapter, similar to Chapter 3, just one random effect per area is considered.

4.3 The Spatial Existence Theorem

Model (4.2) may be extended to a spatial model by substituting the distribution of the random effect by the CAR structure (2.20):

$$u|\sigma_u^2 \sim N_k(0, \sigma_u^2(I - p\tilde{Q})^{-1}W),$$

where $W = \text{diag}(1/\sum_{j=1}^k Q_{i,j})$ and $\tilde{Q} = \{Q_{i,l}/\sum_{j=1}^k Q_{i,j}\}_{il}$. This leads to the model:

$$\begin{aligned} y_i|u &\sim \text{Bernoulli}\left(h(x_i\beta + z_i u)\right) \\ u|\sigma_u^2 &\sim N_k(0, \sigma_u^2(I - p\tilde{Q})^{-1}W) \\ \pi_{\sigma_u^2}(\sigma_u^2|a) &\propto \frac{1}{(\sigma_u^2)^{a+1}}, \end{aligned} \tag{4.4}$$

where p is a constant between the smallest and largest eigenvalue of $(I - p\tilde{Q})^{-1}W$.

Then, the result below holds.

Theorem 4.3: *For model (4.4)*

- *The posterior distribution of σ_u^2 exists only when Π_1 is satisfied and $-\frac{k}{2} < a < 0$.*
- *When $h(\cdot)$ is the logit or probit function, the posterior distribution of σ_u^2 exists if Π_2 is satisfied and $-\frac{1}{2} < a < 0$.*

There will be two proofs presented in this work. The first proof follows the lines of Natarajan and McCulloch (1995). In this case knowledge of the original Theorem 4.1 is not necessary. The second proof makes use of the Natarajan and McCulloch (1995) theorem and is much shorter.

Proof*:* Denote the denominator in (4.3) by I. The posterior distribution with density $f(\sigma_u^2|y_1, \cdots, y_n)$ is proper if and only if I converges.

(i) Necessity. First it is shown that the condition $-\frac{k}{2} < a < 0$ is necessary by re-writing I as:

$$\begin{aligned} I &= \int_{\sigma_u^2}\int_u \prod_{i=1}^{n}\{1 - h(x_i^*\beta + z_i^* u)\}((2\pi)^k \det((I - p\tilde{Q})^{-1}W\sigma_u^2)^{-\frac{1}{2}}) \cdot \\ &\cdot \exp\left(-\frac{1}{2}u'\frac{W^{-1}(I - p\tilde{Q})}{\sigma_u^2}u\right) du \frac{d\sigma_u^2}{(\sigma_u^2)^{a+1}}. \end{aligned} \tag{4.5}$$

Since $(\det(\sigma_u^2(I - p\tilde{Q})^{-1}W))^{-\frac{1}{2}} = (\sigma_u^2)^{-k/2}(\det((I - p\tilde{Q})^{-1}W))^{-\frac{1}{2}}$ the integral (4.5) is proportional to:

$$I \propto \int_{\sigma_u^2}\int_u \prod_{i=1}^{n}\{1-h(x_i^*\beta+z_i^* u)\}(\det((I-p\tilde{Q})^{-1}W))^{-\frac{1}{2}} \exp\left(-\frac{1}{2}u'\frac{W^{-1}(I - p\tilde{Q})}{\sigma_u^2}u\right) du \frac{d\sigma_u^2}{(\sigma_u^2)^{a+k/2+1}}.$$

Since $(\det((I - p\tilde{Q})^{-1}W))^{-\frac{1}{2}}$ is finite and nonzero it follows that:

$$I \propto \int_{\sigma_u^2}\int_u \prod_{i=1}^{n}\{1 - h(x_i^*\beta + z_i^* u)\} \exp\left(-\frac{1}{2}u'\frac{W^{-1}(I - p\tilde{Q})}{\sigma_u^2}u\right) du \frac{d\sigma_u^2}{(\sigma_u^2)^{a+k/2+1}}. \tag{4.6}$$

Substitute $u = v(\sigma_u^2)^{\frac{1}{2}}$ gives:

$$I \propto \int_{\sigma_u^2} \int_v \prod_{i=1}^{n} \{1 - h(x_i^*\beta + (\sigma_u^2)^{\frac{1}{2}} z_i^* v)\} \exp\left(-\frac{1}{2} v' W^{-1}(I - p\tilde{Q})v\right) dv \frac{d\sigma_u^2}{(\sigma_u^2)^{a+1}}. \tag{4.7}$$

The integral (4.7) diverges for σ_u^2 in a neighborhood of zero if $a > 0$.

For large σ_u^2, let $x = \frac{1}{\sigma_u^2}$. Then:

$$\int_{\sigma_u^2} e^{-\frac{c}{\sigma_u^2}} (\frac{1}{\sigma_u^2})^{b+1} d\sigma_u^2 = -\int_0^{\infty} e^{-cx} x^{b+1} \frac{dx}{x^2} = -\int_0^{\infty} e^{-cx} x^{b-1} dx.$$

The above integral exists if $c > 0$ and $b-1 > -1$. Thus, for the integral (4.6) this condition is equivalent to $\dfrac{u' W^{-1}(I - p\tilde{Q})u}{2} > 0$ (fulfilled since positive definite) and $a + k/2 > 0$.
It is now proved that Π_1 is necessary. Suppose that Π_1 is not satisfied. Since the integrand in (4.7) is nonnegative, $h(\cdot)$ is monotone, and because σ_u^2 is nonnegative it follows that:

$$I \geq \int_{\sigma_u^2} \int_{\varphi_1} \sum_{i=1}^{n} \{1 - h(x_i^*\beta)\} \exp\left(-\frac{1}{2} v' (W^{-1}(I - p\tilde{Q}))v\right) dv \frac{d\sigma_u^2}{(\sigma_u^2)^{a+1}}. \tag{4.8}$$

The right hand side of (4.8) diverges due to the integral over σ_u^2.

(ii) Sufficiency. Since:

$$\int_0^{\infty} x^n e^{-ax} dx = \frac{\Gamma(n+1)}{a^{n+1}}$$

if $n > -1$ and $a > 0$, it follows after integrating (4.6) over σ_u^2 that:

$$I \propto \int_u \prod_{i=1}^{n} \{1 - h(x_i^*\beta + z_i^* u)\} \frac{du}{(u'(W^{-1}(I - p\tilde{Q}))u)^{(a+k/2)}}.$$

If Π_2 holds, then for every u there exists some index $j_u \in \{1, \cdots, n\}$ such that $x_{j_u}^*\beta + z_{j_u}^* u > 0$. Remember that u is a k-dimensional vector and the integration area is $[-\infty, \infty]^k$. Since for every u there exists some index j_u such that $x_{j_u}^*\beta + z_{j_u}^* u > 0$, changing the integration area to $x_j^*\beta + z_j^* u > 0$ yields a larger integration area. Therefore, I can be bounded in the following way:

$$I \leq \sum_{j=1}^{n} \int_{\{u: x_j^*\beta + z_j^* u > 0\}} \{1 - h(x_i^*\beta + z_i^* u)\} \frac{du}{(u' W^{-1}(I - p\tilde{Q})u)^{(a+k/2)}},$$

where the fact is used that $1 - h(\cdot) \leq 1$, and therefore, it suffices to retain only one term in the product. It is, therefore, enough to inspect the convergence of:

$$I_j = \int_{\{u: x_j^*\beta + z_j^* u > 0\}} \{1 - h(x_i^*\beta + z_i^* u)\} \frac{du}{(u' W^{-1}(I - p\tilde{Q})u)^{(a+k/2)}}.$$

If $h(\cdot)$ is the logit link, it follows that:

$$h(x_i^*\beta + z_i^*u) = \frac{\exp(x_i^*\beta + z_i^*u)}{1+\exp(x_i^*\beta + z_i^*u)}$$

and thus:

$$1 - h(\cdot) = 1 - \frac{\exp(\cdot)}{1+\exp(\cdot)} = \frac{1}{1+\exp(\cdot)} \leq \frac{1}{\exp(\cdot)}.$$

Therefore:

$$I_j \leq \int_{\{u: x_j^*\beta + z_j^*u > 0\}} \exp\{-(x_j^*\beta + z_j^*u)\} \frac{du}{(u'W^{-1}(I - p\tilde{Q})u)^{(a+k/2)}}.$$

Without loss of generality, it is assumed that $z_j^* = (1, 0, \cdots, 0)$, since this merely corresponds to the transformation $v_1 = z_j^*u$ and $v_i = u_i$ $(i = 2, \cdots, k)$. Thus:

$$I_j \leq \int_{\{u_1: x_j^*\beta + u_1 > 0\}} \int_{u_2} \cdots \int_{u_k} \exp\{-(x_j^*\beta + u_1)\} \frac{du}{(u'(W^{-1}(I - p\tilde{Q}))u)^{(a+k/2)}}.$$

Using the fact, that $W^{-1}(I - p\tilde{Q}) = W^{-1} - pQ$ and Q is symmetric it follows that:

$$\begin{aligned}
u'Qu &= (u_2Q_{2,1} + \cdots + u_kQ_{k,1})u_1 + (u_1Q_{1,2} + u_3Q_{3,2} + \cdots + u_kQ_{k,2})u_2 + \\
&\quad + \cdots + (u_1Q_{1,k} + \cdots + u_kQ_{k-1,k})u_k \\
&= u_1(u_2Q_{2,1} + \cdots + u_kQ_{k1} + u_2Q_{1,2} + \cdots + u_kQ_{1,k}) + (u_3Q_{3,2} + \cdots + u_kQ_{k,2})u_2 + \\
&\quad + \cdots + (u_2Q_{2,k} + \cdots + u_{k-1}Q_{k-1,k})u_k \\
&= u_1^2[2(\frac{u_2}{u_1}Q_{2,1} + \cdots + \frac{u_k}{u_1}Q_{k,1}) + (\frac{u_2}{u_1}\frac{u_3}{u_1}Q_{3,2} + \cdots + \frac{u_2}{u_1}\frac{u_k}{u_1}Q_{k,2}) + \\
&\quad + \cdots + (\frac{u_2}{u_1}\frac{u_k}{u_1}Q_{2,k} + \cdots + \frac{u_{k-1}}{u_1}\frac{u_k}{u_1}Q_{k-1,k})].
\end{aligned}$$

Since $u'W^{-1}u = w_1^{-1}u_1^2 + \cdots + w_k^{-1}u_k^2 = u_1^2[w_1^{-1} + \frac{u_2^2w_2^{-1}}{u_1^2} + \cdots + \frac{u_k^2w_k^{-1}}{u_1^2}]$ it follows for I_j after substituting $u_i = y_iu_1$ $(i = 2, \cdots, k)$:

$$I_j \leq \int_{\{u_1: x_j^*\beta + u_1 > 0\}} \exp\{-(x_j^*\beta + u_1)\} \frac{du_1}{u_1^{2a+1}} \int_{y_2} \cdots \int_{y_k} \frac{dy_2 \cdots dy_k}{(*)}, \tag{4.9}$$

where:

$$\begin{aligned}
(*) &= (w_1^{-1} + w_2^{-1}y_2^2 + \cdots + w_k^{-1}y_k^2) - 2p\underbrace{[(y_2Q_{2,1} + \cdots + y_kQ_{k,1})]}_{(***)} \qquad (4.10) \\
&\quad -2p[\underbrace{(y_2y_3Q_{3,2} + \cdots + y_ky_2Q_{k,2}) + \cdots + y_{k-1}y_kQ_{k-1,k}}_{(**)}].
\end{aligned}$$

In $(**)$ every element of the lower diagonal matrix of Q (without the diagonal elements) is only in the sum once with the according y_i and y_j. This is due to symmetry.

Now the existence of the integral (4.9) is shown. Since the integral:

$$\int e^{-x} x^{-(2a+1)} dx$$

exists if:

$$-(2a+1) > 1 \Leftrightarrow a < 0,$$

the integral over u_1 converges.

Therefore, we only need to deal with the integrals over $y_2, \cdots, y_k$ in formula (4.9). The integration area is $[-\infty, \infty]^{k-1}$. But since $(W^{-1} - pQ)$ is positive definite, the denominator is greater than 0. Therefore, there are no singularities, and the integrals exist in a bounded area. It follows, that:

$$\int_{y_2} \cdots \int_{y_k} \frac{dy_2 \cdots dy_k}{(1 + y_2^2 + \cdots + y_k^2)^{a+k/2}} \tag{4.11}$$

exists if $a > -\frac{1}{2}$. We now show that the integrals in (4.9) have the same properties than the integrals (4.11) in Natarajan and McCulloch (1995).

Using the Binomial formula, it follows that:

$$xy < \frac{1}{2}(x^2 + y^2) < x^2 + y^2. \tag{4.12}$$

Therefore $(**)$ is bounded above by:

$$((y_2^2 + y_3^2)Q_{3,2} + \cdots + (y_k^2 + y_2^2)Q_{k,2}) + \cdots + ((y_{k-1}^2 + y_k^2)Q_{k-1,k}).$$

Thus, it follows for $(*)$ (not dealing with $(***)$ at the moment):

$$w_1^{-1} + \sum_{i=2}^{k} w_i^{-1} y_i^2 - 2p(**) - 2p(***) \geq \sum_{j=2}^{k} [(w_i^{-1} - 2p \sum_{i=2, i \neq j}^{k} Q_{ij}) y_j^2] - 2p(***). \tag{4.13}$$

There is still something we must handle in formula (4.13). The coefficient of some y_j^2 might be negative for some j, yielding a singularity. Therefore, we must adjust the inequality in (4.12) to that of (4.14), after dealing with $(***)$ first.

To deal with $(***)$ the integration area is split up into smaller parts, starting with $[0, \infty]^{k-1}$. Since the integral $(*)$ contains no singularities and exists in every bounded integration area, it suffices to consider the area $[M, \infty]^{k-1}$, where $M > 0$ is chosen such that:

$$MQ_{j,1} < y_j Q_{j,1} < y_j^2 \epsilon \text{for} \epsilon > 0, j = 2, \cdots, k.$$

To yield positive coefficients in (4.13) it is assumed that:

$$MQ_{ij} \leq y_i y_j Q_{ij} \leq (y_i^2 + y_j^2)\epsilon \tag{4.14}$$

for $0 < \epsilon < \frac{1}{2}\frac{1}{2p(2k-3)}, i = 2, \cdots, k; \quad j = 2, \cdots, k; \quad j \neq i.$

Therefore, (***) is bounded above by:

$$[(y_2 Q_{2,1} + \cdots + y_k Q_{k,1})] \leq y_2^2\epsilon + \cdots + y_k^2\epsilon. \tag{4.15}$$

And (**) is bounded above by the following alternative expression:

$$w_1^{-1} + \sum_{i=2}^{k} y_i^2 - 2p(**) - 2p(***) \geq \sum_{j=2}^{k}[(w_i^{-1} - 2p(k-2)\epsilon y_j^2] - 2p(***). \tag{4.16}$$

Combining the inequality (4.16) with (4.15) yields (see 4.10):

$$(*) \geq w_1^{-1} + \sum_{j=2}^{k}[\underbrace{(w_i^{-1} - 2p(2k-3)\epsilon)}_{=:a_j} y_j^2] = w_1^{-1} + \sum_{j=2}^{k} a_j y_j^2.$$

Substituting $a_j \cdot w_1 y_j^2 = (y_j^*)^2$ shows that the integrals in (4.9) have the same structure in the integration area $[M, \infty]^{k-1}$ as the integral in (4.11). Using $\epsilon < \frac{1}{2}\frac{1}{2p(2k-3)}$ shows that $a_j > 0$ for all j. Therefore, the integral in (4.9) exists in the special integration area if $a > -\frac{1}{2}$.

Now the entire integration area is considered. The integration area $[-\infty, \infty]^{k-1}$ can be split up into all combinations of the intervals $[-\infty, 0]$ and $[0, \infty]$. All the integrals over these integration areas exist, since the inequality of the Binomial formula (4.12) is true for all values. Also, if for any i, y_i takes on negative values, this only enlarges the denominator, since all entries of the matrix Q are positive. Therefore, substituting y_i by 0 if $y_i < 0$ in (***) shows the existence of the integrals. In other words, the integration area $[0, \infty]^{k-1}$ is the worst possible case, which means that the integrated function is largest here. Since the integral exists in this case, this completes the proof. Note that if $h(\cdot)$ is the logit link, Natarajan and McCulloch (1995) bounded $1 - h(\cdot)$ by $\exp\{-\frac{1}{2}(\cdot)^2\}$ and the same proof applies. □

4.4 Alternative Proof of the Existence Theorem

In this section an alternative, much easier proof for Theorem 4.3 is presented.

Proof: Since the covariance matrix $\sigma_u^2(I - p\tilde{Q})^{-1}W$ is symmetric and positive definite there exists an orthogonal transformation such that:

$$O'((I - p\tilde{Q})^{-1}W)O = \text{diag}(\lambda_i),$$

where λ_i are the eigenvalues of the covariance matrix. Hence:

$$(I - p\tilde{Q})^{-1}W = ODDO' =: AA'$$

with $D = \text{diag}(\sqrt{\lambda_i})$ and $A = OD$. It follows that:

$$A^{-1}u \sim N\left(0, A^{-1}\sigma_u^2(I - p\tilde{Q})^{-1}W(A^{-1})'\right) = N(0, \sigma_u^2 I).$$

Therefore, defining $u^{**} = A^{-1}u$ and $z_i^{**} = z_i A$ yields:

$$\begin{aligned} y_i|u^{**} &\sim \text{Bernoulli}\{h(x_i\beta + z_i^{**}u^{**})\} \\ u^{**}|\sigma_u^2 &\sim N_k(0, \sigma_u^2 I) \\ \pi_{\sigma_u^2}(\sigma_u^2|a) &\propto \frac{1}{(\sigma_u^2)^{a+1}}, \end{aligned}$$

and thus, Theorem 4.1 may be applied. □

Theorem 4.3 allows statisticians to use the spatial Bernoulli mixed models. Within the German Census 2011 many variables will be categorical, and those models are applicable in this framework. However, just because it is possible to use and implement a model does not mean that the analysis will be meaningful. Therefore, in the next section the spatial and non-spatial Bernoulli models are tested on unemployment data in Germany. In addition, the models are compared to the normal models of Chapter 3.

4.5 Application

In this section, the spatial and non-spatial Bernoulli models (4.2) and (4.4) are tested against each other via the same simulation setup as in Section 3.5. Then, the findings are

		Bias of results		RRMSE of results		
Federal state	Data	Non-sp.	Spatial	Non-sp.	Spatial	Diff.
Berlin	15.5	−0.0248	−0.0073	0.3437	0.3885	−0.0448
Bremen	12.7	−0.0122	−0.0059	0.3411	0.4168	−0.0757
Brandenburg	14.9	−0.0241	−0.0180	0.3488	0.2944	0.0544
Baden-Württ.	4.9	0.0303	0.0234	0.8058	0.6956	0.1103
Bavaria	5.3	0.0279	0.0289	0.7270	0.7129	0.0141
Hamburg	9.2	0.0069	0.0071	0.4033	0.4404	−0.0370
Hesse	7.6	0.0157	0.0095	0.4445	0.3366	0.1079
Mecklenburg-Vorp.	16.5	−0.031	−0.0302	0.3566	0.3319	0.0246
Lower Saxony	8.9	0.0077	0.0116	0.3913	0.3025	0.0888
North Rhine Westph.	9.5	0.0062	0.00002	0.3642	0.3421	0.0221
Rhineland-Palatinate	6.5	0.0223	0.0133	0.5645	0.4580	0.0166
Saxony-Anhalt	16	−0.0267	−0.0251	0.3512	0.3109	0.0404
Saxony	14.7	−0.0204	−0.0192	0.3435	0.3141	0.0294
Schleswig-Holstein	8.4	0.0104	0.0143	0.3972	0.4192	−0.0220
Saarland	8.4	0.0095	0.0013	0.4012	0.4870	−0.0858
Thuringia	13.2	−0.011	−0.0157	0.3649	0.3152	0.0497

Table 5: Unemployment Rate for the 16 German Federal States, as well as Bias, RRMSE, and Difference of the RRMSE for the Bernoulli Models; Data Source: Bundesagentur für Arbeit (2008).

compared to the results of the normal models of Section 3.5. Simulation is done using R and WinBUGS, and the WinBUGS models are included in Appendix B.

The logit link is assumed and $a = -0.05$, and $p = 0.95$. Other choices of the parameters

are discussed in Sections 6.4.4 and 6.4.5. Thus, the spatial model is as follows:

$$\begin{aligned} y_i|u &\sim \text{Bernoulli}\{\theta_{i,\text{model}}\} \\ \text{logit}(\theta_{i,\text{model}}) &= \beta + u_i \\ u|\sigma_u^2 &\sim N_k(0, \sigma_u^2(I - 0.95\tilde{Q})^{-1}W) \\ \pi_{\sigma_u^2}(\sigma_u^2) &\propto \frac{1}{(\sigma_u^2)^{-0.05+1}}, \end{aligned} \tag{4.17}$$

where y_i is of length $16 \cdot 20$ containing $0's$ and $1's$ representing employed and unemployed. The neighborhood matrix is constructed via the nearest neighbor structure (see Section 2.4.3). The non-spatial model is equal to (4.17) except that:

$$u|\sigma_u^2 \sim N_k(0, \sigma_u^2 I).$$

Both models assign a proper, non informative normal distribution on β. It is difficult to include the adjacency information into the non-spatial model. The estimator of the true unemployment rate θ_i for the i^{th} state given the spatial model is denoted by $\hat{\theta}_{i,\text{spatial}}$ and for the non-spatial model by $\hat{\theta}_{i,\text{non-spatial}}$. Assumption Π_2 of Theorems 4.1 and 4.3 is satisfied since there is at least one level of random effects for which there is a success and a failure (cf. Natarajan and McCulloch, 1995, p. 642). Thus, Theorem 4.3 ensures posterior propriety.

As in Section 3.5 the two models are compared by the three methods previously discussed: RRMSE, bias, and average Gain.

The average Gain of 1.12 indicates an average gain for the spatial model is 12%. In the simulation study, we find that in about 75% of the samples there has been a gain greater than 1. Furthermore, the deviance of the RRMSE is 0.38, showing the better overall performance of the spatial model. For some regions the RRMSE of the non-spatial model is smaller (see Table 5), however all those regions are city states or unusually small federal states with a few number of neighbors (see Figure 10 on page 64). This may be avoided by choosing an alternative neighborhood structure, where each state has at least two neighbors. The topic is further discussed in Section 6.4.3. The sum of the absolute values of the bias is 0.29 for the non-spatial model and 0.23 for the spatial one. All of

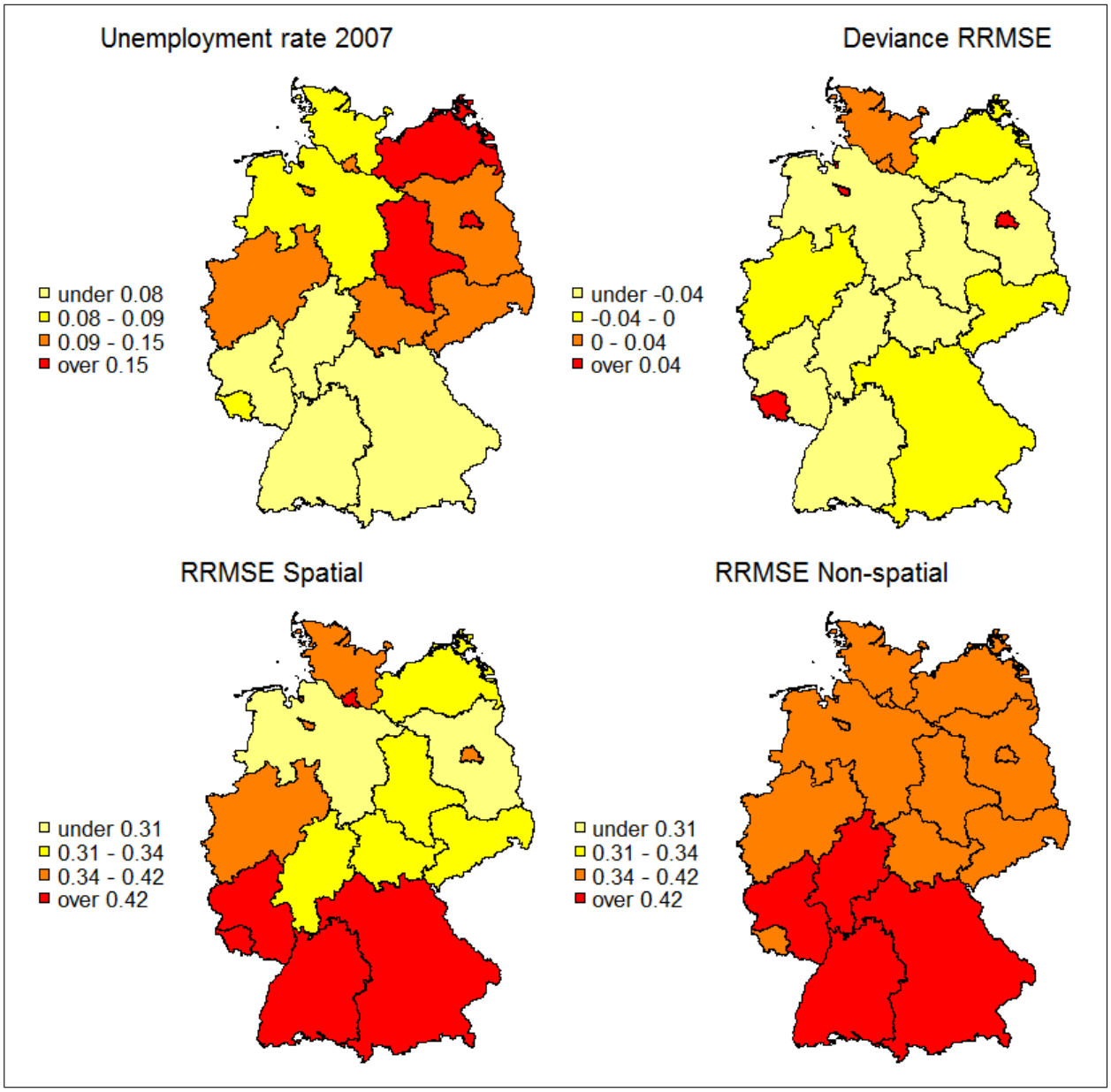

Figure 9: Unemployment Rate 2007, RRMSE, and Deviance of the RRMSE for the Spatial and Non-spatial Bernoulli Model.

the above measures indicate a better overall performance of the spatial model. Figure 9 shows and compares the RRMSE on the map for the spatial and non-spatial model. These plots reflect the better performance of the spatial model as noted before. The sum of the RRMSE and the bias of the Bernoulli models of this section as well as those quantities for the normal models of Section 3.5 are given in Table 6. The RRMSE of the spatial normal model is the lowest, followed by the RRMSE of the spatial Bernoulli model. We notice a

similar occurrence for the non-spatial models. Thus, in terms of the RRMSE, the normal models beat the Bernoulli models, but at the cost of a higher bias. Here, the spatial as well as the non-spatial Bernoulli model beat both normal models.

	Normal models		Bernoulli models	
Measure	Non-sp.	spatial	Non-sp.	Spatial
RRMSE	6.72	6.27	6.95	6.57
Bias	0.32	0.32	0.29	0.23

Table 6: RRMSE and Bias for the Spatial and Non-spatial Normal and Bernoulli models.

4.6 Conclusion

With the development of more powerful computers and MCMC methods, Bayesian modeling has become powerful and popular. Since the results of MCMC methods may be misleading if the posterior distribution is improper, Natarajan and McCulloch (1995) proved an existence theorem for a class of Bernoulli mixed models. However, in many applications spatial correlations are present, and thus, the independence assumption of the random effects is replaced by the conditional autoregressive approach. Within this chapter, we proved an extension of the Natarajan and McCulloch (1995) theorem, which incorporates spatially correlated random effects. We showed for the unemployment rate example that estimation results may improve if the CAR structure is utilized. Thus, if correlation is present it is beneficial to apply spatial modeling. This happens to be the case for many applications in economics, social sciences, or geography, and is likely to happen in the German Census 2011 as well.

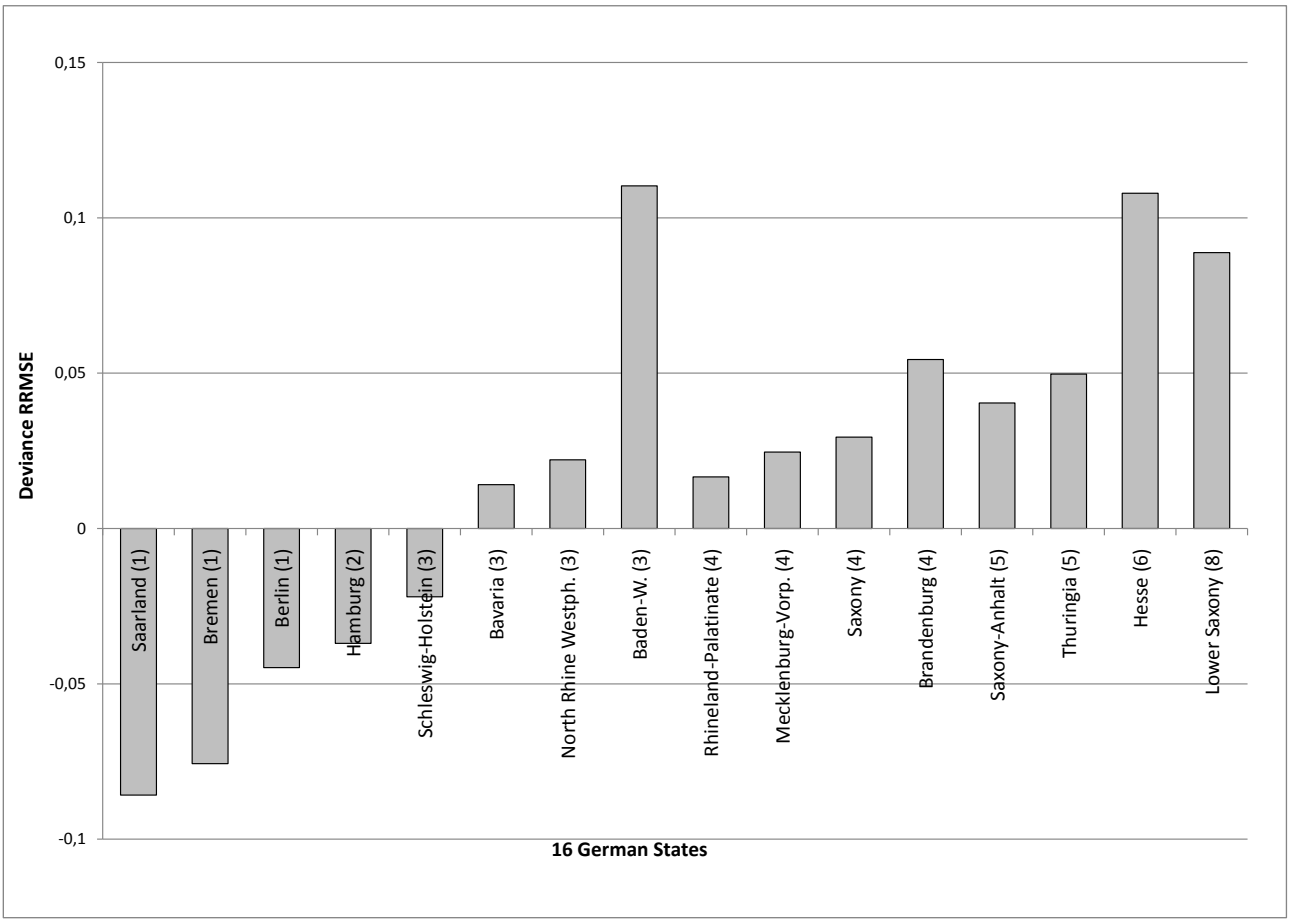

Figure 10: Deviance of the RRMSE for the Federal States, Sorted by Number of Neighbors (Ascending Order).

5 The Spatial Fay-Herriot Model: Propriety and Prediction

In Chapters 3 and 4 propriety theorems for two important spatial model classes were proved, which include a wide variety of unit as well as area-level models, like the spatial hierarchical extension of the Battese-Harter-Fuller model (2.6). However, one of the most commonly applied small area models, the Fay-Herriot model (2.3), is not part of these model classes. This is due to unequal sampling variances, which are not allowed in Chapter 3. The gap is closed in this chapter. Like in the previous chapters, we prove the propriety of the posterior for the spatial extension under certain conditions. In the examples of Chapters 3 and 4, possible gains of spatial modeling could be shown when no covariates are present. In this chapter, the effect of covariates on the benefits of spatial modeling will be analyzed. Therefore, covariates of different types will be included in our example. Moreover, we consider the situation of one unsampled area.

5.1 Introduction

Due to budgetary constraints, it is generally not feasible to sample all small areas of interest. In the framework of the German Census 2011, all geographical areas, like municipalities, are sampled. This might not be the case for areas (domains) arising from social factors, perhaps single fathers below 20 years of age in a certain region. The estimation for these unplanned small areas is problematic if the survey does not produce any samples for these areas. One possible way to obtain estimates for unsampled areas, is to apply small area models which borrow strength from other areas. However, standard small area models borrow strength from every area without a weighting. If spatial correlations are present, this approach may be subject to considerable bias. One potential way to reduce the bias is to model the spatial interactions explicitly via the CAR structure (2.20) and to increase the weight of neighboring areas.

In Section 5.2, an extension of the Fay-Herriot model that incorporates spatial correlations

using an intrinsic spatial CAR model is proposed. The propriety of the posterior distribution under certain regularity conditions is shown. In addition, we derive the necessary theory for predicting one unsampled area and show by means of a simple example how the CAR structure influences the estimates. The spatial extension is tested in Section 5.3 using the Small Area Income and Poverty (SAIPE) data of the United States Census Bureau.

5.2 Theory

In this section the Fay-Herriot model is extended by including prior distributions on the regression coefficient and the variance component. The independence assumption of the random coefficients is replaced by the conditional autoregressive structure. Then we derive a formula for the mean of the unsampled area, and prove the propriety of the posterior distribution.

5.2.1 Spatial Hierarchical Extension of the Fay-Herriot Model

The Fay-Herriot model (2.3) is given by:

$$\begin{aligned} Y_i &\sim N(\theta_i, \sigma^2_{\epsilon,i}) \\ \theta_i &= X_i\beta + u_i, \quad i = 1, \cdots, k \\ u &\sim N(0, \sigma^2_u I), \end{aligned}$$

where $\sigma^2_{\epsilon,i}$ are assumed to be known. This model can be extended to a hierarchical model by assuming prior distributions on the coefficients β and σ^2_u. Here, the prior distributions of Example 2.9 are taken to be:

$$\begin{aligned} \pi_\beta(\beta) &\propto 1 \\ 1/\sigma^2_u &\sim \Gamma(0.5, 0.0005). \end{aligned}$$

Note that it is quite difficult to choose appropriate prior distributions (cf. Kass and Wassermann, 1996). For this analysis standard distributions are assumed (cf. Kelsall

and Wakefield, 2008, p. 151 and GeoBUGS User Manual, 2004, p. 41). We do not address prior selection in this work. The Fay-Herriot model (2.3) can be further extended by assuming dependent random effects:

$$\begin{aligned} Y_i &\sim N(\theta_i, \sigma^2_{\epsilon,i}) \\ \theta_i &= X_i\beta + u_i \\ u &\sim N(0, \sigma^2_u(I - p\tilde{Q})^{-1}W) \\ \pi_\beta(\beta) &\propto 1 \\ 1/\sigma^2_u &\sim \Gamma(0.5, 0.0005), \end{aligned} \tag{5.1}$$

where $W = \text{diag}\left(1/\sum_{j=1}^{k}(Q_{i,j})\right)$ and $\tilde{Q}_{i,j} = \dfrac{Q_{i,j}}{\sum_{j=1}^{k}(Q_{i,j})}$ are the weights (2.20) suggested by Banerjee et al., 2004, p. 79. The neighborhood matrix Q is symmetric with $Q_{ii} = 0$.

The overall goal of this section is to analyze how effective the spatial hierarchical Fay-Herriot model (5.1) is in terms of prediction for one unsampled area, compared to the hierarchical Bayes methodology. To achieve this, we derive a formula for the mean of the unsampled area in the following section.

5.2.2 The Predicted Mean of One Unsampled Area

In order to derive a formula for the mean of the unsampled area under the model (5.1), the conditional distribution of the unsampled area, Y_{US}, given the sampled areas, Y_{S}, is needed. Assume that β, $\sigma^2_{\epsilon,i}$, and σ^2_u are known. Then, model (5.1) may be written in the form:

$$Y \sim N(\mu, \Sigma) \Leftrightarrow \begin{pmatrix} Y_{\text{US}} \\ Y_{\text{S}} \end{pmatrix} \sim N\left(\begin{pmatrix} X_{\text{US}}\beta + u_{\text{US}} \\ X_{\text{S}}\beta + u_S + \epsilon_{\text{S}} \end{pmatrix}, \begin{bmatrix} \Sigma_{\text{US}} & \Sigma_{\text{US,S}} \\ \Sigma_{\text{S,US}} & \Sigma_{\text{S}} \end{bmatrix}\right), \tag{5.2}$$

where $\Sigma_{\text{US}} = ((I_k - p\tilde{Q})^{-1}W\sigma^2_u)_{\text{US}}$, $\Sigma_{\text{US,S}} = ((I_k - p\tilde{Q})^{-1}W\sigma^2_u)_{\text{US,S}}$, $\Sigma_{\text{S,US}} = ((I_k - p\tilde{Q})^{-1}W\sigma^2_u)_{\text{S,US}}$, $\Sigma_{\text{S}} = \sigma^2_{\epsilon,i,S}I_{k-1} + ((I_k - p\tilde{Q})^{-1}W\sigma^2_u)_S$.

Using standard results, the conditional distribution is given by:

$$Y_{\text{US}} \mid (Y_{\text{S}} = y_{\text{S}}, \beta, \sigma^2_u, \sigma^2_\epsilon, p) \sim N(\bar{\mu}, \bar{\Sigma}),$$

$$\text{where } \bar{\mu} = \mu_{\text{US}} + \Sigma_{\text{US},S}\Sigma_{\text{S}}^{-1}(y_{\text{S}} - \mu_{\text{S}})$$

$$\text{and } \bar{\Sigma} = \Sigma_{\text{US}} - \Sigma_{\text{US},S}\Sigma_{\text{S}}^{-1}\Sigma_{S,\text{US}}.$$

The conditional mean $\bar{\mu}$ may be used to predict one unsampled area. Using the block matrix inversion formula:

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \Rightarrow M^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}, \tag{5.3}$$

and so an alternative formulation of $\bar{\mu}$ can be derived. The new specification shows explicitly how the spatial correlations enter the model.

Lemma 5.1: *Consider the model (5.1) with known β, $\sigma^2_{\epsilon,i}$, and σ^2_u in the form (5.2). Then, the mean $\bar{\mu}$ of the conditional distribution of the unsampled area Y_{US} given the other areas Y_S, may be written as:*

$$\bar{\mu} = X_{US}\beta - \sigma^2_u B \left[\sigma^2_{\epsilon,S} I_S W_S^{-1}(D - CA^{-1}B) + \sigma^2_u I_S\right]^{-1}(y_S - X_S\beta), \tag{5.4}$$

where A is the $1,1$ element, B the $1,2:n$ elements, C the $2:n,1$ and D the $2:n,2:n$ elements of $\Sigma_T = I - p\tilde{Q}$. Thus, A represents the variance of the first area, B and C the correlation between the unsampled and the sampled areas, and D the correlation between the sampled areas.

Proof*:* Model (5.2) follows with the block matrix inversion formula (5.3):

$$\begin{aligned}
\Sigma &= \left(0, \text{diag}_{k-1}(\sigma^2_{\epsilon,i})\right) + \Sigma_T^{-1} W \sigma^2_u \\
&= \begin{bmatrix} (A - BD^{-1}C)^{-1}W_{\text{US}}\sigma^2_u & -A^{-1}B(D - CA^{-1}B)^{-1}W_{\text{S}}\sigma^2_u \\ -D^{-1}C(A - BD^{-1}C)^{-1}W_{\text{US}}\sigma^2_u & \sigma^2_{\epsilon,\text{S}} I_{\text{S}} + (D - CA^{-1}B)^{-1}W_{\text{S}}\sigma^2_u \end{bmatrix} \\
&=: \begin{bmatrix} \Sigma_{\text{US}} & \Sigma_{\text{US,S}} \\ \Sigma_{\text{S,US}} & \Sigma_{\text{S}} \end{bmatrix}.
\end{aligned}$$

Thus:

$$\begin{aligned}
\bar{\mu} &= X_{\text{US}}\beta + \Sigma_{\text{US},S}\Sigma_{\text{S}}^{-1}(y_{\text{S}} - X_{\text{S}}\beta) \\
&= X_{\text{US}}\beta - A^{-1}B(D - CA^{-1}B)^{-1}W_{\text{S}}\sigma^2_u \left[\sigma^2_{\epsilon,\text{S}} I_{\text{S}} + (D - CA^{-1}B)^{-1}W_{\text{S}}\sigma^2_u\right]^{-1}(y_{\text{S}} - X_{\text{S}}\beta).
\end{aligned}$$

Using the fact that $M^{-1}N^{-1} = (NM)^{-1}$ with $M = (D - CA^{-1}B)^{-1}W_{\mathrm{S}}$, $N = \left[\sigma^2_{\epsilon,\mathrm{S}} I_{\mathrm{S}} + (D - CA^{-1}B)^{-1}W_{\mathrm{S}}\sigma^2_u\right]^{-1}$ and $A^{-1} = 1$ it follows that:

$$\begin{aligned}\bar{\mu} &= X_{\mathrm{US}}\beta - \sigma^2_u A^{-1}B\left[(\sigma^2_{\epsilon,\mathrm{S}} I_{\mathrm{S}} + (D - CA^{-1}B)^{-1}\sigma^2_u W_{\mathrm{S}}) \cdot W_{\mathrm{S}}^{-1}(D - CA^{-1}B)\right]^{-1}(y_{\mathrm{S}} - X_{\mathrm{S}}\beta) \\ &= X_{\mathrm{US}}\beta - \sigma^2_u B\left[\sigma^2_{\epsilon,\mathrm{S}} I_{\mathrm{S}} W_{\mathrm{S}}^{-1}(D - CB) + \sigma^2_u I_{\mathrm{S}}\right]^{-1}(y_{\mathrm{S}} - X_{\mathrm{S}}\beta).\end{aligned}$$

The following example clarifies the meaning of formula (5.4).

Example 5.2: In this example, the mean $\bar{\mu}$ of formula (5.4) will be calculated for a situation with 3 areas, where the first area is unsampled. A nearest neighbor structure is assumed, which means that the first area is a neighbor of the second, the second area is a neighbor of the first and the third area, and finally the third area has got area two as a neighbor (see Figure 6 on page 35). Therefore, the neighborhood matrix Q is as follows:

$$Q = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Dividing each row of Q by the number of neighbors $\tilde{Q}_{i,j} = \dfrac{Q_{i,j}}{\sum_{j=1}^{3} Q_{i,j}}$ yields:

$$\tilde{Q} = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}.$$

The weight matrix $W = \mathrm{diag}\left(\dfrac{1}{\sum_{j=1}^{3} Q_{i,j}}\right)$ is given by:

$$W = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The mean of one unsampled area shall be calculated using formula (5.4). Since the first area is unsampled, W_{S} are the weights for the second and third area, given by:

$$W_{\mathrm{S}} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}.$$

Now $\bar{\mu}$ may be calculated using formula (5.4). Note that $\sigma^2_{\epsilon,\mathrm{S}} = \begin{bmatrix} \sigma^2_{\epsilon,2} & \sigma^2_{\epsilon,3} \end{bmatrix}$ and then:

$$\sigma^2_{\epsilon,\mathrm{S}} I_{\mathrm{S}} W_{\mathrm{S}}^{-1} = \begin{bmatrix} 2\sigma^2_{\epsilon,2} & 0 \\ 0 & \sigma^2_{\epsilon,3} \end{bmatrix}. \tag{5.5}$$

It follows that:

$$I - p\tilde{Q} = \begin{bmatrix} 1 & -p & 0 \\ -\frac{p}{2} & 1 & -\frac{p}{2} \\ 0 & -p & 1 \end{bmatrix} =: \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \tag{5.6}$$

where $A = 1$, $B = \begin{bmatrix} -p & 0 \end{bmatrix}$, $C = \begin{bmatrix} -\frac{p}{2} \\ 0 \end{bmatrix}$, and $D = \begin{bmatrix} 1 & -\frac{p}{2} \\ -p & 1 \end{bmatrix}$.

Using formula (5.6) yields:

$$D - CA^{-1}B = \begin{bmatrix} 1 - p^2/2 & -p/2 \\ -p & 1 \end{bmatrix}. \tag{5.7}$$

Using (5.5) and (5.7) it follows that:

$$\sigma^2_{\epsilon,\mathrm{S}} I_{\mathrm{S}} W_{\mathrm{S}}^{-1}(D - CA^{-1}B) + \sigma^2_u I_{\mathrm{S}} = \begin{bmatrix} 2\sigma^2_{\epsilon,2}(1 - p^2/2) + \sigma^2_u & -\sigma^2_{\epsilon,2} p \\ -p\sigma^2_{\epsilon,3} & \sigma^2_{\epsilon,3} + \sigma^2_u \end{bmatrix}.$$

Therefore:

$$(\sigma^2_{\epsilon,\mathrm{S}} I_{\mathrm{S}} W_{\mathrm{S}}^{-1}(D - CA^{-1}B) + \sigma^2_u I_{\mathrm{S}})^{-1} = \frac{1}{m} \begin{bmatrix} \sigma^2_{\epsilon,3} + \sigma^2_u & \sigma^2_{\epsilon,2} p \\ p\sigma^2_{\epsilon,3} & 2\sigma^2_{\epsilon,2}(1 - p^2/2) + \sigma^2_u \end{bmatrix},$$

where:

$$m = (2\sigma^2_{\epsilon,2}(1 - p^2/2) + \sigma^2_u)(\sigma^2_{\epsilon,3} + \sigma^2_u) - \sigma^2_{\epsilon,2}\sigma^2_{\epsilon,3} p^2.$$

Now all the necessary parts to calculate $\bar{\mu}$ have been found. Thus:

$$\begin{aligned} \bar{\mu} &= X_{\mathrm{US}}\beta - \sigma^2_u B \left[\sigma_{\epsilon,\mathrm{S}} I_{\mathrm{S}} W_{\mathrm{S}}^{-1}(D - CA^{-1}B) + \sigma^2_u I_{\mathrm{S}}\right]^{-1} (y_{\mathrm{S}} - X_{\mathrm{S}}\beta) \\ &= X_{\mathrm{US}}\beta - \sigma^2_u \begin{bmatrix} -p & 0 \end{bmatrix} \frac{1}{m} \begin{bmatrix} \sigma^2_{\epsilon,3} + \sigma^2_u & \sigma^2_{\epsilon,2} p \\ p\sigma^2_{\epsilon,3} & 2\sigma^2_{\epsilon,2}(1 - p^2/2) + \sigma^2_u \end{bmatrix} (y_{\mathrm{S}} - \mu_{\mathrm{S}}). \end{aligned}$$

Matrix calculation yields:

$$\bar{\mu} = X_{\mathrm{US}}\beta + \frac{p}{m} \begin{bmatrix} (\sigma^2_{\epsilon,3} + \sigma^2_u) & p\sigma^2_{\epsilon,2} \end{bmatrix} (y_{\mathrm{S}} - X_{\mathrm{S}}\beta). \tag{5.8}$$

Out of formula (5.8) the following observations can be made:

1. The resulting estimate is a linear combination between the synthetic estimate $X_{\mathrm{US}}\beta$ and information of the other areas $(y_{\mathrm{S}} - X_{\mathrm{S}}\beta)$.
2. Weight is given to neighbors (area 2) and to non-neighbors (area 3).
3. Since $\sigma_u^2 > 0$ and $p < 1$, it follows that if $\sigma_{\epsilon,3}^2$ and $\sigma_{\epsilon,2}^2$ are of equal size, more power is given to the neighborhood area 2.
4. If $\sigma_{\epsilon,2}^2$ is large and the information of the second area is low, then more strength is taken from area 3 and vice versa.
5. If $p = 0$ and independence is assumed, just the synthetic estimate will be used.

5.2.3 Propriety of the Posterior Distribution

In applications the spatial correlation term, p, is frequently assumed to equal 1. Unfortunately, this leads to an improper prior distribution on the random effects u, the so called intrinsic CAR model (cf. Besag et al., 1991 and Besag and Kooperberg, 1995). Thus, the propriety of the posterior distribution is not ensured and similar to Chapters 3 and 4 an existence theorem is needed.

Remark 5.3: The structure of the spatial hierarchical Fay-Herriot model slightly differs from the structure of the models of Chapters 3 and 4. The prior distribution of the variance component of the Fay-Herriot model is proper. Possible propriety problems enter through the improper intrinsic CAR structure. The other two models assume a different, proper, CAR structure. Problems might occur because the prior distributions of the variance components are improper.

Sun et al., 1999, p. 346 considered the propriety for the intrinsic CAR model with unknown, but equal sampling variances. Let $Y = (Y_1, \cdots, Y_n)$ be the vector of n observations and let X and Z be the $n \times p$ and $n \times k$ design matrices. The least squares estimator for (β', u')

is given by $(\hat{\beta}, \hat{u}) = ((X,Z)'(X,Z))^{-}(X,Z)'Y$, where $((X,Z)'(X,Z))^{-}$ is a generalized inverse of $(X,Z)'(X,Z)$. Finally, let $SSE = Y'\{I_n - (X,Z)((X,Z)'(X,Z))^{-}\}Y$ be the sum of squared errors. Then, the following theorem holds (cf. Sun et al., 1999, p. 346).

Theorem 5.4: *Consider the linear mixed model:*

$$Y = X\beta + Zu + \epsilon,$$

where $\epsilon \sim N(0, \sigma_\epsilon^2 I)$. *Assume the prior densities* $f(u) \propto \exp\left(-\frac{1}{2\sigma_u^2}u'Bu\right)$, *where* B *is nonnegative definite but not positive definite,* $\pi_\beta(\beta) \propto 1$, $\pi_{\sigma_\epsilon^2}(\sigma_\epsilon^2) \propto (\sigma_\epsilon^2)^{-(a_\epsilon+1)}\exp(-b_\epsilon/\sigma_\epsilon^2)$ *and* $\pi_{\sigma_u^2}(\sigma_u^2) \propto (\sigma_u^2)^{-(a_u+1)}\exp(-b_u/\sigma_u^2)$. *The variance components are assumed to be a priori independent. Assume the following conditions:*

- $rank(X) = p$ *and* $rank(u'R_1u + B) = k$, *where* $R_1 = I_n + X(X'X)^{-1}X'$
- $a_u > 0$ *and* $b_u > 0$
- $n - p - k - 2a_\epsilon > 0$ *and* $SSE + 2b_\epsilon > 0$

Then, the joint posterior distribution of $(\beta, Z, \sigma_\epsilon^2, \sigma_u^2)$ *given* Y *is proper.*

In this theorem, the sampling variances are assumed to be *unknown* but *equal*. Therefore, this theorem does not ensure propriety for a Fay-Herriot type model with *known* but *unequal* sampling variances. However, Theorem 5.4 can be adapted to the spatial general linear mixed model with known, unequal sampling variances, which includes the spatial hierarchical Fay-Herriot model (5.1).

Theorem 5.5: *Consider the linear mixed model* $Y = X\beta + Zu + \epsilon$, *where* $\epsilon \sim N(0, \Sigma_\epsilon)$ *with known sampling variance matrix* Σ_ϵ. *In addition, we assume the following prior distributions:* $f(u) \propto \exp\left(-\frac{1}{2\sigma_u^2}u'Bu\right)$, *where* B *is nonnegative definite but not positive definite,* $\pi_\beta(\beta) \propto 1$, *and* $\pi_{\sigma_u^2}(\sigma_u^2) \propto (\sigma_u^2)^{-(a+1)}\exp(-b/\sigma_u^2)$. *Assume the following conditions:*

- $rank(X) = p$ *and* $rank(u'R_1u + B) = k$, *where* $R_1 = \Sigma_\epsilon^{-1} + \Sigma_\epsilon^{-1}X(X'\Sigma_\epsilon^{-1}X)^{-1}X'\Sigma_\epsilon^{-1}$

- $a > 0$ *and* $b > 0$.

Then, the joint posterior distribution of (β, u, σ_u^2) *given* Y *is proper.*

Proof*:* The idea is to integrate the joint posterior density of (β, u, σ_u^2) with respect to the three variables: β, u, and σ_u^2. The joint posterior density is proportional to:

$$G = (\sigma_u^2)^{-\frac{1}{2}k} \cdot \exp\left(-\frac{1}{2}(Y - X\beta - Zu)'\Sigma_\epsilon^{-1}(Y - X\beta - Zu) - \frac{u^T B u}{2\sigma_u^2}\right) \cdot \pi_{\sigma_u^2}(\sigma_u^2).$$

The proof is split into six parts. In parts 1 and 2 the joint posterior is transformed and then integrated with respect to β. In parts 2 and 3 the integrated posterior is rearranged to allow for an easier integration with respect to u. Finally, in parts 5 and 6 the joint posterior is bounded, and thus, the propriety is shown.

1. First, G is transformed since this helps to better handle the integration with respect to β. This is done by adding and subtracting $X\hat{\beta}$ and $Z\hat{u}$. It follows that:

$$\begin{aligned} & (Y - X\beta - Zu)'\Sigma_\epsilon^{-1}(Y - X\beta - Zu) \\ = \; & (Y - X\hat{\beta} - Z\hat{u} - X(\beta - \hat{\beta}) - Z(u - \hat{u}))'\Sigma_\epsilon^{-1} \cdot \\ & \cdot(Y - X\hat{\beta} - Z\hat{u} - X(\beta - \hat{\beta}) - Z(u - \hat{u})). \end{aligned} \tag{5.9}$$

   Expanding (5.9) yields:

$$\begin{aligned} & (Y - X\beta - Zu)'\Sigma_\epsilon^{-1}(Y - X\beta - Zu) \\ = \; & e'\Sigma_\epsilon^{-1}e - \\ & - e'\Sigma_\epsilon^{-1}X(\beta - \hat{\beta}) + \\ & + (\beta - \hat{\beta})'X'\Sigma_\epsilon^{-1}X(\beta - \hat{\beta}) - \\ & - (\beta - \hat{\beta})'X'\Sigma_\epsilon^{-1}e + \\ & + (\beta - \hat{\beta})'X'\Sigma_\epsilon^{-1}Z(u - \hat{u}) + (u - \hat{u})'Z'\Sigma_\epsilon^{-1}X(\beta - \hat{\beta}) - \\ & - e'\Sigma_\epsilon^{-1}Z(u - \hat{u}) - \\ & - (u - \hat{u})'Z'\Sigma_\epsilon^{-1}e + \\ & + (u - \hat{u})'Z'\Sigma_\epsilon^{-1}Z(u - \hat{u}), \end{aligned} \tag{5.10}$$

   where $e = Y - X\hat{\beta} - Z\hat{u}$.

2. Now all of the factors in (5.10) containing $(\beta - \hat{\beta})$ are collected, such that:

$$(Y - X\beta - Zu)'\Sigma_\epsilon^{-1}(Y - X\beta - Zu) = (\beta - \hat{\beta} - C_0 - C_1)'X'\Sigma_\epsilon^{-1}X(\beta - \hat{\beta} - C_0 - C_1) + \mathrm{K}_0,$$

where

$$\begin{aligned} C_0 &= (X'\Sigma_\epsilon^{-1}X)^{-1}X'\Sigma_\epsilon^{-1}e \\ C_1 &= (X'\Sigma_\epsilon^{-1}X)^{-1}X'\Sigma_\epsilon^{-1}Z(u - \hat{u}). \end{aligned}$$

Note that $X'\Sigma_\epsilon^{-1}X$ is symmetric. K_0 is a constant which contains all the factors independent of β. We find that:

$$\begin{aligned} \mathrm{K}_0 &= (u - \hat{u})'Z'\Sigma_\epsilon^{-1}Z(u - \hat{u}) - (u - \hat{u})'Z'\Sigma_\epsilon^{-1}e - \\ &- e'\Sigma_\epsilon^{-1}Z(u - \hat{u}) - C_1'X'\Sigma_\epsilon^{-1}XC_0 - \\ &- C_1'X'\Sigma_\epsilon^{-1}XC_1 - C_0'X'\Sigma_\epsilon^{-1}XC_1 - C_0'X'\Sigma_\epsilon^{-1}XC_0 \end{aligned}$$

Integrating G with respect to β yields:

$$\int_{\mathbb{R}^p} G \; d\beta = \frac{(2\pi)^{\frac{1}{2}p}|X'\Sigma_\epsilon^{-1}X|^{-\frac{1}{2}}}{(\sigma_u^2)^{\frac{1}{2}k}} \exp\left\{-\frac{1}{2}\mathrm{K}_0 - \frac{u'Bu}{2\sigma_u^2}\right\} \cdot \pi_{\sigma_u^2}(\sigma_u^2). \tag{5.11}$$

3. Now, we can compute the integration with respect to u. Therefore, the exponential function (5.11) is calculated and transformed by collecting the terms containing u in K_0. The following facts are useful:

   (a) $C_0'X'\Sigma_\epsilon^{-1}XC_0$ which is independent of all integration variables and will be seen as a constant.

   (b) $C_0'X'\Sigma_\epsilon^{-1}XC_1 = ((X'\Sigma_\epsilon^{-1}X)^{-1}X'\Sigma_\epsilon^{-1}e)'X'\Sigma_\epsilon^{-1}Z(u - \hat{u})$

   (c) $C_1'X'\Sigma_\epsilon^{-1}XC_1 = (u - \hat{u})'Z'\Sigma_\epsilon^{-1}X(X'\Sigma_\epsilon^{-1}X)^{-1}X'\Sigma_\epsilon^{-1}Z(u - \hat{u})$

   (d) $C_1'X'\Sigma_\epsilon^{-1}XC_0 = -((X'\Sigma_\epsilon^{-1}X)^{-1}X'\Sigma_\epsilon^{-1}Z(u - \hat{u}))'X'\Sigma_\epsilon^{-1}e$

   (e) $e'\Sigma_\epsilon^{-1}Z(u - \hat{u})$

   (f) $(u - \hat{u})'Z'\Sigma_\epsilon^{-1}e$

   (g) $(u - \hat{u})'Z'\Sigma_\epsilon^{-1}Z(u - \hat{u})$.

This leads to :

$$\mathrm{K}_0 = (u - \hat{u} - C_2)' Z' R_1 Z (u - \hat{u} - C_2) + \mathrm{K}_1, \tag{5.12}$$

where $C_2 = (Z' R_1 Z)^{-1} Z' (R_1 + \Sigma_\epsilon^{-1}) e$ includes the above terms with one $(u - \hat{u})$. Then K_1 contains all terms independent of u which occur by including C_2 in the formula.

4. Now, we consider $\dfrac{u'Bu}{\sigma_u^2}$. Therefore, formula (5.12) of the previous step is utilized, since:

$$\mathrm{K}_0 + \frac{u'Bu}{\sigma_u^2} = (u - \hat{u} - C_2)' Z' R_1 Z (u - \hat{u} - C_2) + \mathrm{K}_1 + \frac{u'Bu}{\sigma_u^2}.$$

This leads to:

$$u' Z' R_1 Z u + \frac{u'Bu}{\sigma_u^2} + \text{other terms}. \tag{5.13}$$

Rearranging (5.13) yields:

$$(u - C_3)' R_2 (u - C_3) - (C_3)' R_2 C_3, \tag{5.14}$$

where $R_2 = Z' R_1 Z + \dfrac{B}{\sigma_u^2}$ and $C_3 = R_2^{-1} Z' R_1 Z (\hat{u} + C_2)$.
Using the integrated G in (5.11) together with (5.14) it follows that:

$$\int_{\mathbb{R}^k} \int_{\mathbb{R}^p} G \; d\beta du = \frac{(2\pi)^{\frac{1}{2}(p+k)} |X' \Sigma_\epsilon^{-1} X|^{-\frac{1}{2}}}{(\sigma_u^2)^{\frac{1}{2}k + a + 1} |R_2|^{\frac{1}{2}}} \exp\left\{ -(C_3)' R_2 C_3 - \frac{b}{\sigma_u^2} \right\}. \tag{5.15}$$

Integration leads to the factor $|R_2|^{-\frac{1}{2}}$ in front of the exponential function. Since $Z' R_1 Z R_2^{-1} Z' R_1 Z$ in $(C_3)' R_2 C_3$ is nonnegative definite this term can be bounded by 0 and thus discarded from the integral.

5. Using the argument by Sun et al., 1999, p. 346 it follows that:

$$\begin{aligned} |R_2|^{-\frac{1}{2}} &\leq \{\min(1, (\sigma_u^2)^{-1})^k \cdot |Z' R_1 Z + B|\}^{-\frac{1}{2}} \\ &< (1 + (\sigma_u^2)^{\frac{k}{2}}) \cdot |Z' R_1 Z + B|^{-\frac{1}{2}}. \end{aligned} \tag{5.16}$$

6. Finally, combining (5.15) and (5.16) yields:

$$\int_{\mathbb{R}^k} \int_{\mathbb{R}^p} G \; d\beta du \leq (2\pi)^{p+k} |X' \Sigma_\epsilon^{-1} X|^{-\frac{1}{2}} |Z' R_1 Z + B|^{-\frac{1}{2}} (J_1 + J_2),$$

where

$$J_1 = \frac{1}{(\sigma_u^2)^{\frac{1}{2}k+a+1}} \exp\left(-\frac{b}{\sigma_u^2}\right),$$

and

$$J_2 = \frac{1}{(\sigma_u^2)^{a+1}} \exp\left(-\frac{b}{\sigma_u^2}\right).$$

Since $a > 0$ and $b > 0$ the integrals J_1 and J_2 exist with respect to σ_u^2.

This completes the proof. □

Notes 5.6:

1. In Theorem 5.4 by Sun et al. (1999), the sum of squared errors (SSE) arises in the assumptions. This term is not important here, since it depends only on the known variance-covariance matrix Σ_ϵ.

2. In the proof of Theorem 5.4, all terms containing e cancel out of formula (5.10), since SSE is orthogonal to e. This need not to be the case if Σ_ϵ is non-diagonal.

5.3 Application

We have now proved the propriety of the posterior distribution of the spatial hierarchical Fay-Herriot model and the necessary theory for spatial prediction has also been derived. In the following the spatial hierarchical Fay-Herriot model (5.1) is compared to the according non-spatial version in terms of prediction using real data.

5.3.1 Setup

The poverty ratio in children ages $5-17$ in families in poverty will be analyzed using data from the Current Population Survey (CPS) for the years 1989 and 1993. Four covariates are available and include the Internal Revenue Service (IRS) pseudo child poverty rate ($x1$), the IRS non-filer rate ($x2$), a food stamp participation rate ($x3$), and the census residuals ($x4$). Combining data with the covariates the U.S. Census Bureau produces Small Area

Income and Poverty Estimates (SAIPE). In this application those estimates are seen as a gold standard. Since the data is from the state level, a Fay-Herriot type model can be applied. In this application, 48 US States and the District of Columbia (Hawaii and Alaska omitted) are considered. Every area is left out once and is predicted by means of a spatial and non-spatial model. This procedure is repeated for covariates $(0-4)$ and for the years 1989 and 1993. The estimation results are compared to the official estimates (gold standard).

5.3.2 Results

Since the results are similar for the years 1989 and 1993, we will simply just analyze the data for 1993. Tables 7 and 8 contain simulation results for 1993 and 1989 respectively. Column 1 of Table 7 shows different measures of comparison, based on the squared deviance between the predicted estimator and the official value, the absolute deviance, and the maximum of the deviance. Since each of the 49 areas is left out once, the deviances are averaged over all states. The deviances are constructed for the model containing all four covariates, no covariate or each of the covariates alone. Columns 2 and 3 contain the corresponding values for the spatial and non-spatial model. The last two columns compare the deviances of the spatial and non-spatial model with each other (difference and ratio). The following observations can be made:

1. If no covariates are included, the deviances for the spatial and non-spatial model are large. These values decrease as the quality and number of covariates increases. The lowest value is reached when all covariates are included.

2. The ratio of the non-spatial deviance compared to the spatial is large if no or weak covariates are included. The ratio decreases if the quality of the covariates improves.

3. If all covariates are included there is no gain by using the spatial model.

The same effects can be seen, in Figures 11, 12, and 13. These figures compare the predicted values of the spatial and non-spatial model with varying numbers of covariates.

	Spatial	**Non-sp.**	**Spatial − Non-sp.**	$\frac{\textbf{Non-sp.}}{\textbf{Spatial}}$
$\frac{1}{49} \cdot \sum_{i=1}^{49}((\textbf{Estimator}_i - \textbf{Official}_i)^2)$				
all	1.08	1.08	0.00	1.00
x1	7.76	12.51	-4.75	1.61
x2	17.04	27.43	-10.38	1.61
x3	4.74	4.59	0.15	0.97
x4	23.42	34.06	-10.64	1.44
without	25.14	34.18	-9.04	1.36
$\frac{1}{49} \cdot \sum_{i=1}^{49}(\lvert\textbf{Estimator}_i - \textbf{Official}_i\rvert)$				
all	0.79	0.79	0.00	1.00
x1	2.20	2.71	-0.51	1.23
x2	3.00	4.04	-1.04	1.35
x3	1.76	1.76	0.00	1.00
x4	3.76	4.77	-1.01	1.27
without	3.85	4.84	-0.99	1.26
$\textbf{max}(\lvert\textbf{Estimator} - \textbf{Official}\rvert)$				
all	0.07	0.07	0.00	1.00
x1	0.12	0.17	-0.05	1.40
x2	0.24	0.29	-0.05	1.21
x3	0.12	0.12	0.00	1.00
x4	0.26	0.32	-0.06	1.23
without	0.29	0.27	0.02	0.93

Table 7: SAIPE Simulation Results for 1993.

Figure 11 on page 80 shows that if all covariates are included, there is no visible difference between the spatial and non-spatial model. However, if no covariates are included, then the predicted values of the non-spatial model compared to the official values are almost constant. On the other hand, the spatial model improves the relationship. The same effect can be observed if covariates of a different quality are included (Figures 12 and 13). Figure

14 on page 83 underlines these results, by showing the squared deviance of the spatial and non-spatial model for all 4 covariates (upper plots) and no covariates (lower plots) on the map. If no covariates are included, the spatial model performs better than the non-spatial model. This effect diminishes if all covariates are included.

	Spatial	**Non-sp.**	**Spatial − Non-sp.**	$\frac{\textbf{Non-sp.}}{\textbf{Spatial}}$
$\frac{1}{49} \cdot \sum_{i=1}^{49}((\textbf{Estimator}_i - \textbf{True}_i)^2)$				
all	0.97	0.95	0.02	0.98
x1	4.53	4.67	-0.14	1.03
x2	12.14	20,95	-8,81	1.73
x3	3.89	5.23	-1.34	1.34
x4	17.27	28.12	-10.85	1.63
without	16.11	26.83	-10.72	1.67
$\frac{1}{49} \cdot \sum_{i=1}^{49}(\lvert\textbf{Estimator}_i - \textbf{True}_i\rvert)$				
all	0.82	0.81	0.01	0.98
x1	1.51	1.52	-0.01	1.01
x2	2.64	3.43	-0.80	1.30
x3	1.58	1.82	-0.24	1.15
x4	3.13	4.02	-0.89	1.28
without	3.16	3.98	-0.82	1.26
$\textbf{max}(\lvert\textbf{Estimator} - \textbf{True}\rvert)$				
all	0.05	0.05	0.00	1.00
x1	0.14	0.14	0.00	1.00
x2	0.25	0.32	-0.07	1.27
x3	0.09	0.10	-0.01	1.09
x4	0.30	0.36	-0.06	1.19
without	0.25	0.34	-0.09	1.35

Table 8: SAIPE Simulation Results for 1989.

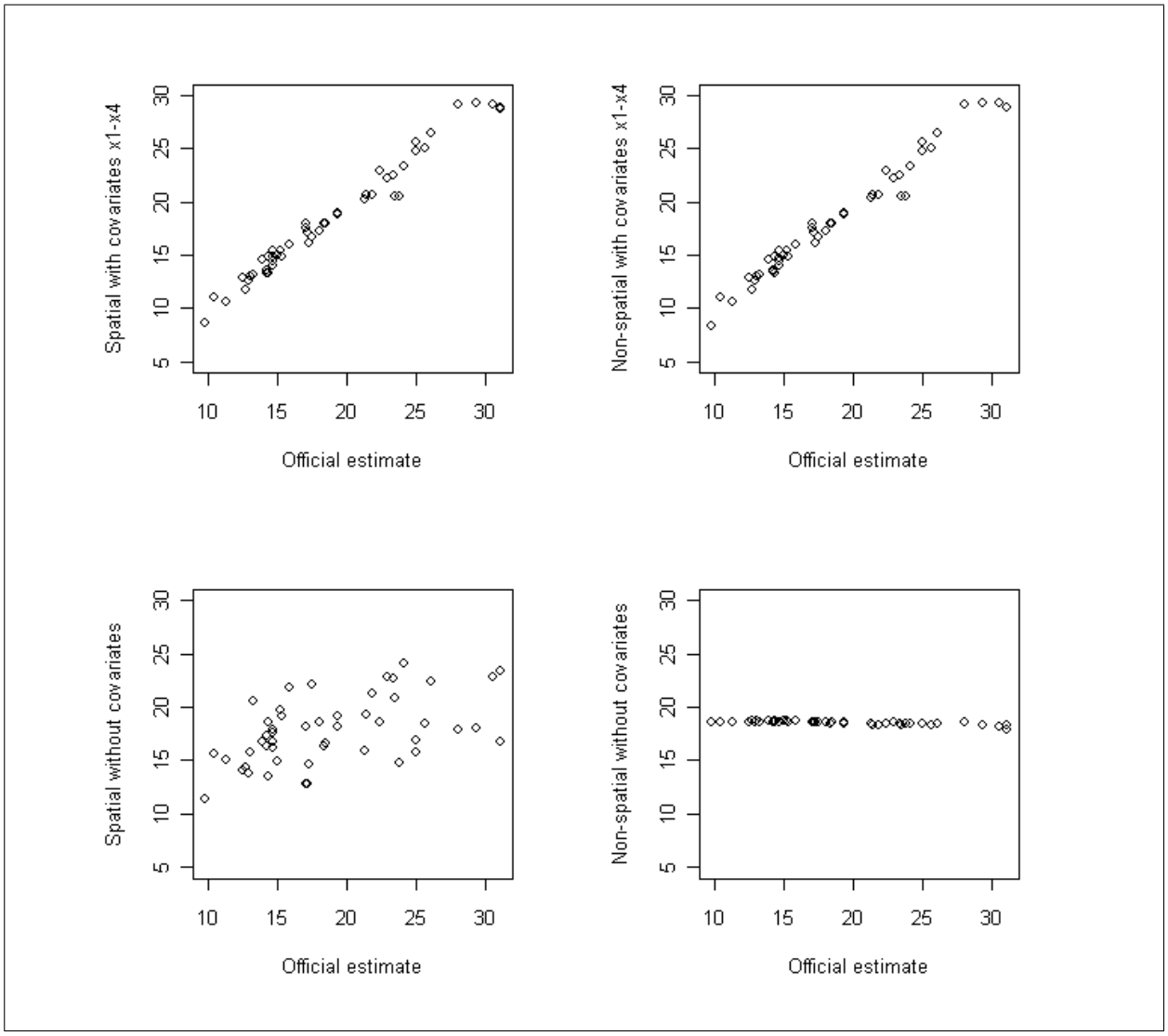

Figure 11: Predicted Values of the Spatial and Non-spatial FH Model Compared to the Official Estimates from 1993 : 4 and No Covariates.

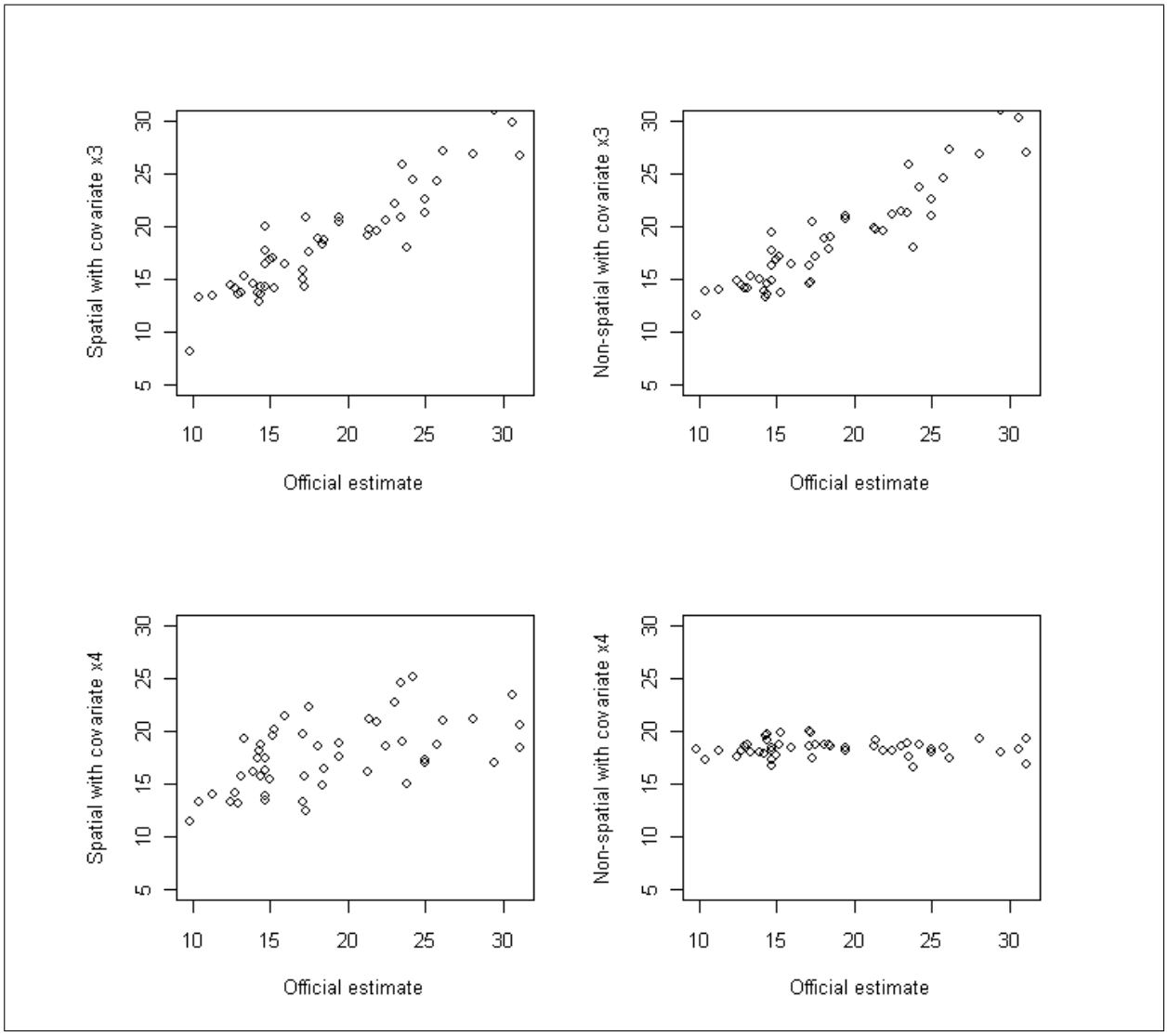

Figure 12: Predicted Values of the Spatial and Non-spatial FH Model Compared to the Official Estimates from 1993 : Covariates $x3$, $x4$.

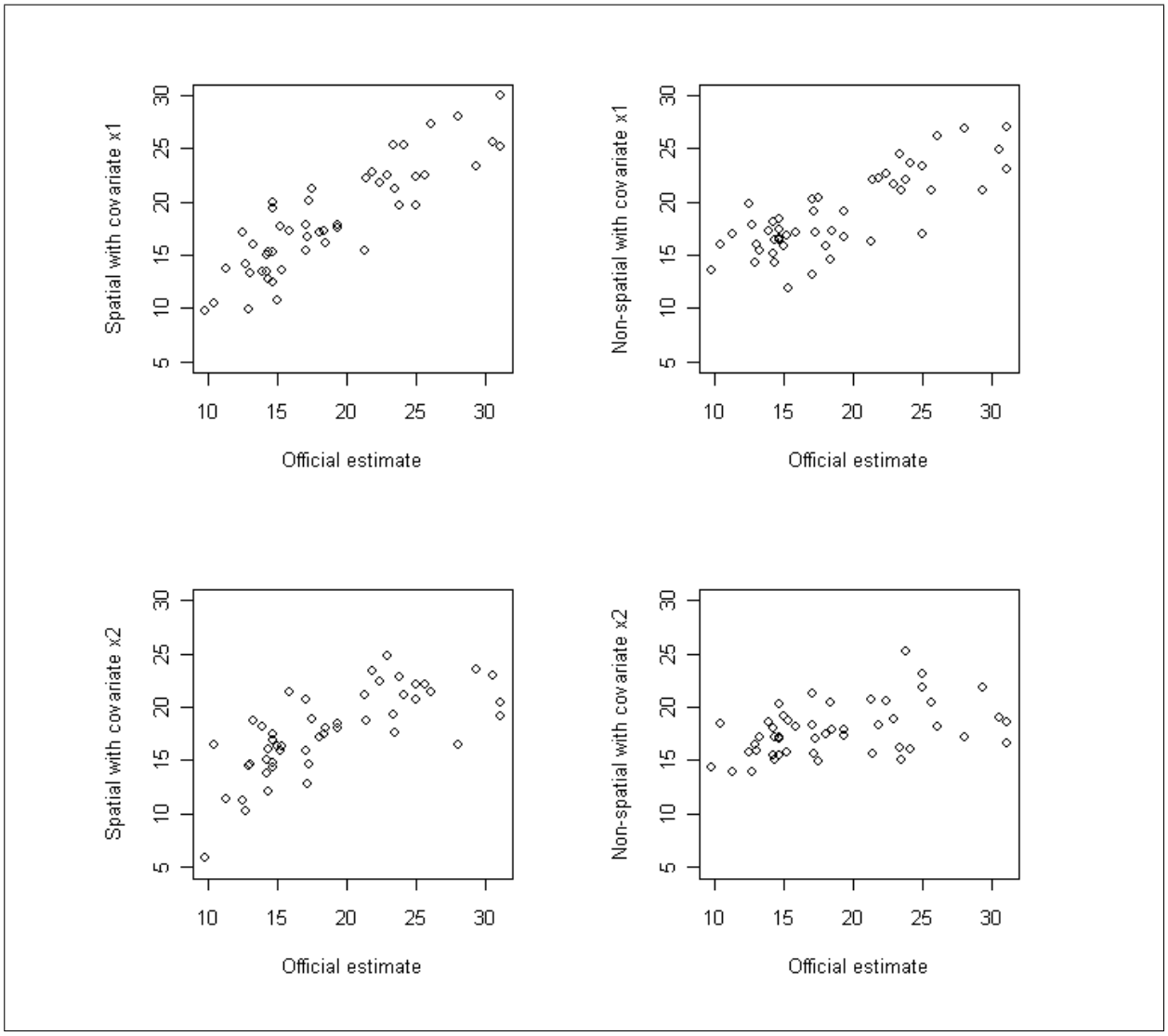

Figure 13: Predicted Values of the Spatial and Non-spatial FH Model Compared to the Official Estimates from 1993 : Covariates $x1$, $x2$.

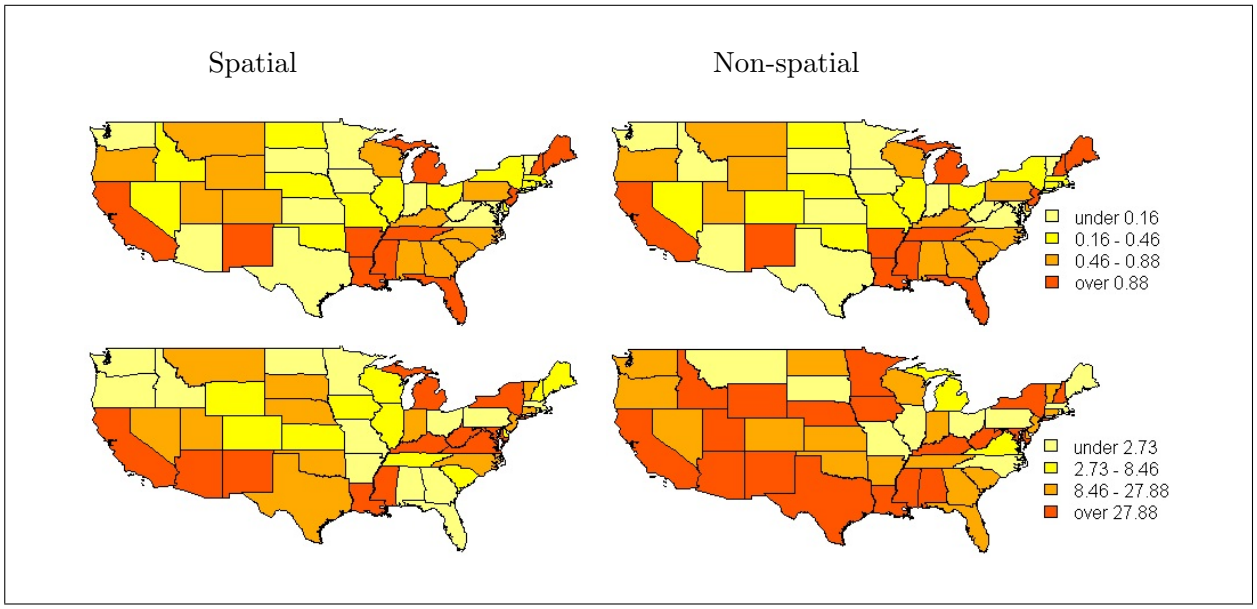

Figure 14: Squared Deviance of the Spatial and Non-spatial FH Model for 4 (Upper Plots) and No (Lower Plots) Covariates.

5.4 Conclusion

In Chapter 3, we proved propriety of the posterior distribution for the spatial general linear mixed model. This model class includes many area as well as unit-level models. However, the Fay-Herriot model is not part of this class. We have reached a resolution in this chapter: the Fay-Herriot model has been extended to allow for spatial correlation and the propriety of the posterior distribution has been proved. Moreover, the effect of covariates on the estimation results has been tested. An application to SAIPE data revealed that modeling spatial correlation can considerably improve the associated hierarchical Bayes methodology if the area specific auxiliary data are either weak or not available. This effect diminishes if the quality of the covariates improves. Within the German Census 2011 every geographical area will be sampled, but this might not be the case for areas formed for example by social factors, called small domains. One application revealed that in this case spatial modeling is an interesting alternative to standard modeling.

6 Simulation Study

Throughout our work, different spatial models have been applied by proving the propriety of the posterior distribution. Such models include the spatial hierarchical linear mixed models in Chapter 3, the spatial Bernoulli mixed models in Chapter 4, and the spatial hierarchical Fay-Herriot model in Chapter 5. These classes include the spatial hierarchical extension of the Fay-Herriot and Battese-Harter-Fuller model. Both model types are examined for a possible use within the German Census 2011 (cf. Münnich et al., 2007, p. 2). Since many variables within the German Census 2011 are categorical, they may be also modeled via the spatial Bernoulli mixed model. In this chapter, the normal and Bernoulli unit-level models, as well as the hierarchical Fay-Herriot model are tested and compared with their spatial versions, using data from the public use file of the German Mikrocensus (2002).

6.1 Introduction and Aim

The German Mikrocensus (MC) is an annual, representative one percent sample survey where 1% of the households are sampled. The aim of the Mikrocensus is to provide statistical information about the economic and social situation of Germany, as well as information about employment, labor market and education. Further details are given on the websites Destatis (2010) and GESIS (2010). The MC data set is chosen due to the following four reasons:

1. The structure of the Microcensus resembles the unavailable census data. Therefore, similar problems, except outdated and missing entries, can be modeled.

2. Unit-level information is available, which is necessary for the implementation of the models of Chapters 3 and 4.

3. It is possible to identify areas, like federal states, and locate them on a map. This allows the use of spatial models.

4. The data is available for free (cf. German Mikrocensus, 2002). Together with the provided WinBUGS models in Appendix B, this allows scientists to reproduce the presented results.

Throughout the work we showed that spatial models include extra geographical information in an elegant way. The aim of the simulation study is to analyze how the extra spatial information influences the estimation results. In addition, we shall test how the estimation results are affected by different choices of the sample size, the neighborhood structure, the prior distributions, and the spatial correlation parameter.

6.2 Dataset Description

The public use file German Mikrocensus (2002) contains 335 variables for 25,137 individuals out of 11,655 households. In order to allow an implementation of the normal as well as the Bernoulli model of Chapters 3, 4, and 5 a categorical variable with two realizations is chosen. This variable has number EF288 and stands for:

Beruflicher Ausbildungs- oder Hochschul-/ Fachhochschulabschluss vorhanden?

(Professional training or University/College degree?)

To shorten, this variable will be denoted by education. The categories of this variable are yes, no, no answer, and not applicable (children below 15 years). Since missing values and imputation extend the intension of this simulation study, persons between 16 and 65 years old are considered. This leads to a data set containing 16,153 individuals. Table 9 shows how the individuals are divided between the federal states.

The area mean of the variable education at the federal state level is shown in Figure 15. In the plot, red colors represent highly educated states. It may be noted that people from former Eastern Germany tend to be less educated than the Southern regions, and thus, a spatial trend is visible. This is probably due to the relocation of the highly educated from East to West Germany after the reunion of the states in 1990 (cf. Spiegel, 2003, p. 24).

Federal state	Size	Federal state	Size
Berlin	634	Lower Saxony	1506
Bremen	128	North Rhine Westph.	3374
Brandenburg	538	Rhineland-Palatinate	817
Baden-Württ.	2014	Saxony-Anhalt	548
Bavaria	2530	Saxony	886
Hamburg	328	Schleswig-Holstein	592
Hesse	1242	Saarland	210
Mecklenburg-Vorp.	321	Thuringia	485

Table 9: Size of the Transformed Mikrocensus Dataset per German Federal State.

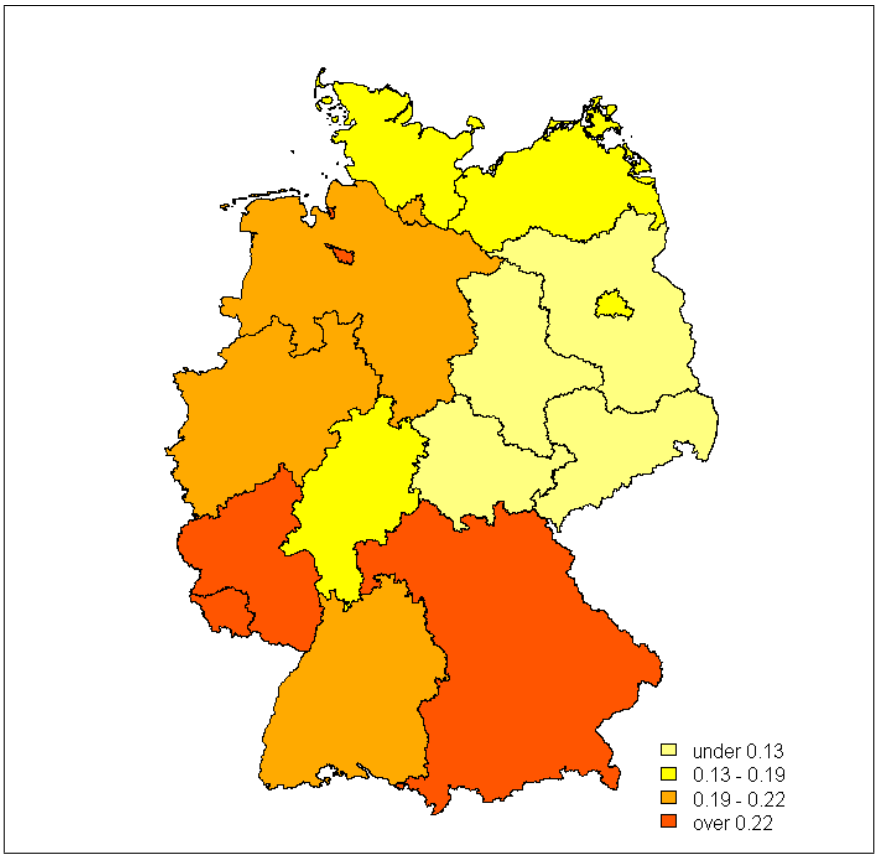

Figure 15: Professional Training or University/College Degree?

In all models of the simulation study, the following two covariates are used: gender (EF32)

and age (EF33). Both auxiliary variables are chosen, because they are typically available in administrative registers and are thus usable within the context of the German Census 2011. The variable gender is categorized into male and female. As denoted above, the range of age is from 16 to 65.

6.3 Simulation Setup

In this section the models, setups, and measures are described.

6.3.1 Models and Propriety

In the simulation study, three models and their spatial extensions are considered. The first model is a hierarchical version of the Battese-Harter-Fuller model (2.6), contained in model (2.11). Recall that:

$$\begin{aligned}
Y_{ij} &= X_{ij}\beta + u_i + \epsilon_{ij} \\
u \mid \sigma_u^2 &\sim N_k(0, \sigma_u^2 I) \\
\epsilon_{ij} &\overset{\text{i.i.d.}}{\sim} N(0, \sigma_\epsilon^2) \qquad i = 1, \cdots, k; \quad j = 1, \cdots, n_i \\
\pi_\beta(\beta) &\propto 1 \\
\pi_{\sigma_\epsilon^2}(\sigma_\epsilon^2 \mid b) &\propto (\sigma_\epsilon^2)^{-(b+1)} \\
\pi_{\sigma_u^2}(\sigma_u^2 \mid a) &\propto (\sigma_u^2)^{-(a+1)}.
\end{aligned}$$

The spatial extension, see model (3.1), is the same except that the distribution of the random effects is substituted by the CAR structure (2.20), where:

$$u \mid \sigma_u^2, p \sim N_k(0, \sigma_u^2 (I - p\tilde{Q})^{-1} W). \tag{6.1}$$

The propriety of these models is discussed and proved in Chapter 3. The second model type is the Bernoulli model of Chapter 4, where:

$$\begin{aligned}
y_i | u &\sim \text{Bernoulli}\{h(x_i\beta + z_i u)\} \\
u | \sigma_u^2 &\sim N_k(0, \sigma_u^2 I) \\
\pi_{\sigma_u^2}(\sigma_u^2 | a) &\propto \frac{1}{(\sigma_u^2)^{a+1}},
\end{aligned}$$

and is introduced by Natarajan and McCulloch (1995). In the spatial version, the random effect is distributed according to (6.1). In this study, both models utilize the logit link. In addition, a proper normal prior is assumed on the fixed effect β. The last model pair considered are the hierarchical Fay-Herriot models of Chapter 5. The non-spatial model is given by:

$$\begin{aligned} Y_i &\sim N(\theta_i, \sigma^2_{\epsilon,i}) \\ \theta_i &= X_i\beta + u_i \\ u|\sigma^2_u &\sim N(0, \sigma^2_u I) \\ \pi_\beta(\beta) &\propto 1 \\ 1/\sigma^2_u &\sim \Gamma(0.5, 0.0005). \end{aligned}$$

For the spatial extension (5.1), the distribution of the random effect is substituted by (6.1), where $p = 1$. The propriety is ensured by Theorem 5.5. In the Fay-Herriot model, the sampling variances are assumed to be known. This strong assumption will be retained throughout the simulation study, giving the Fay-Herriot model an advantage over the unit-level models. Thus, comparisons between the corresponding non-spatial and spatial versions are valid, but care should be taken when comparing different model types.

6.3.2 Setups

The following setups are tested:

1. Standard (Section 6.4.1): This setting is obtained by drawing a sample of size 20 out of each area, where the areas are the 16 German federal states. In addition, the two covariates age and gender are included. Furthermore, for the unit-level models $a = (b =) - 0.9$ and the spatial correlation parameter is set to 0.95. The parameters of the standard model have been chosen, because preliminary results for theses values behaved well (Sections 3.5 and 4.5). Throughout the simulation study this setting will be taken as standard. All the different setups are compared to the results of the six standard models. This is done, since due to capacity reasons (the computing time for one setup had been about 9 days), the number of setups needs to be limited.

2. In this setting three different simulations are done with 160 (10 sampled per area), 320 (20 per area) and 640 (40 per area) total sample sizes to check the influence of the sample size (Section 6.4.2).

3. The effect of different neighborhood structures is considered (Section 6.4.3). In this setting the neighborhood structure of Example 2.25 is compared to the nearest neighbor structure. The basic principle is still a nearest neighbor structure, but the city states and Saarland are separated from the other states and chosen to be neighbors of each other.

4. Also, the influence of the prior distributions (Section 6.4.4) is checked. The prior distributions of the variance components of the spatial and non-spatial Battese-Harter-Fuller model contain the unknown constants a and b. Similarly, the Bernoulli model includes the parameter a. In this setting the estimates are calculated for different values of a and b.

5. In this setting the impact of different values of the spatial correlation parameter p is tested (Section 6.4.5).

6. In this setting a dummy variable for East Germany (Section 6.4.6) is included into the non-spatial and spatial models and the results are compared to the models without dummy variable.

Common to all setups is that the German Mikrocensus data is seen as a gold standard, which means that the area means are known, where the areas are the 16 German federal states. In each setup 1000 samples are drawn. The sampling design is simple random sample and individuals (not households) are drawn. Further, 3 chains with a burn-in period of 10,000 are used. The iteration number is 20,000 and 2,000 simulations are saved per chain. Convergence is assessed if the scale reduction factor ($\hat{R}$) is below 1.1 (cf. Example 2.16). The simulation study is done on five computers with operating systems Linux, Windows XP, and Windows 7. The used simulation software is R and WinBUGS (via R2WinBUGS). During the simulation study WinBUGS tended to run unstable if more than one process was run parallel. Therefore, WinBUGS was installed twice in different

paths to ensure stability. In addition, WinBUGS under Windows performed more stable than under Linux.

6.3.3 Measures

The following will be used to measure the estimation results:

1. Relative Root Mean-Squared Error (RRMSE) for each area i is defined as:

$$\text{RRMSE}_i = \sqrt{\frac{1}{1000} \cdot \sum_{l=1}^{1000} \left(\frac{\hat{\theta}_{l,i} - \theta_i}{\theta_i} \right)^2}.$$

   The RRMSE measures the difference between the true value and the estimated value for each area on the scale of the estimated value and can take values between 0 and ∞.

2. The relative Bias is defined as:

$$\text{RBias}_i = \frac{\frac{1}{1000} \sum_{l=1}^{1000} \hat{\theta}_{l,i} - \theta_i}{\theta_i}.$$

   The relative Bias illustrates the bias of the estimated values over the different samples. RBias can take values between $-\infty$ and ∞, where negative values indicate in average underestimation of the true value, positive values overestimation, and 0 indicates unbiased estimation.

3. The relative Dispersion is defined as:

$$\text{RDisp}_i = \frac{\text{Quantile}(\hat{\theta}_{l=1\cdots1000,i}, 0.95) - \text{Quantile}(\hat{\theta}_{l=1\cdots1000,i}, 0.05)}{\theta_i}.$$

   The relative Dispersion measures the difference of the 0.05 quantile to the 0.95 quantile relative to the true value. RDisp can take values between 0 and ∞. The larger the relative Dispersion, the more scattered the estimated values per area over different samples will be.

6.4 Results

6.4.1 The Standard Models

The standard is obtained by drawing a sample of size 20 out of each area. Additionally, for the unit-level models, $a = (b =) - 0.9$, and the spatial correlation parameter is set to 0.95. Figure 16 shows the boxplots of the RRMSE for the 16 federal states for the six standard models, which are the non-spatial and spatial Battese-Harter-Fuller (BHF), the Bernoulli (Bern), and the Fay-Herriot (FH) models.

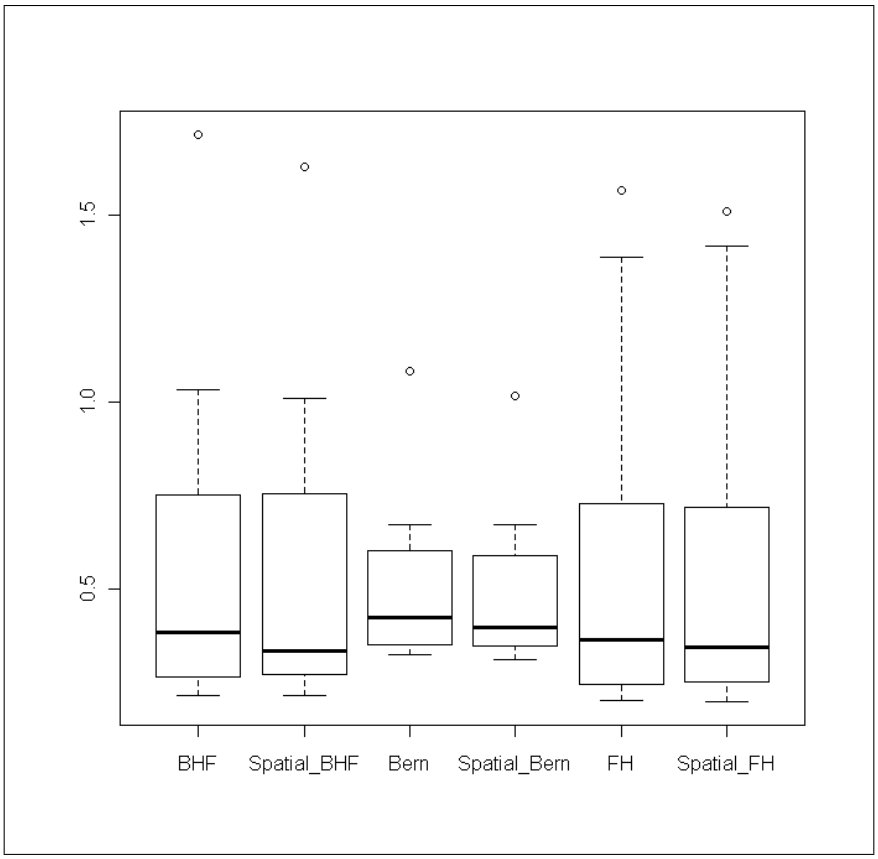

Figure 16: Boxplots of the RRMSE for the Standard Models.

The plot reveals that the median of the RRMSE is lower for the spatial models, compared to the corresponding non-spatial version. Table 10 shows the maximum of the RRMSE for the six models. Similar to the median, the maximum is lower for the spatial models. This is especially of interest if quality constraints, such as an allowed maximum of the RRMSE

per area are defined. The maximum is adopted in Saxony for the unit-level models, while it is adopted for the Fay-Herriot models in Brandenburg.

RRMSE	BHF	Bernoulli	FH
Non-sp.	1.72	1.09	1.57
Spatial	1.63	1.02	1.51

Table 10: Maximum of the RRMSE of the BHF, Bernoulli, and FH Standard Models.

Figures 17 and 18 plot the RRMSE of the spatial against the non-spatial models. The plots indicate that spatial modeling decreases the RRMSE for most areas (federal states) for the unit-level models, whereas the RRMSE of the spatial Fay-Herriot model is similar to the RRMSE of the non-spatial version.

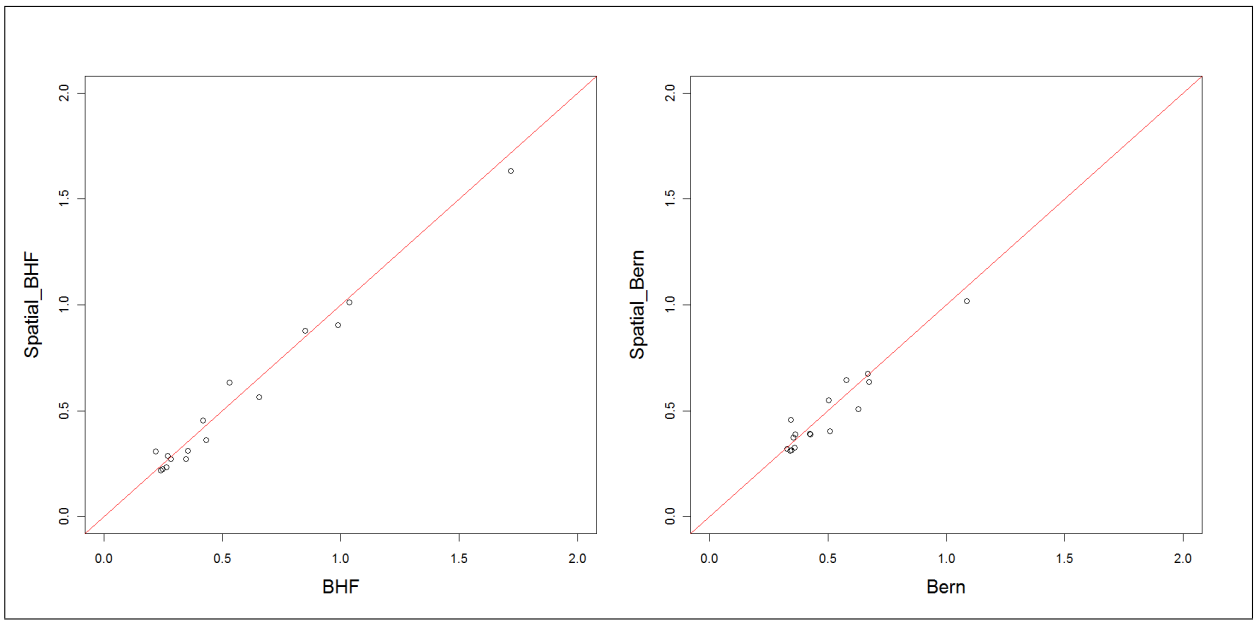

Figure 17: RRMSE for the 16 Federal States of the Spatial Against the Non-spatial BHF and Bern Model.

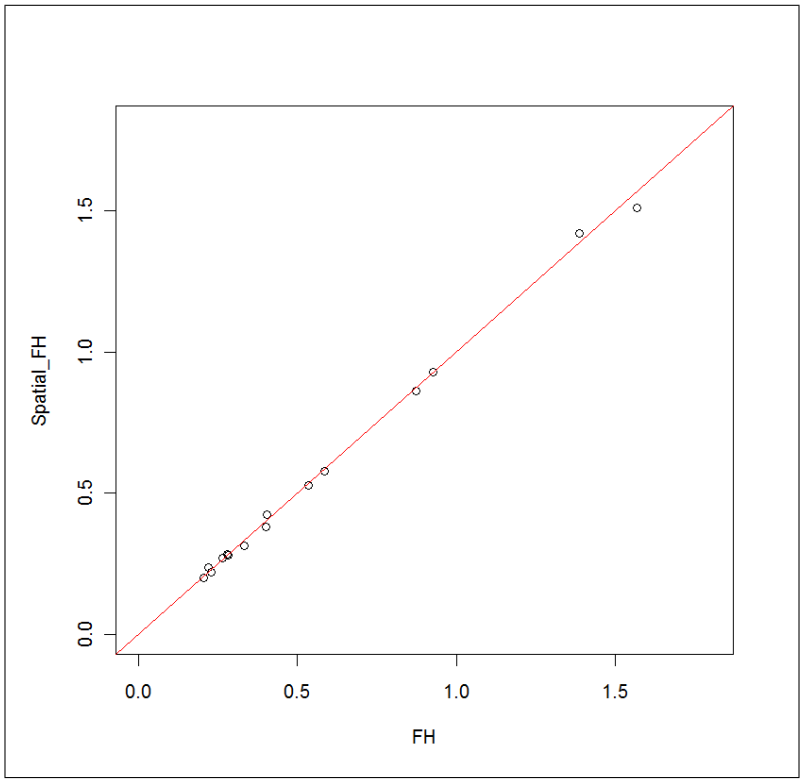

Figure 18: RRMSE for the 16 Federal States of the Spatial Against the Non-spatial FH Model.

Tables 11, 12, and 13 show the RRMSE and the deviance of the RRMSE of the non-spatial and spatial Battese-Harter-Fuller, Bernoulli, and Fay-Herriot models. We can observe that the RRMSE of the spatial BHF model is lower compared to the RRMSE of the non-spatial version in 10 of the 16 states. Like in Section 4.5 (Figure 10 on page 64), the results of the non-spatial model are better in the small states and Bavaria. This will be further discussed in Section 6.4.3. The results of the RRMSE of the spatial Bernoulli model mimic this observation. The spatial Fay-Herriot model improves the RRMSE slightly in 11 states.

		RBias of results		RRMSE of results		
Federal state	Data	Non-sp.	Spatial	Non-sp.	Spatial	Diff.
Berlin	15.2	−0.0667	−0.0780	0.5306	0.6334	−0.1028
Bremen	23.1	0.0773	0.0812	0.4172	0.4554	−0.0382
Brandenburg	8.4	0.0113	0.0050	0.4304	0.3620	0.0684
Baden-Württ.	21.5	−0.0362	−0.0259	0.2633	0.2329	0.0304
Bavaria	22.3	−0.0385	−0.0509	0.2699	0.2883	−0.0184
Hamburg	19.1	−0.1496	−0.1540	0.8494	0.8772	−0.0278
Hesse	17.2	−0.0434	−0.0342	0.3458	0.2712	0.0746
Mecklenburg-Vorp.	13.5	0.0781	0.0655	0.6558	0.5638	0.0920
Lower Saxony	19.6	−0.0219	−0.0348	0.2475	0.2233	0.0243
North Rhine Westph.	20.6	−0.0209	−0.0119	0.2407	0.2177	0.0230
Rhineland-Palatinate	26.1	−0.0783	−0.0679	0.3547	0.3113	0.0434
Saxony-Anhalt	8.5	0.0760	0.0700	0.9884	0.9031	0.0853
Saxony	6.7	0.1093	0.1037	1.7179	1.6319	0.0861
Schleswig-Holstein	13.0	0.0097	0.0073	0.2826	0.2720	0.0106
Saarland	29.0	0.0171	0.0505	0.2177	0.3073	−0.0896
Thuringia	9.4	0.0896	0.0889	1.0354	1.0125	0.0229

Table 11: Professional Training or University/College Degree, as well as RBias, RRMSE, and Difference of the RRMSE for the BHF Models; Data Source: German Mikrocensus (2002).

Figure 19 on page 97 shows the deviance of the RRMSE of the BHF and Bern models on the map. Yellow colored states indicate an improvement of the RRMSE through spatial modeling. For both unit-level models the only red colored states are small city states and the Saarland. The deviance of the RRMSE of the Fay-Herriot models is negative for 5 states. In contrast to the unit-level models those states are not mainly small or city states. Figure 20 on page 97 shows the deviance on the map. It may be observed that the states with negative deviance form a belt in the south of Germany.

	RBias of results		RRMSE of results		
Federal state	Non-sp.	Spatial	Non-sp.	Spatial	Diff.
Berlin	−0.0672	−0.0721	0.5044	0.5487	−0.0443
Bremen	0.0992	0.1090	0.5773	0.6457	−0.0684
Brandenburg	−0.0049	−0.0105	0.3475	0.3139	0.0336
Baden-Württ.	−0.0572	−0.0470	0.3600	0.3248	0.0351
Bavaria	−0.0591	−0.0729	0.3620	0.3897	−0.0277
Hamburg	−0.1219	−0.1225	0.6698	0.6748	−0.0050
Hesse	−0.0601	−0.0574	0.4255	0.3889	0.0366
Mecklenburg-Vorp.	0.0438	0.0296	0.5084	0.4047	0.1037
Lower Saxony	−0.0463	−0.0657	0.3543	0.3747	−0.0203
North Rhine Westph.	−0.0434	−0.0355	0.3408	0.3113	0.0294
Rhineland-Palatinate	−0.0949	−0.0879	0.4242	0.3913	0.0329
Saxony-Anhalt	0.0371	0.0294	0.6293	0.5071	0.1222
Saxony	0.0616	0.0585	1.0858	1.0173	0.0685
Schleswig-Holstein	−0.0175	−0.0190	0.3267	0.3199	0.0068
Saarland	0.0453	0.0862	0.3427	0.4586	−0.1159
Thuringia	0.0474	0.0481	0.6739	0.6347	0.0392

Table 12: Professional Training or University/College Degree, as well as RBias, RRMSE, and Difference of the RRMSE for the Bern Models; Data Source: German Mikrocensus (2002).

	RBias of results		RRMSE of results		
Federal state	Non-sp.	Spatial	Non-sp.	Spatial	Diff.
Berlin	0.0048	0.0007	0.2812	0.2791	0.0020
Bremen	−0.0435	−0.0442	0.2803	0.2839	−0.0036
Brandenburg	0.1219	0.1169	1.5677	1.5110	0.0567
Baden-Württ.	−0.0191	−0.0203	0.2290	0.2201	0.0089
Bavaria	−0.0381	−0.0407	0.2203	0.2363	−0.0159
Hamburg	−0.0834	−0.0811	0.5855	0.5787	0.0068
Hesse	0.0415	0.0435	0.4043	0.4243	−0.0199
Mecklenburg-Vorp.	0.0321	0.0288	0.3323	0.3125	0.0198
Lower Saxony	0.0058	0.0024	0.2034	0.2007	0.0027
North Rhine Westph.	−0.0274	−0.0273	0.2035	0.1995	0.0040
Rhineland-Palatinate	−0.0373	−0.0344	0.2647	0.2684	−0.0037
Saxony-Anhalt	0.0686	0.0675	0.8728	0.8614	0.0114
Saxony	0.0884	0.0907	1.3880	1.4197	−0.0318
Schleswig-Holstein	0.0604	0.0592	0.5354	0.5260	0.0094
Saarland	−0.1099	−0.1019	0.4021	0.3785	0.0236
Thuringia	0.0823	0.0823	0.9281	0.9279	0.0002

Table 13: Professional Training or University/College Degree, as well as Bias, RRMSE, and Difference of the RRMSE for the FH Models; Data Source: German Mikrocensus (2002).

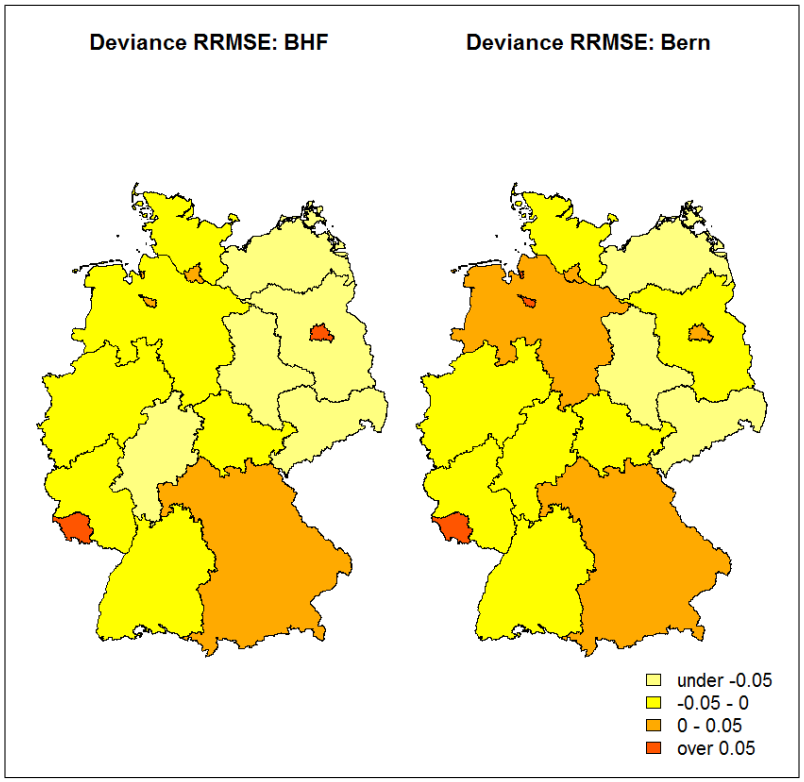

Figure 19: Deviance of the RRMSE for the Standard BHF and Bernoulli Models.

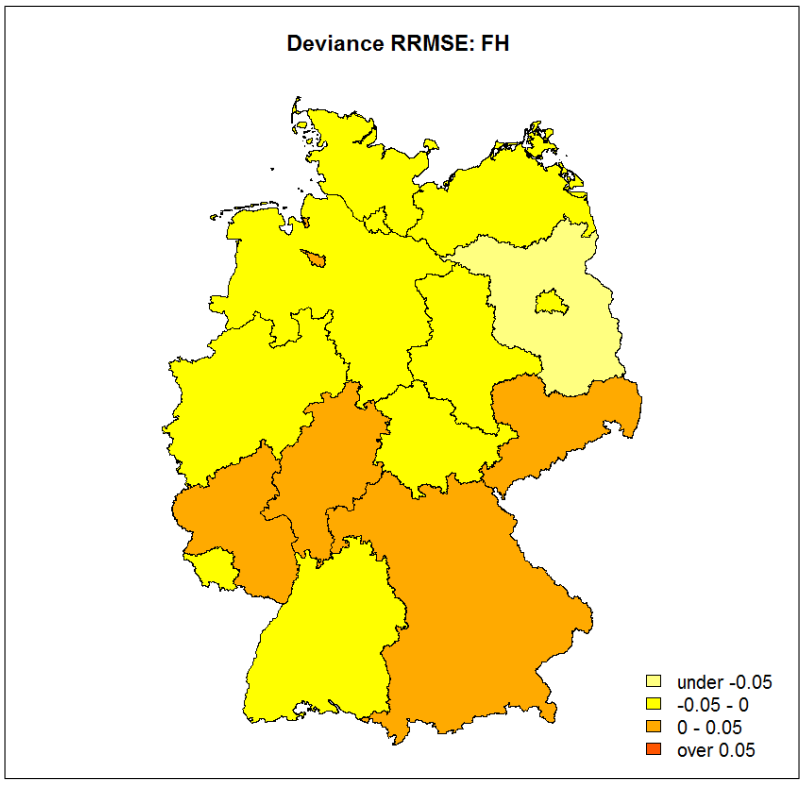

Figure 20: Deviance of the RRMSE for the Standard Fay-Herriot Model.

Figure 21 shows the boxplots of the relative Bias for the six models. The boxplots for the Fay-Herriot models are symmetric, whereas the boxplots for the Battese-Harter-Fuller and Bernoulli models are skewed. It can be observed that the skewness is reduced for the unit-level models, when using the spatial version. In addition for all model types the median of the RBias is closer to 0 under the spatial version. The Bernoulli models underestimate the true value, which has been observed before (cf. MÜNNICH et al., 2010).

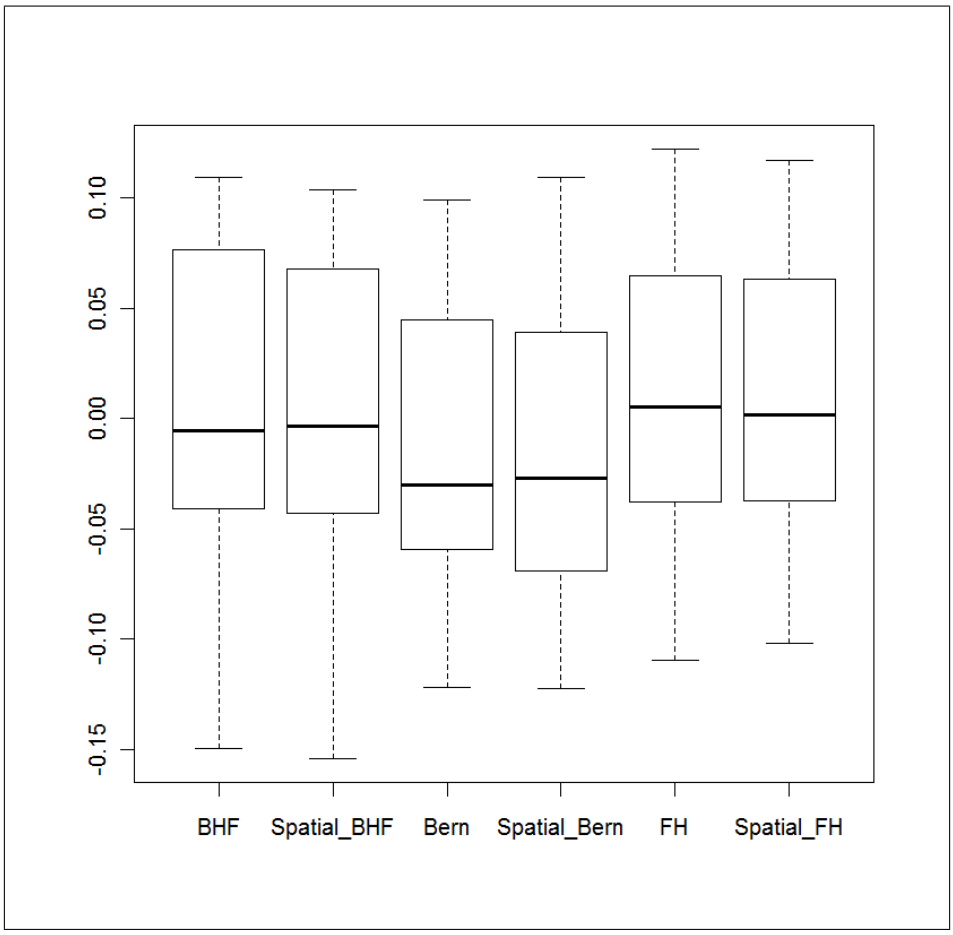

Figure 21: Boxplots of the Relative Bias for the 6 Standard Models.

Figures 22, 23, and 24 show the RBias on the map. It can be seen that all models tend to overestimate the federal states in East Germany and underestimate in West Germany. This might be due to borrowing strength (see Example 2.1 on page 5), since the gold standard values (see Figure 15) for the federal states in East Germany are lower than the values for West Germany. The technique of borrowing strength combines the values of the Eastern and Western states. This leads to increased estimated values for East Germany

compared to the sample and decreased values in West Germany. Stronger covariates than gender and age may reduce this effect. This can be observed in Figure 56 where a dummy variable for East Germany is included in the model.

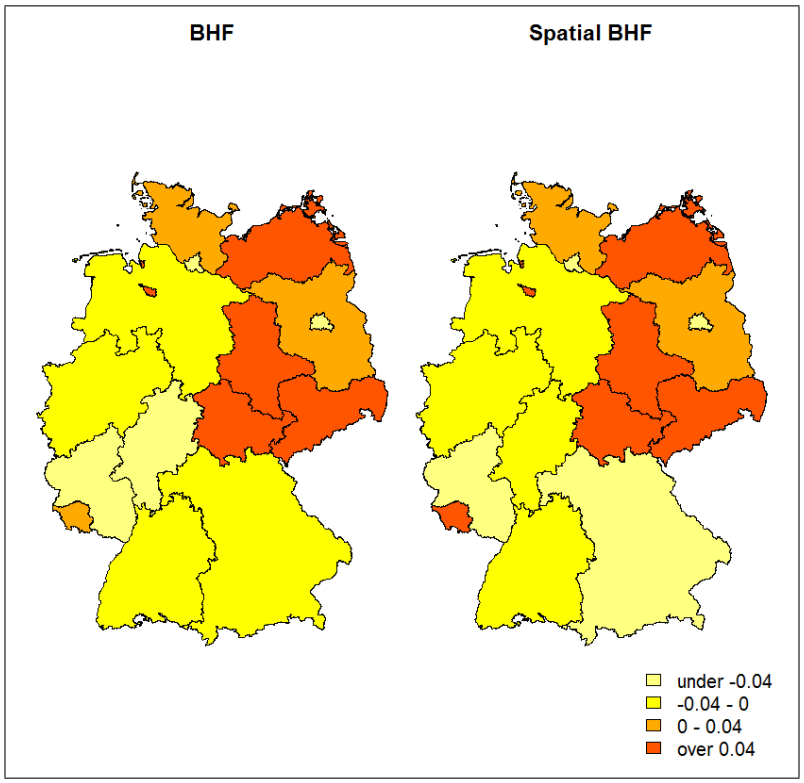

Figure 22: Relative Bias for the Spatial and Non-spatial BHF Models on the Map.

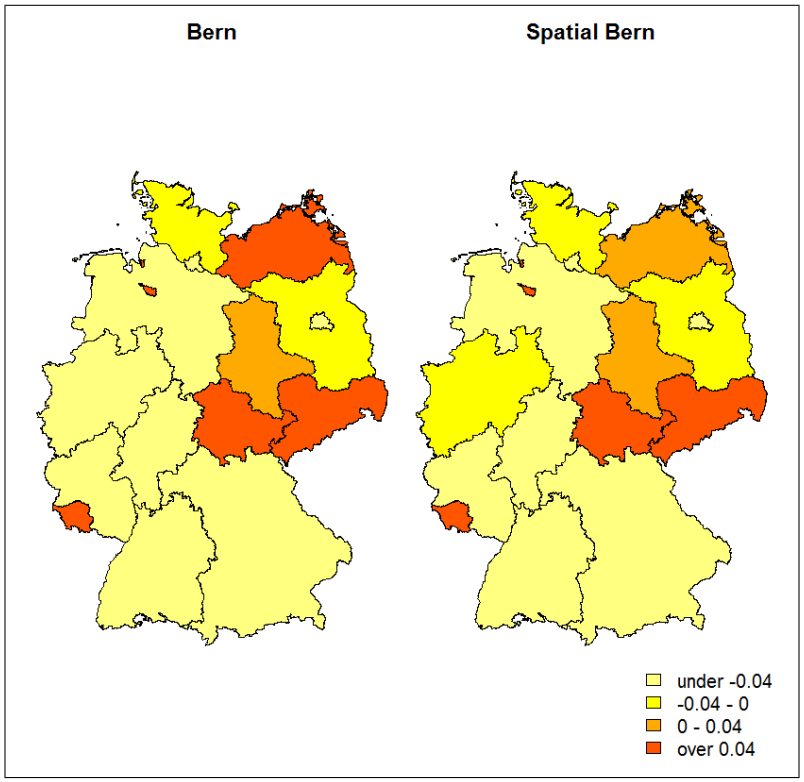

Figure 23: Relative Bias for the Spatial and Non-spatial Bern Models on the Map.

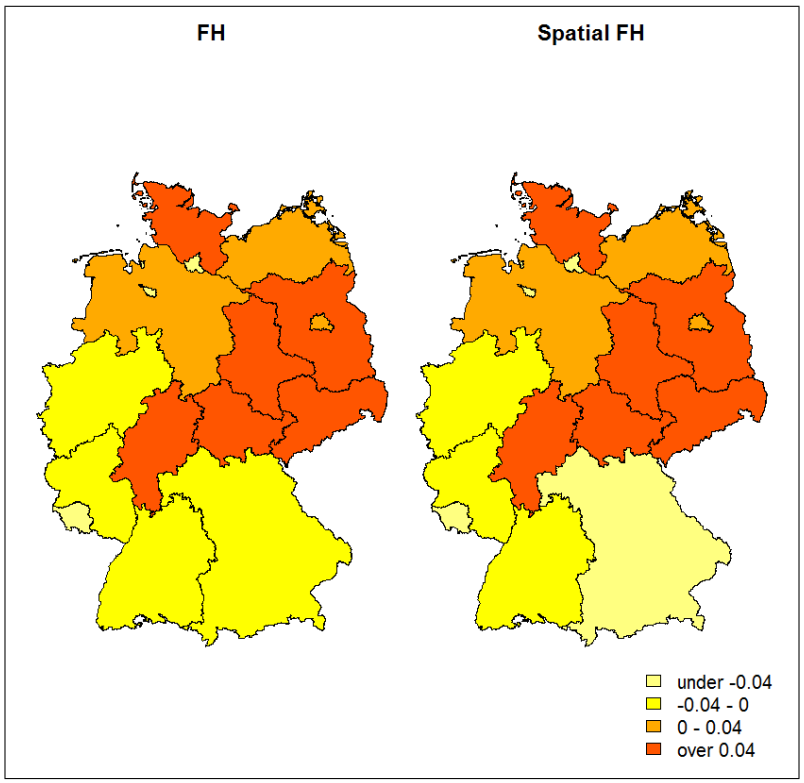

Figure 24: Relative Bias for the Spatial and Non-spatial FH Models on the Map.

Figure 25 shows the relative Dispersion of the six standard models. It may be observed that for the unit-level models the maximum and minimum of RDisp of the spatial models is lower, with a higher median. Whereas, the relative Dispersion of the Fay-Herriot models is quite similar.

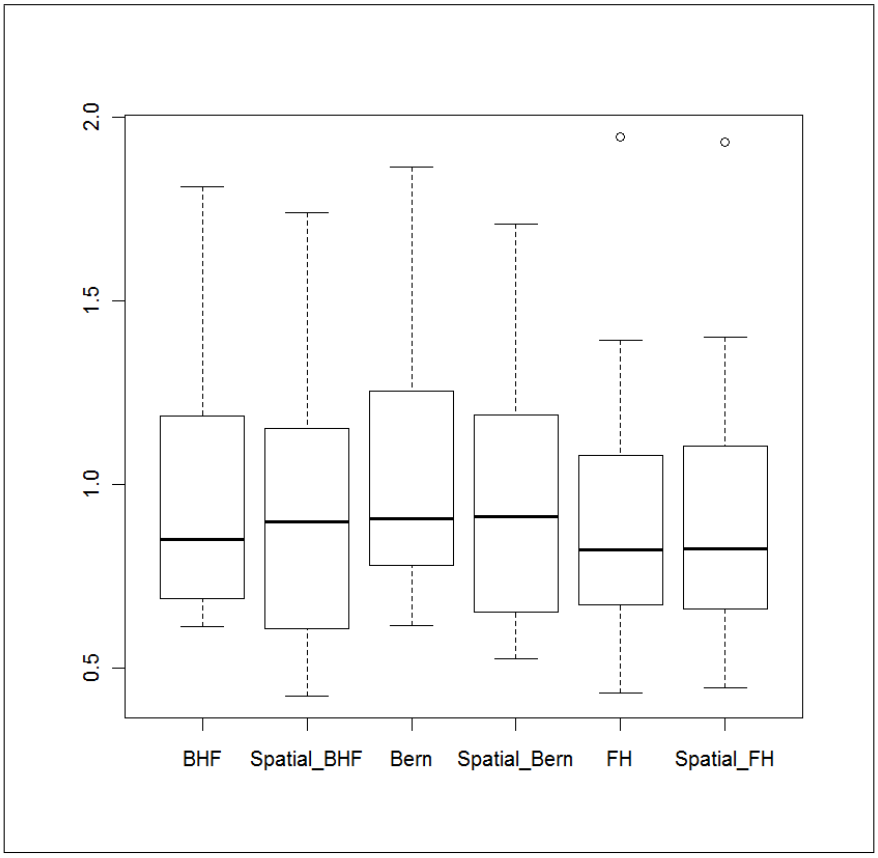

Figure 25: Boxplots of the Relative Dispersion for the 6 Standard Models.

6.4.2 The Influence of the Sample Size

In this section the influence of the sample size on the estimation results is described. To do this, three different simulations are done with 160 (10 sampled per area), 320 (20 per area), and 640 (40 per area) as total sample size.

Figures 26, 27, and 28 show the boxplots of the RRMSE of the spatial and non-spatial BHF, Bernoulli, and FH models for varying sample size. Each plot reveals that the RRMSE decreases as the sample size increases for all models. In addition, Table 14 shows that the

mean and the maximum of the RRMSE decrease for all models and is lower for the spatial models in all cases.

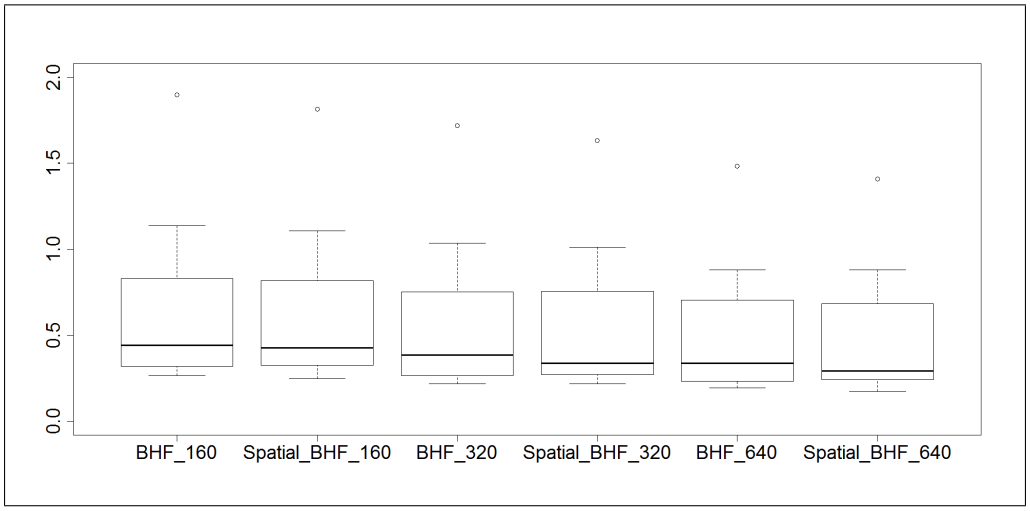

Figure 26: Boxplots of the RRMSE for the Spatial and Non-spatial BHF Models with Varying Sampling Sizes.

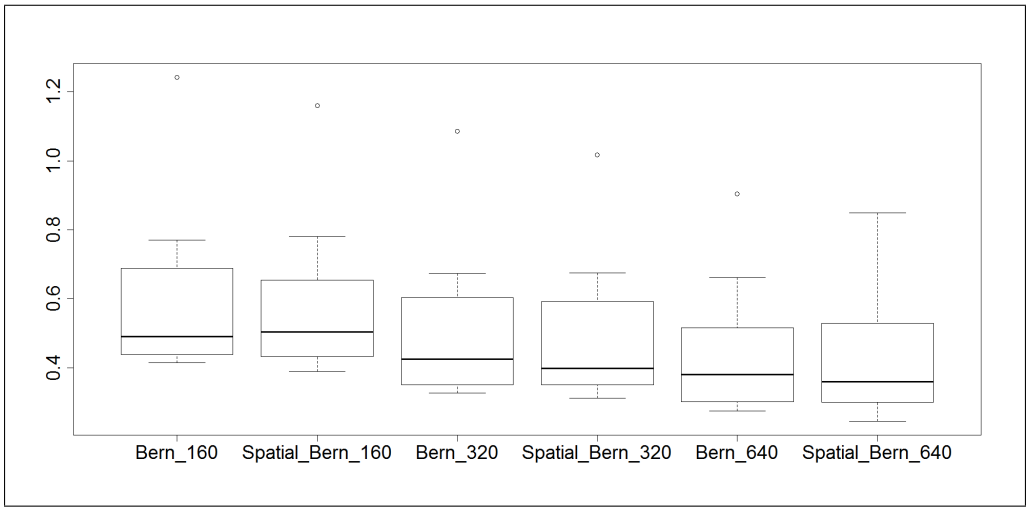

Figure 27: Boxplots of the RRMSE for the Spatial and Non-spatial Bernoulli Models with Varying Sampling Sizes.

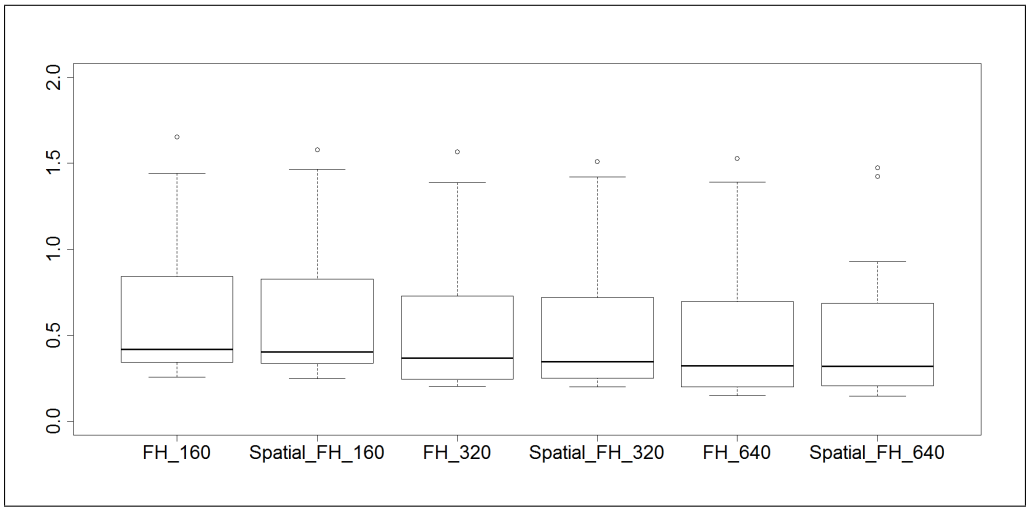

Figure 28: Boxplots of the RRMSE for the Spatial and Non-spatial FH Models with Varying Sampling Sizes.

Measure	Sample size	BHF	Spatial_BHF	Bern	Spatial_Bern	FH	Spatial_FH
Mean	160	0.63	0.61	0.59	0.57	0.62	0.61
	320	0.55	0.54	0.50	0.48	0.54	0.54
	640	0.49	0.48	0.43	0.42	0.51	0.50
Max	160	1.90	1.81	1.24	1.16	1.65	1.58
	320	1.72	1.63	1.09	1.02	1.57	1.51
	640	1.48	1.41	0.90	0.85	1.53	1.48

Table 14: Mean and Maximum of the RRMSE for the BHF, Bernoulli, and FH Models under Varying Sample Sizes.

Figures 29 and 30 show the relative Bias and the relative Dispersion of the BHF, Bern, and FH models for varying sample sizes. These plots reveal that the reduction of the RRMSE is due to a reduction in the variance not the relative Bias. In addition the relative Dispersion is reduced.

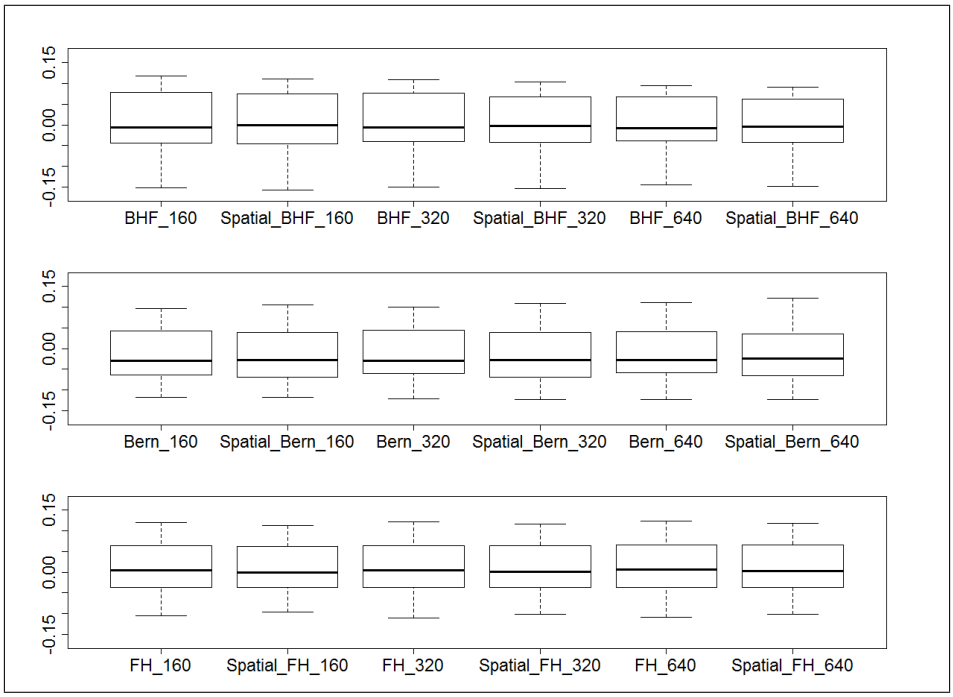

Figure 29: Boxplots of the Relative Bias for the Spatial and Non-spatial Models with Varying Sampling Sizes.

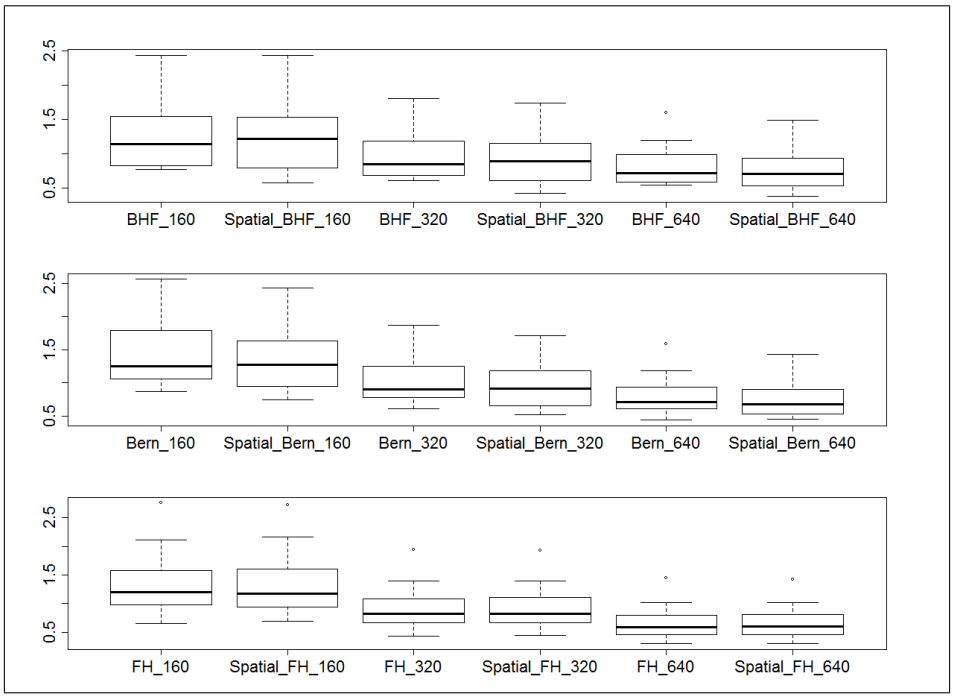

Figure 30: Boxplots of the Relative Dispersion for the Spatial and Non-spatial Models with Varying Sampling Sizes.

6.4.3 The Effect of Different Neighborhood Structures

Until now in every example and application the nearest neighbor structure has been used. Unfortunately, as may be seen in Figures 10 on page 64 and 19 on page 97, the quality of the estimates depends on the number of neighbors. Therefore, this assumption is changed such that every area has at least 2 neighbors. To do this the neighborhood structure of Example 2.25 is used. The basic principle is still a nearest neighbor structure, but the city states and Saarland are separated from the other states and chosen to be neighbors of each other.

Figure 31 shows the RRMSE for the standard and the new models (denoted by City). It can be observed that the new structure reduces the spread of the RRMSE for all models. Moreover, the maximum and the mean of the RRMSE is reduced as may be seen in Table 15.

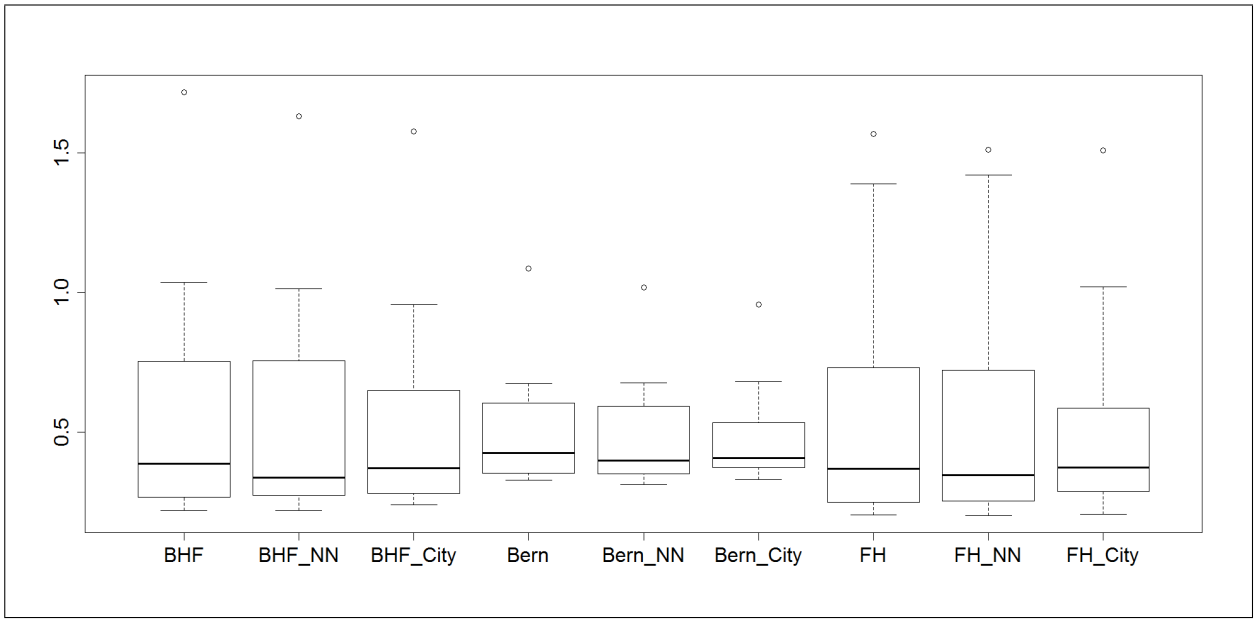

Figure 31: Boxplots of the RRMSE for the Standard and City Structure Models.

Measure	BHF_NN	BHF_City	Bern_NN	Bern_City	FH_NN	FH_City
Mean	0.54	0.52	0.48	0.47	0.54	0.50
Max	1.63	1.58	1.02	0.96	1.51	1.51

Table 15: Mean and Maximum of the RRMSE for the Spatial BHF, Bernoulli, and FH Models Under Nearest Neighbor (NN) and City Structure (City).

Table 16 and Figure 32 on page 109 show the deviance of the RRMSE for the spatial nearest neighbor and city structure Battese-Harter-Fuller and Bernoulli model. It may be observed that the RRMSE is lower for the nearest neighbor model in most states (plotted in yellow). This goes in hand with the higher median of the RRMSE of the new models in Figure 31. However, as noted above the mean as well as the maximum is smaller for the city structure models. This is due to the fact, that the deviance is negative, but close to zero. Whereas the states plotted in red (Figure 32) indicate strong gains of the new structure.

Federal state	BHF_NN	BHF_City	Deviance	Bern_NN	Bern_City	Deviance
Berlin	0.6334	0.4317	0.2017	0.5487	0.4464	0.1023
Bremen	0.4554	0.4886	-0.0332	0.6457	0.6800	-0.0343
Brandenburg	0.3620	0.3821	-0.0202	0.3139	0.3475	-0.0336
Baden-Württ.	0.2329	0.2612	-0.0284	0.3248	0.3489	-0.0240
Bavaria	0.2883	0.3098	-0.0215	0.3897	0.4052	-0.0155
Hamburg	0.8772	0.7626	0.1145	0.6748	0.6325	0.0422
Hesse	0.2712	0.3037	-0.0325	0.3889	0.4093	-0.0204
Mecklenburg-Vorp.	0.5638	0.5333	0.0305	0.4047	0.3876	0.0172
Lower Saxony	0.2233	0.2609	-0.0376	0.3747	0.3977	-0.0230
North Rhine Westph.	0.2177	0.2396	-0.0218	0.3113	0.3300	-0.0186
Rhineland-Palatinate	0.3113	0.3591	-0.0478	0.3913	0.4206	-0.0292
Saxony-Anhalt	0.9031	0.8554	0.0477	0.5071	0.4777	0.0294
Saxony	1.6319	1.5773	0.0546	1.0173	0.9560	0.0613
Schleswig-Holstein	0.2720	0.2956	-0.0237	0.3199	0.3584	-0.0385
Saarland	0.3073	0.2454	0.0620	0.4586	0.3878	0.0708
Thuringia	1.0125	0.9570	0.0555	0.6347	0.5875	0.0473

Table 16: RRMSE for the Spatial BHF and Bernoulli Models Under Nearest Neighbor (NN) and City Structure (City).

What is the effect on areas with few neighbors? Figure 32 shows that in 3 out of the 4 small states there has been a considerable improvement when using the city structure model. We can summarize that the RRMSE under the nearest neighbor structure is smaller than under the city structure for most states, but the gain is small. That is, the mean of the RRMSE under the city structure model is smaller.

Similar to the observations for the unit-level models, the mean of the RRMSE of the spatial and non-spatial Fay-Herriot model decreases when using the new structure (see Table 15), whereas the maximum is equal. Figure 33 on page 109 shows the deviance of

the RRMSE for the two spatial Fay-Herriot models. In this case, unlike for the unit-level models, positive effects for the small states cannot be observed. This goes along with the observations of Figure 20 on page 97, where a different effect of spatial modeling on the small states under the Fay-Herriot model could be observed.

Federal state	FH_NN	FH_City	Deviance
Berlin	0.2791	0.5631	-0.2840
Bremen	0.2839	0.2929	-0.0089
Brandenburg	1.5110	1.5101	0.0009
Baden-Württ.	0.2201	0.2652	-0.0452
Bavaria	0.2363	0.3026	-0.0664
Hamburg	0.5787	0.4688	0.1100
Hesse	0.4243	0.4414	-0.0171
Mecklenburg-Vorp.	0.3125	0.2880	0.0245
Lower Saxony	0.2007	0.2045	-0.0038
North Rhine Westph.	0.1995	0.3012	-0.1017
Rhineland-Palatinate	0.2684	0.2609	0.0075
Saxony-Anhalt	0.8614	0.6053	0.2561
Saxony	1.4197	1.0192	0.4005
Schleswig-Holstein	0.5260	0.4430	0.0831
Saarland	0.3785	0.2844	0.0941
Thuringia	0.9279	0.7195	0.2084

Table 17: RRMSE for the Spatial FH Models Under Nearest Neighbor (NN) and City Structure (City) .

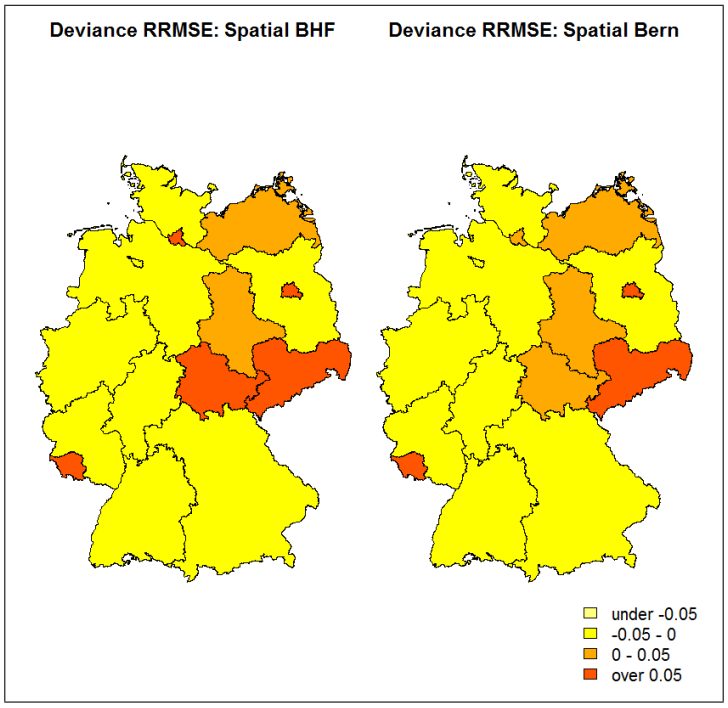

Figure 32: Deviance of the RRMSE for the Nearest Neighbor and City Structure BHF and Bern Models.

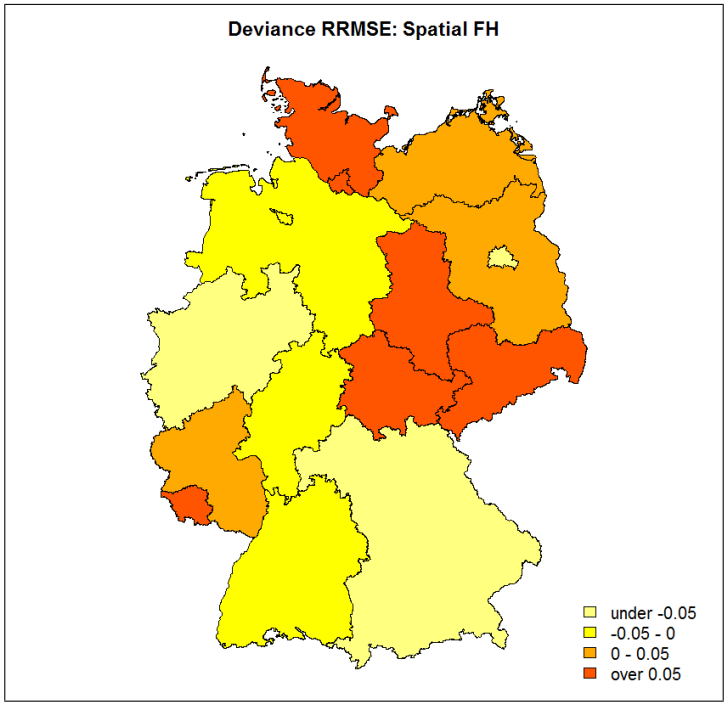

Figure 33: Deviance of the RRMSE for the Nearest Neighbor and City Structure FH Models.

Figure 34 shows the relative Bias for the model under the city structure neighborhood. We can observe that the spread of the relative Bias for the BHF and FH city structure models is smaller than for the nearest neighbor models. However, the relative Dispersion (Figure 35) increases under the new structure.

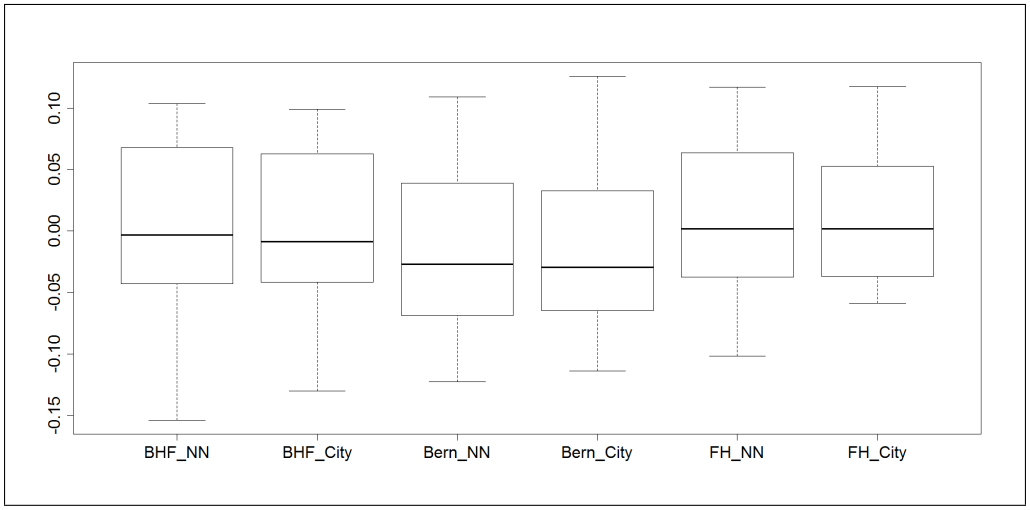

Figure 34: Relative Bias of the BHF, Bern, and FH Models for the City Structure.

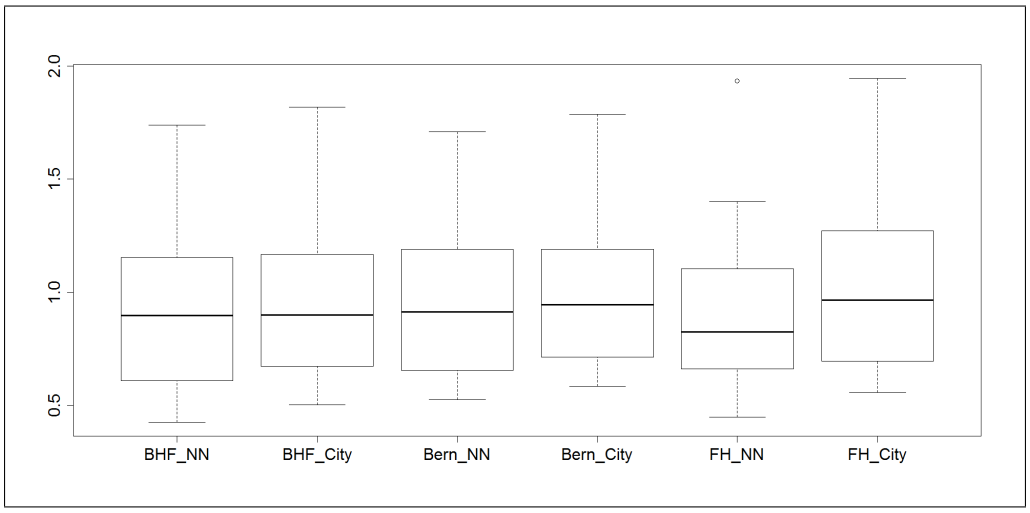

Figure 35: Relative Dispersion of the BHF, Bern, and FH Models for the City Structure.

6.4.4 The Influence of the Prior Distribution

The prior distributions of the variance components of the spatial and non-spatial Battese-Harter-Fuller model contain the unknown constants a and b. Similarly, the Bernoulli models

contain the parameter a. In this section the estimates are calculated for different values of a and b. Figures 36 and 37 show the boxplots of the RRMSE of the spatial and non-spatial BHF and Bernoulli models. These plots are drawn under different values of $a = b$ for the BHF models and a for the Bernoulli models. The values considered are -0.6, -0.7, -0.8, and -0.9.

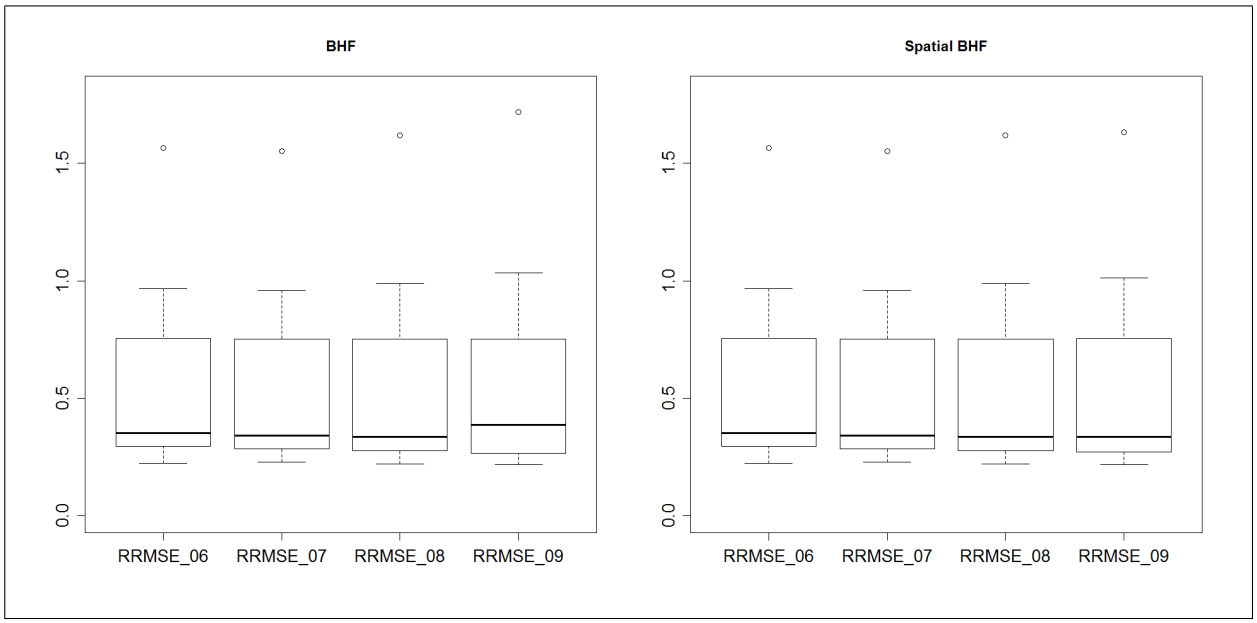

Figure 36: Boxplots of the RRMSE of the BHF Models for Different Prior Distributions.

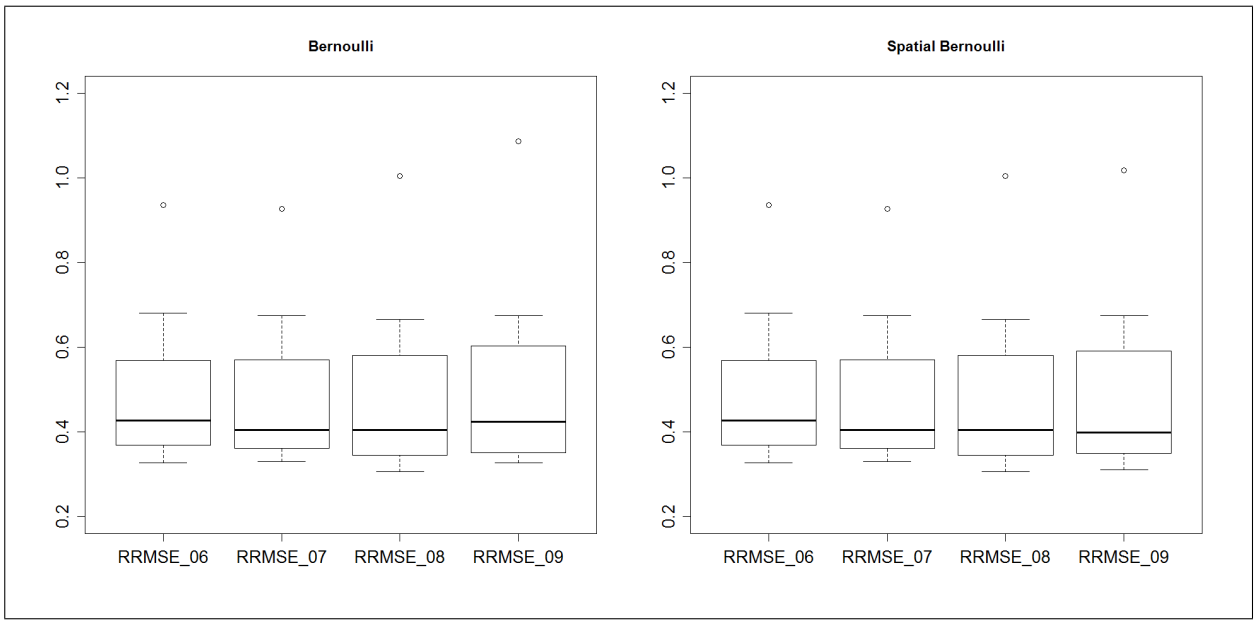

Figure 37: Boxplots of the RRMSE of the Bern Models for Different Prior Distributions.

The maximum of the RRMSE seems to decrease first and then increase again. Other than that, we can observe that the estimation results are not very sensitive under different choices of the prior distribution. The same observation can be made in Figures 38 and 39 showing the relative Bias and the relative Dispersion.

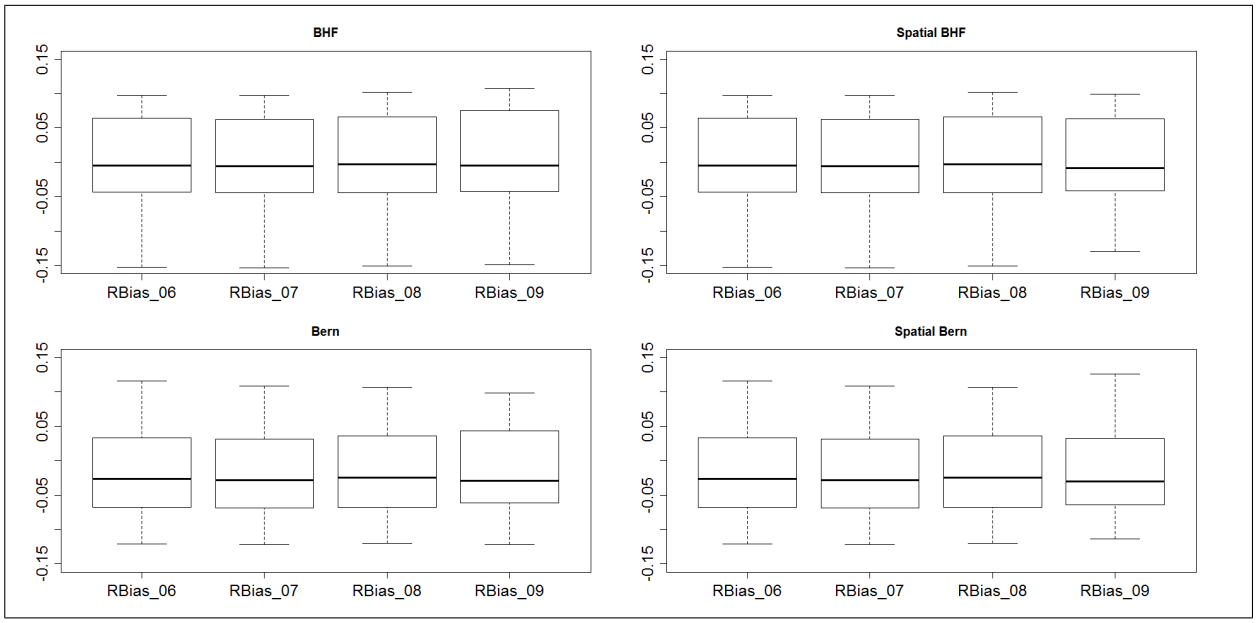

Figure 38: Boxplots of the Relative Bias for Different Prior Distributions.

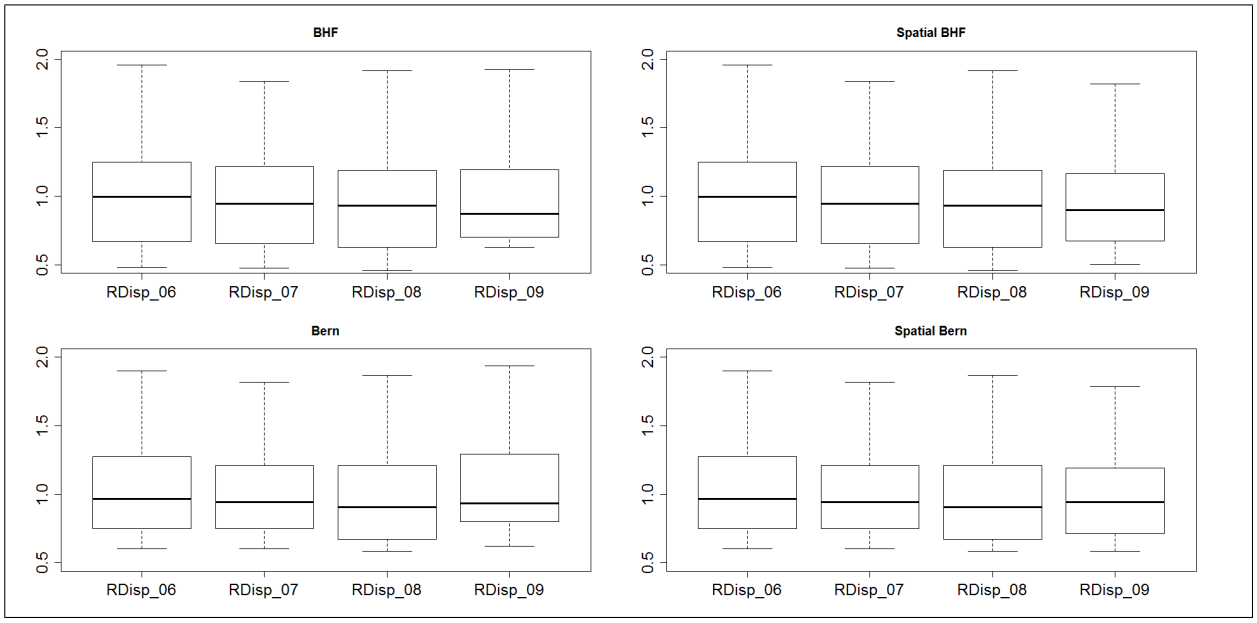

Figure 39: Boxplots of the Relative Dispersion for Different Prior Distributions.

Up until this point, a has been set equal to b. In the following calculation $a = -0.9$

will be held fixed, while b varies between -0.6 and -0.9. Since b is not present in the Bernoulli models, just the BHF models will be considered. Figures 40 and 41 show the corresponding values of the RRMSE, RBias, and RDisp. It can be observed that the results are not sensitive with respect to changes in b.

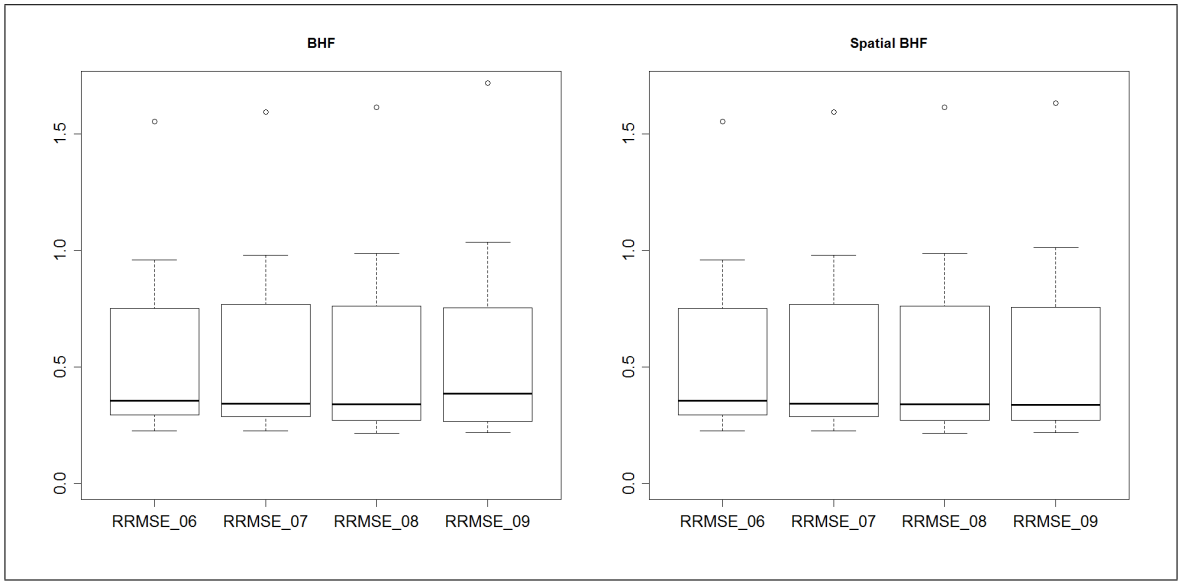

Figure 40: RRMSE for the BHF and Spatial BHF Models with $a = -0.9$ and Varying b.

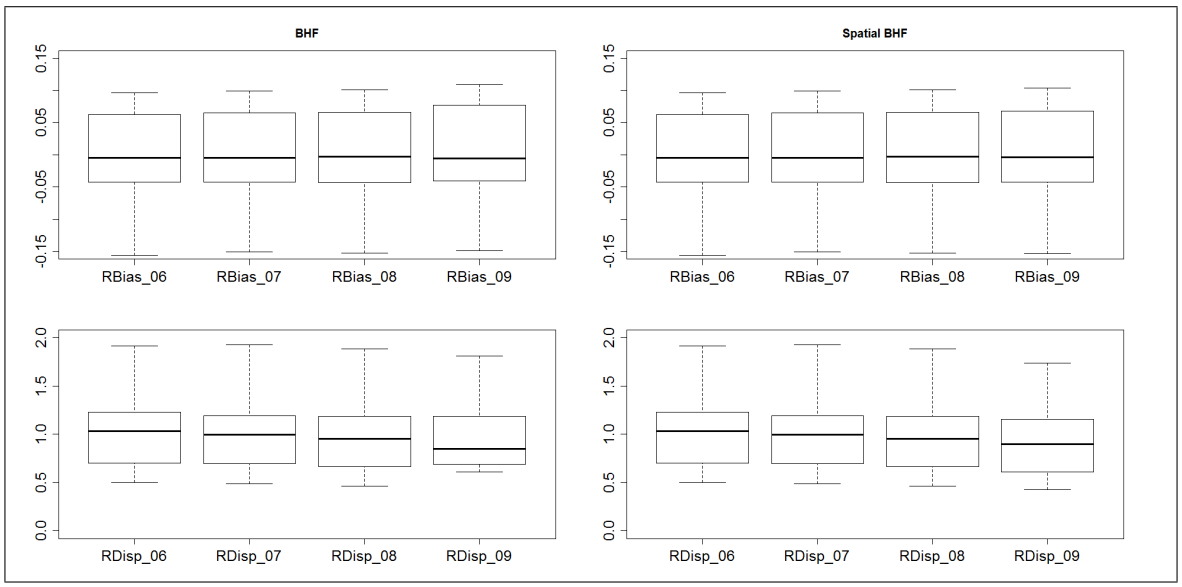

Figure 41: RBias and RDisp for the BHF and Spatial BHF Models with $a = -0.9$ and Varying b.

Figures 42 and 43 show the values of the RRMSE, RBias, and RDisp for the model with fixed $b = -0.9$ and varying a. Similar to the case of $a = -0.9$ the results are insensitive to the choice of the parameters.

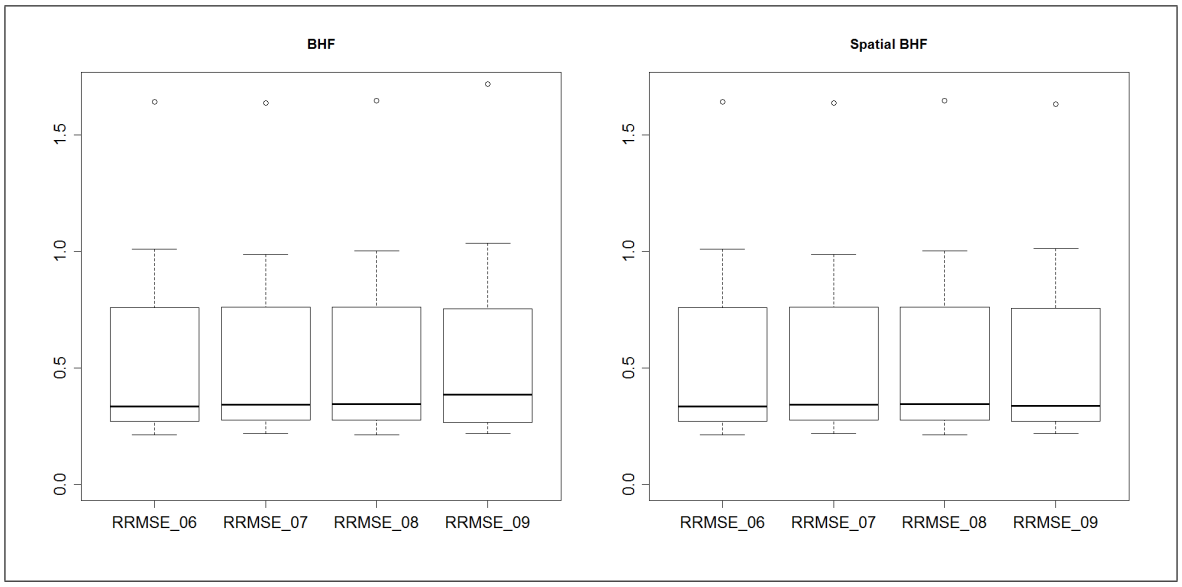

Figure 42: RRMSE for the BHF and Spatial BHF Models with $b = -0.9$ and Varying a.

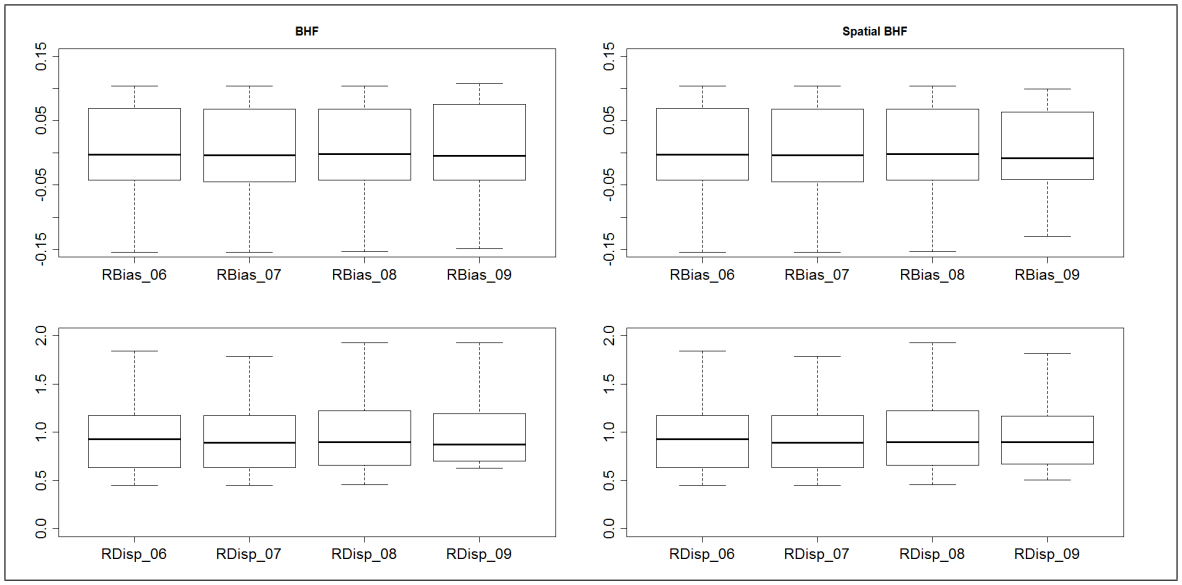

Figure 43: RBias and RDisp for the BHF and Spatial BHF Models with $b = -0.9$ and Varying a.

6.4.5 The Impact of the Spatial Correlation Parameter

When using the CAR structure (2.20) there is a need to specify the spatial correlation parameter p. In this section the effect of different parameters is tested. Figures 44 and 45 show the RRMSE for the spatial Battese-Harter-Fuller and Bernoulli model with spatial correlation coefficients 0.2, 0.5, 0.6, 0.7, 0.8, 0.9, and 0.95. In addition the non-spatial models are included. The Fay-Herriot model is not included in this setting, since the correlation parameter is set to 1.

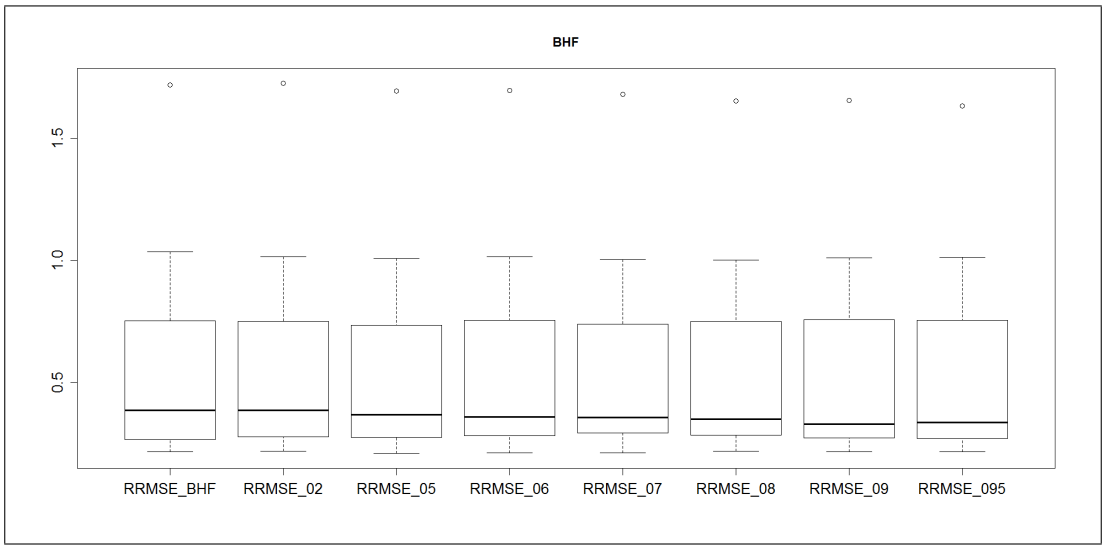

Figure 44: RRMSE for BHF Models with Varying Spatial Correlation Parameter.

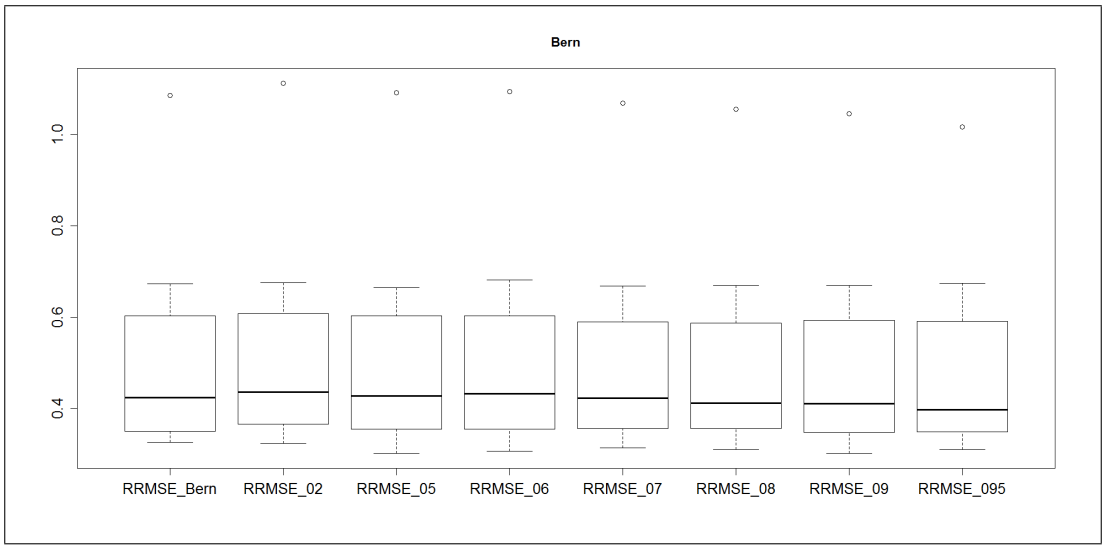

Figure 45: RRMSE for Bern Models with Varying Spatial Correlation Parameter.

Although the boxplots are quite similar, we may observe that the maximum values of the RRMSE decrease when the parameter increases. This effect can be better seen in Figure 46, where the maximum of the RRMSE of the spatial Battese-Harter-Fuller and Bernoulli model is shown. The red line indicates a simple least-squares regression line. The non-spatial model is included at 0. The plots reveal, that values of the spatial correlation parameter above 0.7 seem to improve the results compared to the non-spatial version.

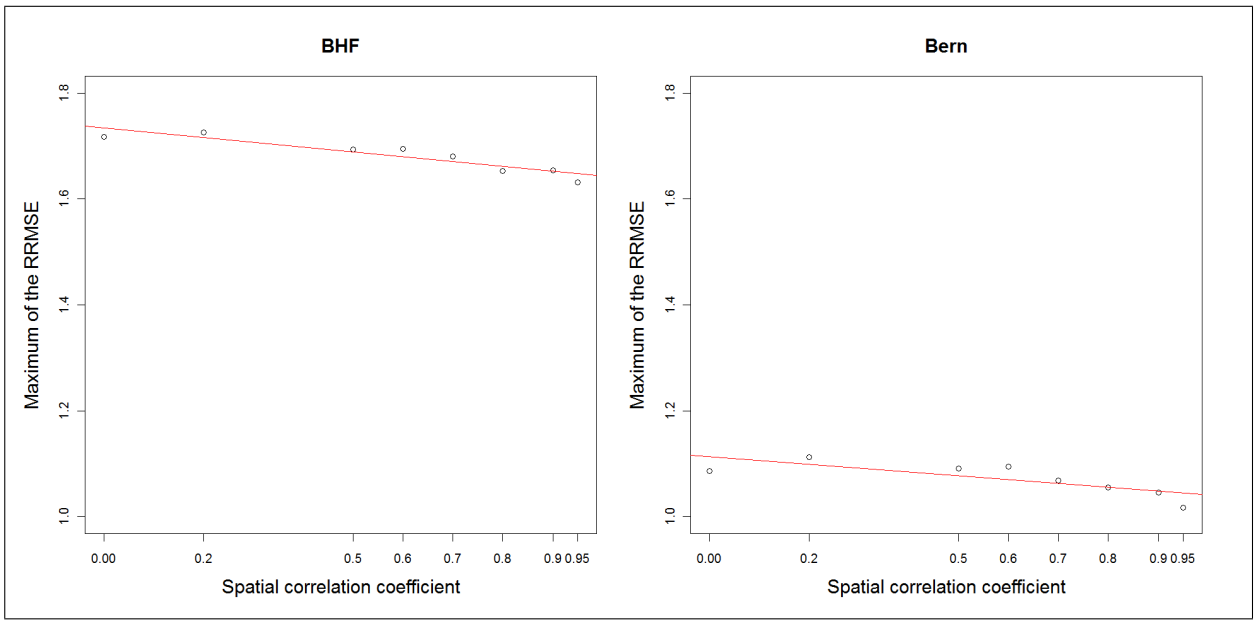

Figure 46: Maximum of the RRMSE for BHF and Bernoulli Models with Varying Spatial Correlation Parameter.

Figures 47 and 48 show the relative Bias and the relative Dispersion of the models under varying spatial correlation parameter. Like the RRMSE the boxplots are quite similar, where the maximum of the RDisp slightly decreases with increasing spatial correlation parameter.

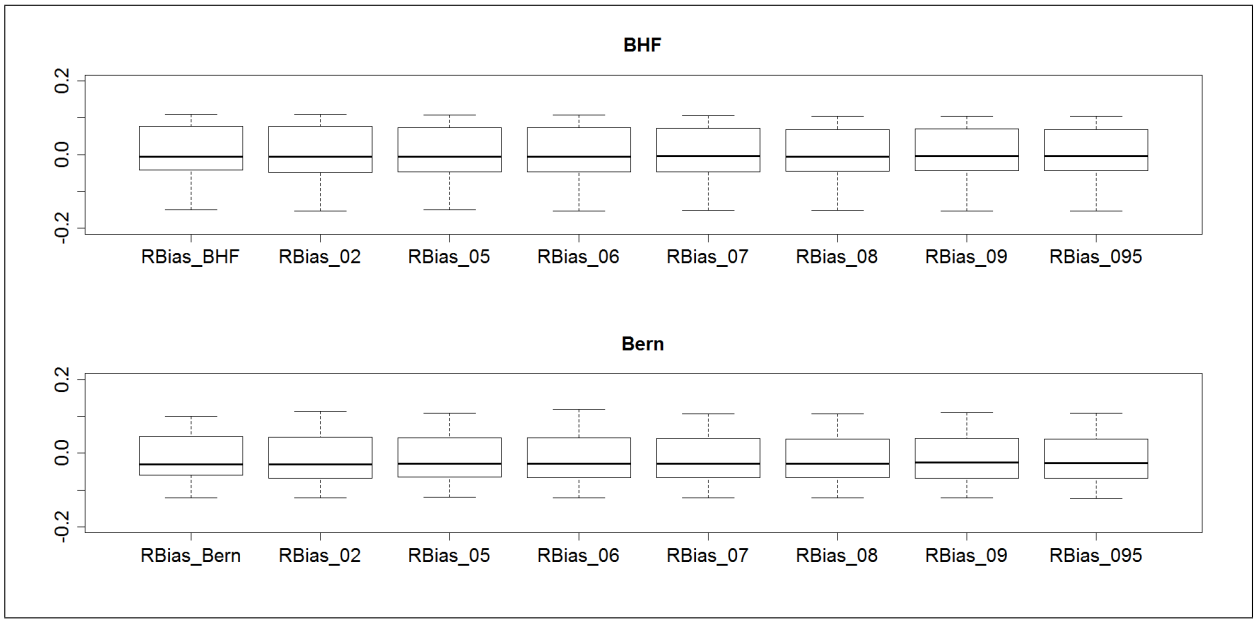

Figure 47: Relative Bias of the BHF and Bern Models with Varying Spatial Correlation Parameter.

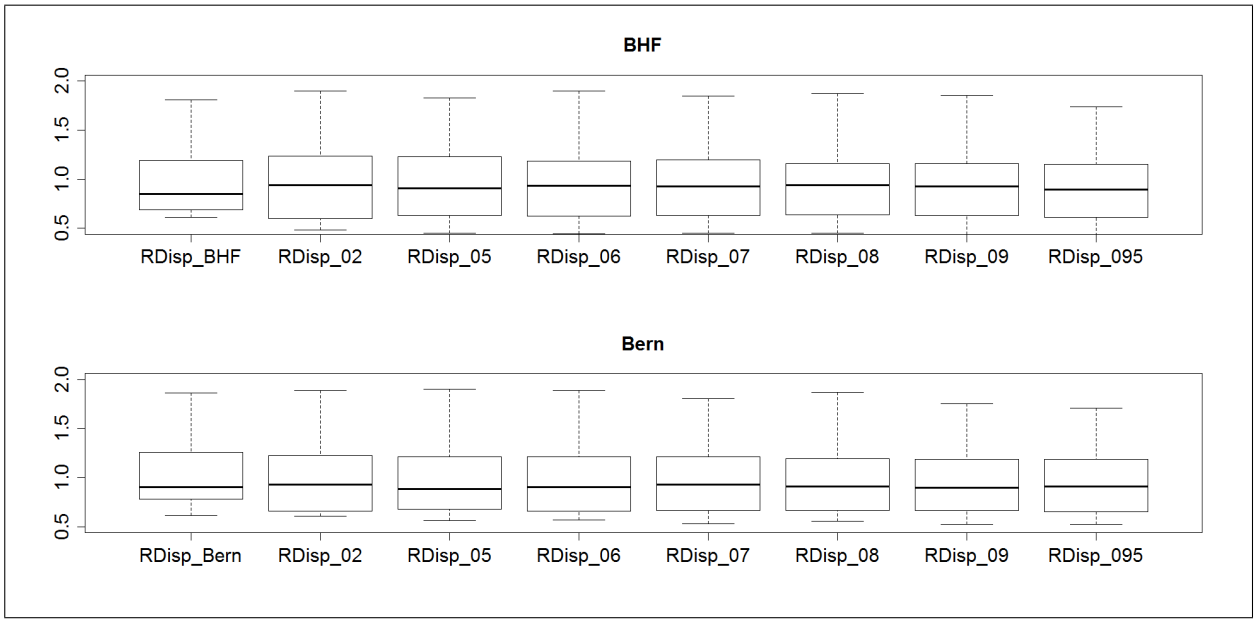

Figure 48: RDisp for BHF and Bern Models with Varying Spatial Correlation Parameter.

6.4.6 Dummy Variable for East Germany

Figure 15 on page 86 reveals differences in the structure of the variable education between Eastern and Western Germany. In the previous setups this has been modeled via spatially correlated random effects. In this setting a dummy variable for East Germany is included. Note that in other applications with a more complex spatial structure this might not be an option.

Figures 49 and 50 show the boxplot of the RRMSE for the Battese-Harter-Fuller, Bernoulli, and Fay-Herriot model with dummy variable, denoted by East, and the corresponding spatial version with and without dummy variable. It may be observed that the inclusion of the dummy variable in the non-spatial model considerably improves the RRMSE compared to the spatial version. However, if the dummy variable is included in the spatial model as well, the RRMSE boxplots are quite similar. This goes along with the observations of Chapter 5, where strong covariates led to a decrease in the gains of spatial modeling.

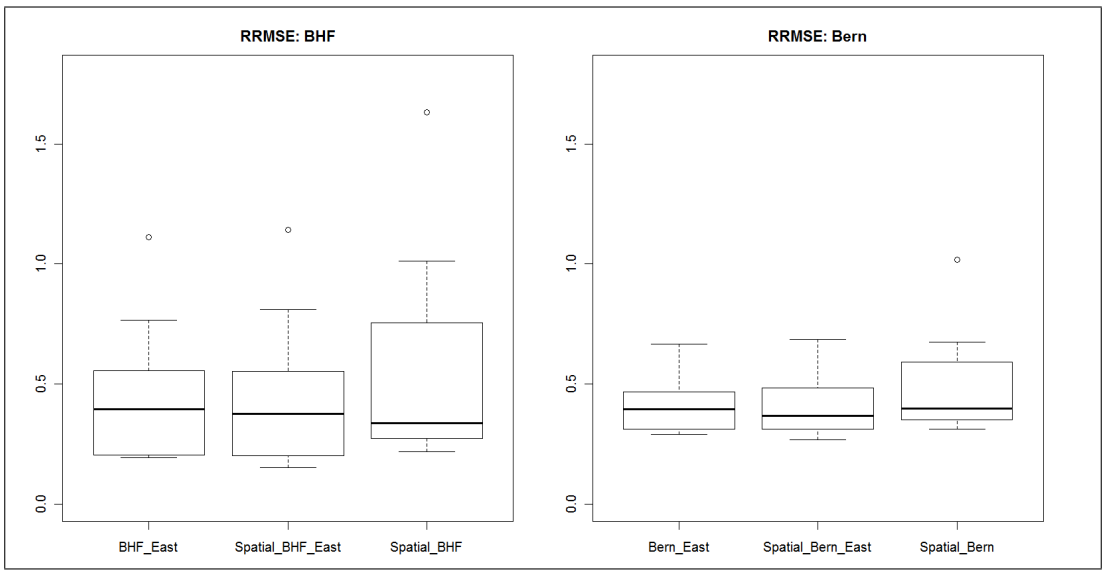

Figure 49: Boxplot of the RRMSE for the BHF and Bern Model with Dummy Variable East and Spatial Versions.

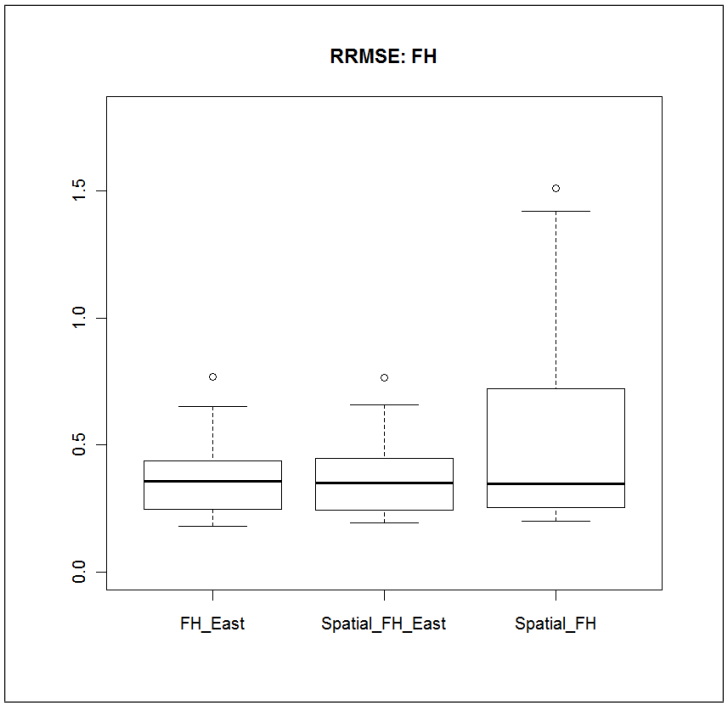

Figure 50: Boxplot of the RRMSE for the FH Model with Dummy Variable East and Spatial Versions.

Figures 51, 52, and 53 show the deviance of the RRMSE for the models on the map. The left plots underline the above mentioned effect that the RRMSE of the dummy variable non-spatial models is mostly lower compared to the standard spatial model without dummy variable. If the dummy variable is included in the spatial models as well (right plots), the deviance of the RRMSE for the unit-level models is negative for most areas, similar to the observations of Section 6.4.1. The deviance of the RRMSE is mostly small for the dummy variable non-spatial and spatial Fay-Herriot model.

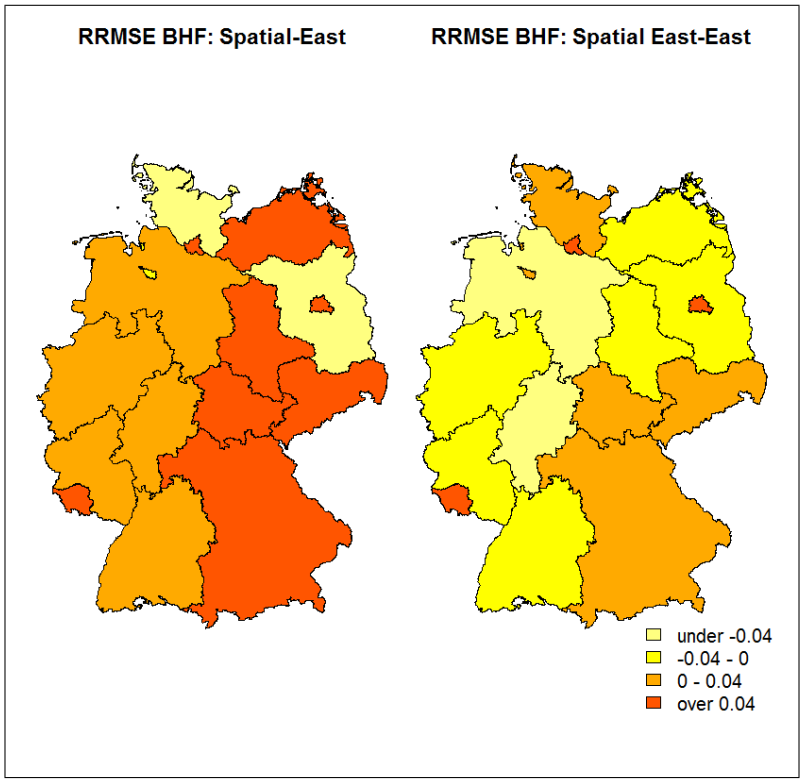

Figure 51: Deviance of the RRMSE for the BHF Model with Dummy Variable East and Spatial Versions.

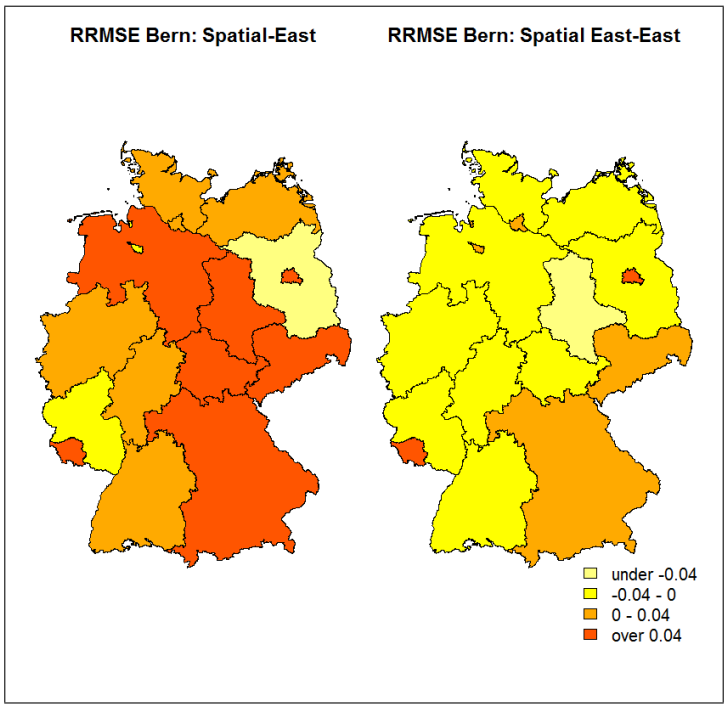

Figure 52: Deviance of the RRMSE for the Bern Model with Dummy Variable East and Spatial Versions.

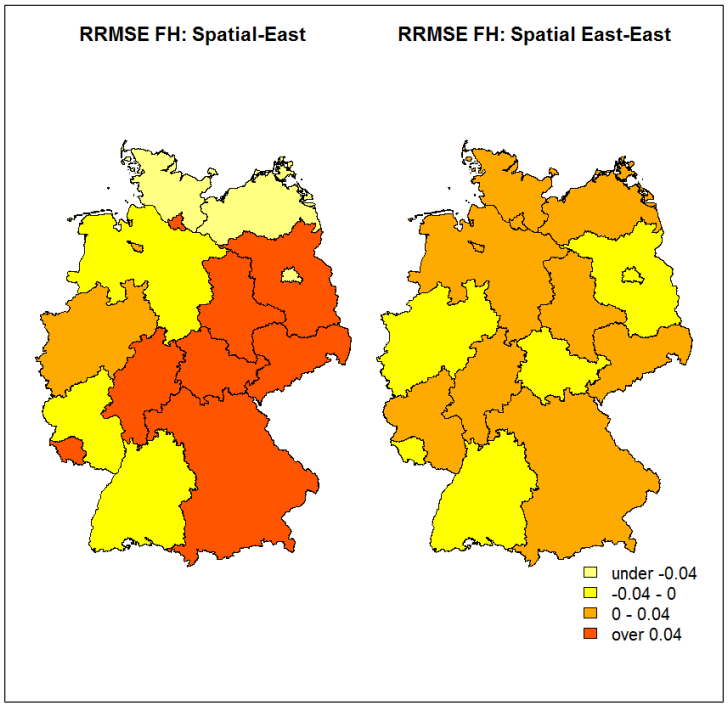

Figure 53: Deviance of the RRMSE for the FH Model with Dummy Variable East and Spatial Versions.

Figures 54 and 55 show the RBias and the RDisp for the models with and without the dummy variable East. The dummy variable reduces the spread of the RBias for all of the models. This cannot be observed for the RDisp. Especially for the Fay-Herriot model, the maximum, median, and the minimum of the RDisp of the spatial non-dummy variable model are lower.

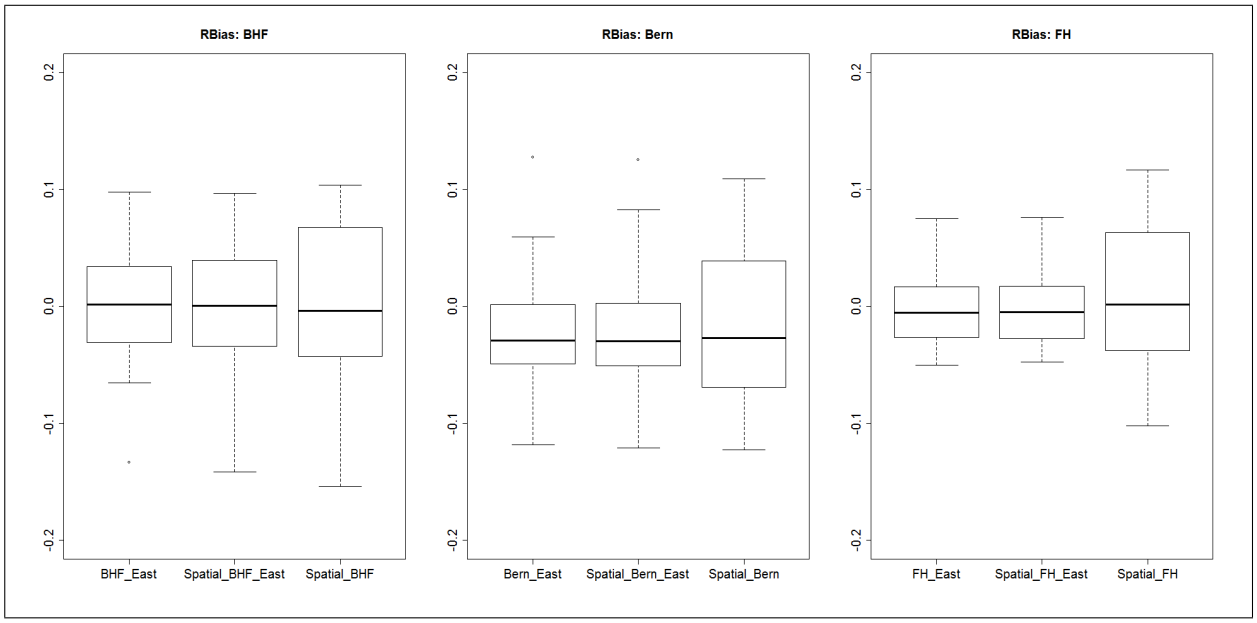

Figure 54: RBias for the Models with Dummy Variable East and the Spatial Versions.

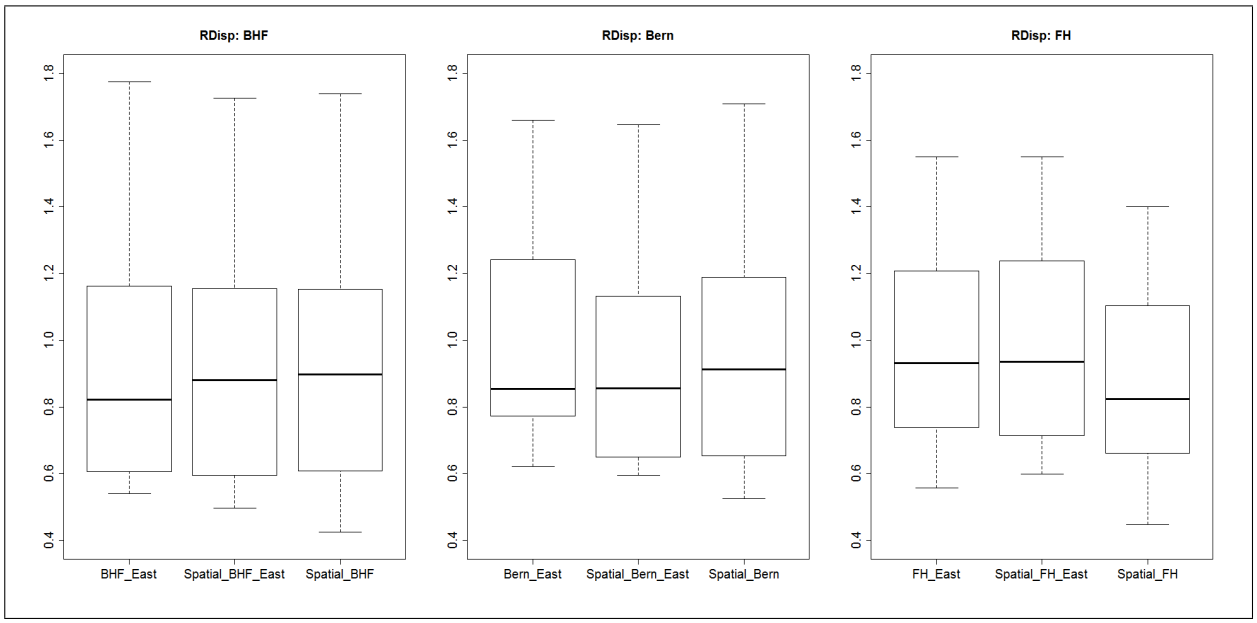

Figure 55: RDisp for the Models with Dummy Variable East and the Spatial Versions.

Figure 56 shows the RBias of the spatial and non-spatial BHF model with dummy variable East on the map[3]. We may observe that compared to the RBias of the standard BHF models (Figure 22 on page 99) the inclusion of the dummy variable removed overestimation in East Germany and underestimation in West Germany.

[3]The RBias for the other model types shows similar results and is omitted.

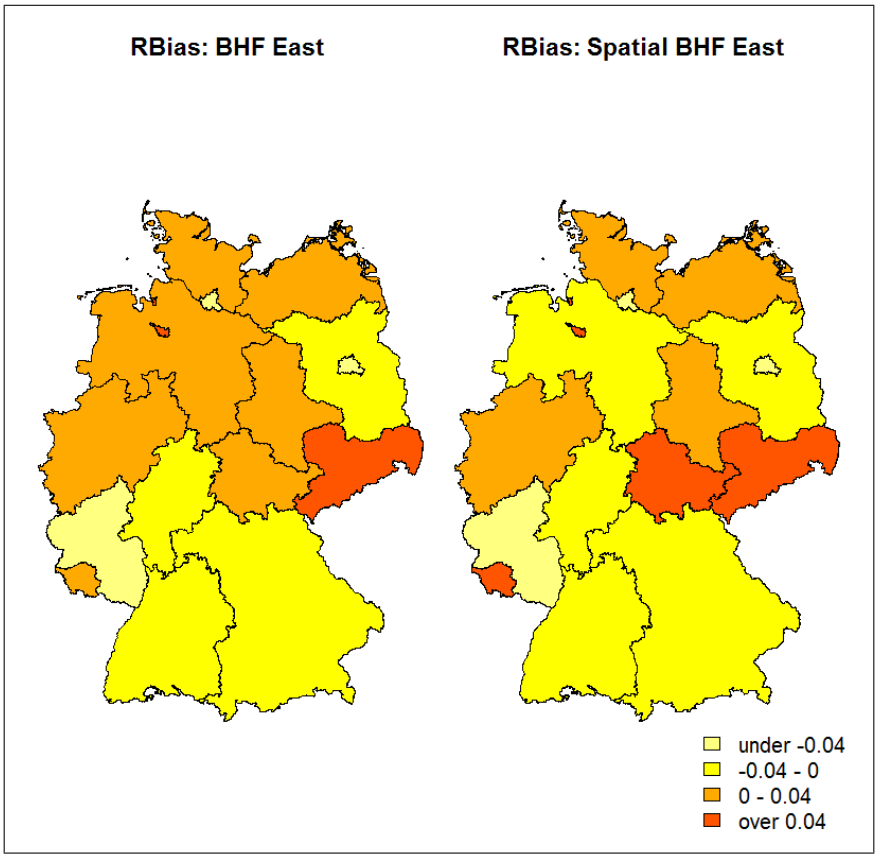

Figure 56: RBias for the Spatial and Non-spatial BHF Model with Dummy Variable on the map.

6.5 Conclusion

In this chapter three models and their spatial extensions have been tested under various settings. The simulation study revealed that especially for the unit-level models spatial modeling improved the estimation results. The maximum of the RRMSE of the spatial model was lower than the maximum for the non-spatial version in all settings. This is important for applications where quality constraints for areas are specified. Furthermore, the estimation results were insensitive with respect to the choice of the prior distribution and the spatial correlation parameter, which is essential for applications with political impact such as the German Census 2011. The estimation results varied under different neighborhood structures. Further improvement of the results of spatial modeling may be achieved by choosing specially tailored neighborhood structures. Also, it could be observed

that similar to the findings of Chapter 5 the inclusion of strong covariates (dummy variable for East Germany), reduces the gains of spatial modeling.

In a next step the simulation study may be enhanced to discuss further settings and variables concerning the German Census 2011. In the German Census 2011, a modern sampling design will be employed (cf. Gabler et al., 2010), and it is of interest to test the performance of the Bayesian spatial models with respect to this design. One challenge hereby is that in the German Census 2011 more areas than the 16 federal states are of interest and the data is high dimensional. The use of WinBUGS may be limited under this framework and specially tailored MCMC methods need to be derived.

7 Summary and Outlook

The motivation of this work has been the German Census 2011, where a new methodology will be employed - a register based census. In addition to the use of administrative registers an additional sample is drawn in order to allow for correcting register errors. This leads to statistical challenges, such as choosing estimation techniques for a large set of small areas, where classical estimators may not be reliable (see Website: Census, 2010c). Possible alternatives include model based small area methods, which have been derived in the last decade. Standard small area models, like the Fay-Herriot or Battese-Harter-Fuller, were reviewed in Section 2.1.2. These methods assume independent random effects. In applications such as the German Census 2011, independence may not be fulfilled and the inclusion of spatial interactions into the model may lead to improved estimates.

In Section 2.4 the independence assumption of the random effects has been dropped to account for spatial correlations between the areas, via the conditional autoregressive (CAR) approach using a Bayesian analysis. Depending on the chosen prior distribution, the posterior distribution is not guaranteed to be proper (cf. Section 2.2.2). Example 2.16 in Section 2.3 showed that this can cause misleading results if MCMC methods are utilized. Therefore, statisticians must ensure the propriety of the posterior distribution before using a model.

In this work, the following three model classes were considered:

1. The spatial general linear mixed model with power prior distributions on the variance components. In Chapter 3, the propriety of the posterior has been proved under certain assumptions for this model class. Further, the Gibbs conditionals, necessary for an implementation of the Gibbs sampler have been derived. An application to unemployment data showed that the spatial model may provide gains over the non-spatial version.

2. The spatial Bernoulli model class with power prior distribution on the variance component. In Chapter 4, two proofs for the propriety of the posterior distribution

under certain assumptions were given. The first proof does not rely on the result for the non-spatial model by Natarajan and McCulloch (Theorem 4.1). The second proof makes extensive use of the result and is much shorter. Similar to the results of Chapter 3, an application to the unemployment data showed possible gains of the proposed model over the non-spatial version.

3. The spatial hierarchical Fay-Herriot model with inverse gamma distribution on the variance component. In Chapter 5, the propriety of the posterior distribution has been proved under certain assumptions. Moreover, the situation of one unsampled area was considered, and the necessary theory for predicting the mean of the unsampled area has been derived. An application to SAIPE data of the U.S. Census Bureau showed that spatial modeling may improve the estimation results in this setting, especially if the utilized covariates are either weak or non-available.

Since the propriety of the posterior distribution is vital for the correct application of the models, the new theorems allow statisticians to implement and use a wide variety of different spatial models applicable within the German Census 2011, including extensions of the frequently applied Battese-Harter-Fuller and Fay-Herriot model. Within the German Census 2011 many variables are categorical; these variables may be also modeled via the spatial Bernoulli mixed model.

A model comparison was performed in Chapter 6 via a simulation study. The aim was to elaborate possible benefits of spatial modeling. The study revealed that the proposed spatial models provide gains over standard small area models under various setups. Moreover, the spatial models were insensitive to different choices of the parameters of the prior distribution and the spatial correlation parameter, which is necessary for applications with political impact such as the German Census 2011. In addition, the study showed that carefully chosen spatial neighborhood structures may offer additional benefits of spatial modeling.

In addition, on basis of the theorems shown in this work, these methods can be implemented in practice. In particular for variables where only few covariate information is

present and suitable model building is challenging, the proposed models may improve the estimation within statistical censuses. For instance, in the context of the research project "Simulation der Strukturerhebung und Kleingebiet-Schätzungen für den schweizer Zensus" (a project with the focus to access the feasibility of small area estimation techniques for the Swiss Census 2010), recent simulations showed that standard small area estimators yield unsatisfactory results when estimating the minutes of the commute (cf. Münnich and Burgard, 2009, p. 19). Initial tests have shown that efficiency gains can be expected in this situation, when applying the proposed spatial models.

A Gibbs Sampler for a Proper and an Improper Model

This section contains the UMACS code of the Gibbs sampler for the proper and improper model of Example 2.16.

```
library(Umacs)
library(rv)
library(LearnBayes)

x.init<-function () rexp(1, 20)
y.init<-function () rexp(1, 20)

x.update <- function () {
    #rexp(1, rate = y)  #for the improper model
    rtruncated(1,0,20,pexp,qexp,y) #for the proper model
}

y.update<- function () {
    #rexp(1, rate = x)  #for the improper model
    rtruncated(1,0,20,pexp,qexp,y)#for the proper model
}

Sam <- Sampler(
    .title = "Improper Gibbs",
    x = Gibbs (x.update, x.init),
    y = Gibbs (y.update, y.init)
)

n.iter=500
Rhat.vec<-c(rep(0,20))

for (i in 1:20){
  Proper<-Sam(n.iter=n.iter,n.chains=2)
  n.iter=n.iter+500
  Rhat<-max(Proper@summary[,8])
  Rhat.vec[i]<-Rhat
}

sim.x<-Proper@chains[[1]][,1] #simulated values of chain 1
sim.y<-Proper@chains[[1]][,2] #simulated values of chain 2
```

Listing 3: Gibbs sampler for Example 2.16

B WinBUGS Models

This section contains the WinBUGS models for the application to unemployment data in Germany of Sections 3.5 and 4.5 and the application to SAIPE data of Section 5.3. These models were used throughout the simulation study of Chapter 6.

```
model {

  for (i in 1:n){
    y[i]~dnorm(theta[i],tau2.e)  # y is Normal distributed
  }

  for (i in 1:n){
    theta[i]<-b+u[region[i]]
  }

  u[1:k]~car.proper(mu2[],C[],adj[],num[],M[], tau2.u,gamma)
     #Spatial distribution

  for (i in 1:k){
    thetas[i]<-theta[n*i]  #Taking every n-th value
  }

  tau2.e<-1/sigma2e
  tau2.u<-1/sigma2u
  sigma2e<-(pow(BBBe,0.9))  #Construction of the power prior
  sigma2u<-(pow(BBBu,0.9))  #Construction of the power prior

  BBBe <-BBe*BBe
  BBBu <-BBu*BBu
  BBe~dflat()
  BBu~dflat()

  b~dflat()  #Prior of the regression coefficient

}
```

Listing 4: Spatial Normal Unit-level Model in WinBUGS

```
model {

  for (i in 1:n){
    y[i]~dnorm(theta[i],tau2.e)  # y is Normal distributed
  }

  for (i in 1:n){
    theta[i]<-b+u[region[i]]
  }

  for (i in 1:k){
    u[i]~ dnorm(0,tau2.u)  #Non-spatial
  }

  for (i in 1:k){
    thetas[i]<-theta[n*i]  #Taking every n-th value
  }

  tau2.e<-1/sigma2e
  tau2.u<-1/sigma2u
  sigma2e<-(pow(BBBe,0.9))  #Construction of the power prior
  sigma2u<-(pow(BBBu,0.9))  #Construction of the power prior

  BBBe <-BBe*BBe
  BBBu <-BBu*BBu
  BBe~dflat()
  BBu~dflat()

  b~dflat()  #Prior of the regression coefficient

}
```

Listing 5: Non-spatial Normal Unit-level Model in WinBUGS

```
model {

 for (i in 1:n){
   y[i]~dbern(theta[i])  # y is Bernoulli distributed
 }

 for (i in 1:n){
   logit(theta[i])<-p[i]  #Logit specification of the model
   p[i]<-b+u[region[i]]
 }

 u[1:k]~car.proper(theta2[],C[],adj[],num[],M[], tau2.u,gamma)
    #Spatial distribution

 for (i in 1:k){
   thetas[i]<-theta[n*i]  #Taking every n-th value
 }

 tau2.u<-1/sigma2u
 sigma2u<-(pow(BBBu,0.9))  #Construction of the power prior
 BBBu <-BBu*BBu
 BBu~dflat()

 b~dnorm(0.0001,0.0001)  #Prior of the regression coefficient

}
```

Listing 6: Spatial Bernoulli Model in WinBUGS

```
model {

  for (i in 1:N) {
    y[i]~dbern(theta[i]) # y is Bernoulli distributed
  }

  for (i in 1:N){
    logit(theta[i])<-p[i] #Logit specification of the model
    p[i]<-b+v[region[i]]
  }

  for (i in 1:k)  {
    v[i]~ dnorm(0,tau) # Non-spatial distribution
  }

  for (i in 1:k){
    thetas[i]<-theta[n*i] #Taking every n-th value
  }

  tau<-1/BBBB
  BBBB<-(pow(BBB,0.9)) #Construction of the power prior
  BB~dflat()

  b~dnorm(0.0001,0.0001) #Prior of the regression coefficient

}
```

Listing 7: Non-spatial Bernoulli Model in WinBUGS

```
model {

  for (i in 1:k) {
    y[i]~ dnorm(theta[i],D[i])
  }

  for (i in 1:k) {
    theta[i]<-b[1]+b[2]*x1[i]+b[3]*x2[i]+b[4]*x3[i]+b[5]*x4[i]+v[i]
  }

  v[1:k]~ car.normal(adj[],weights[],num[],tau)

  for (i in 1:sumNumNeigh) {
    weights[i] <- 1
  }

  for (i in 1:5) {
    b[i]~ dflat()
  }

  tau ~ dgamma(0.5,0.0005)
  var<-1/tau

}
```

Listing 8: Spatial Fay-Herriot Model in WinBUGS

```
model {

  for (i in 1:k) {
    y[i]~ dnorm(theta[i],D[i])
  }


  for (i in 1:k) {
    theta[i]<-b[1]+b[2]*x1[i]+b[3]*x2[i]+b[4]*x3[i]+b[5]*x4[i]+u[i]
  }

  for (i in 1:k)  {
    u[i]~ dnorm(0,tau) #Non-spatial distribution
  }

  for (i in 1:5) {
    b[i]~ dflat()
  }

  tau ~ dgamma(0.5,0.0005)
  sigmau<-1/tau

}
```

Listing 9: Non-spatial Fay-Herriot Model in WinBUGS

References

Albert, J. (**2007**): *Bayesian Computation with R*. Springer Science and Business Media, LLC.

Banerjee, S., **Carlin, B.** and **Gelfand, A. E.** (**2004**): *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman & Hall/CRC.

Battese, G. E., **Harter, R. M.** and **Fuller, W. A.** (**1988**): *An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data*. In: Journal of the American Statistical Association, 83 (401), p. 28–36.

Bavaud, F. (**1998**): *Models for spatial Weights: A systematic look*. In: Geographical Analysis, 30, p. 153–171.

Berger, J. O. (**1985**): *Statistical Decision Theory and Bayesian Analysis*. New York: Springer.

Bernardo, J. M. and **Smith, A. F. M.** (**1994**): *Bayesian Theory*. New York: John Wikey & Sons.

Besag, J. E. (**1974**): *Spatial Interaction and the statistical analysis of lattice systems*. In: Journal of the Royal Statistical Society B, 36, p. 192–225.

Besag, J. E. and **Kooperberg, C. L.** (**1995**): *On conditional and intrinsic autoregressions*. In: Biometrika, 82, p. 733–746.

Besag, J. E., **York, J.** and **Mollie, A.** (**1991**): *Bayesian image restoration with two applications in spatial statistics (with discussion)*. In: Annals of the Institute of Statistical Mathematics, 43, p. 1–59.

Bivand, R. S., **Pebesma, E. J.** and **Gómez-Rubio, V.** (**2008**): *Applied Spatial Data Analysis with R*. New York: Springer.

Box, G. E. P. and **Tiao, G. C.** (**1973**): *Bayesian Inference in Statistical Analysis*. New York: John Wiley & Sons.

Bundesagentur für Arbeit (**2008**): Arbeitsmarkt 2007. *Amtliche Nachrichten der Bundesagentur für Arbeit, 56, Sondernummer 2.*

Carlin, B. P. and **Louis, T. A.** (**2000**): *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton: Chapman & Hall/CRC.

Casella, G. and **George, E. I.** (**1992**): *Explaining the Gibbs Sampler*. In: The American Statistician, 46, p. 167–174.

Chen, M. H., **Shao, Q. M.** and **Xu, D.** (**2002**): *Necessary and sufficient conditions on the propriety of posterior distributions for generalized linear mixed models*. In: Sankhya: The Indian Journal of Statistics, 64, p. 57–85.

Congdon, P. (**2003**): *Applied Bayesian Modelling*. Chichester: John Wiley & Sons.

Congdon, P. (**2005**): *Bayesian Models for Categorical Data*. Chichester: John Wiley & Sons.

Congdon, P. (**2007**): *Bayesian Statistical Modelling*. Chichester: John Wiley & Sons.

Cressie, N. A. C. (**1993**): *Statistics for spatial data*. New York: John Wiley & Sons.

Datta, G. S. and **Ghosh, M.** (**1995**): *Some Remarks on Noninformative Priors*. In: Journal of the American Statistical Association, 90, p. 1357–1363.

Datta, G. S. and **Smith, D. D.** (**2003**): *On propriety of posterior distributions of variance components in small area estimation*. In: Journal of Statistical Planning and Inference, 112, p. 175–183.

Demidenko, E. (**2004**): *Mixed Models Theory and Applications*. New Jersey: John Wiley & Sons.

Destatis (**2010**): Destatis website. *http://www.destatis.de, Download: 20.01.2010.*

Eurarea (**2004**): *Enhancing Small Area Estimation Techniques to meet European Needs*. Project Reference Volume Part I-III.

Faraway, J. J. (**2006**): *Extending the Linear Model with R*. Boca Raton: Chapman & Hall/CRC.

Fay, R. E. and **Herriot, A.** (**1979**): *Estimates of Income for Small Places: An Application of James Stein Procedures to Census Data*. In: Journal of the American Statistical Association, 74 (366), p. 269–277.

Fishman, G. S. (**2006**): *A First Course in Monte Carlo*. Duxbury: Thomson Brooks/-Cole.

Gabler, S., **Ganninger, M.** and **Münnich, R.** (**2010**): *Optimal allocation of the sample size to strata under box constraints.*

Gelfand, A. E. and **Smith, A. F. M.** (**1990**): *Sampling based approaches to calculating marginal densities*. In: Journal of the American Statistical Association, 85, p. 398–406.

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

Gelman, A. and **Rubin, D. B.** (**1992**): *Inference from iterative simulation using multiple sequences*. In: Statistical Science, 7, p. 457–511.

Gelman, A. B., **Carlin, J. S.**, **Stern, H. S.** and **Rubin, D. B.** (**1995**): *Bayesian Data Analysis*. Boca Raton: Chapman & Hall/CRC.

Geman, S. and **Geman, D.** (**1984**): *Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, p. 721–741.

GeoBUGS User Manual (**2004**): GeoBUGS User Manual. Version 1.2.

German Mikrocensus (**2002**): Public use file. *http://www.forschungsdatenzentrum.de/bestand/mikrozensus/cf/2002/index.asp*, *Download: 05.11.2009.*

GESIS (**2010**): GESIS website. *http://www.gesis.org/dienstleistungen/daten/amtliche-mikrodaten/mikrozensus/*, *Download: 20.01.2010.*

Ghosh, J. K., **Delampady, M.** and **Samanta, T.** (**2006**): *An Introduction to Bayesian Analysis: Theory and Methods*. New York: Springer.

Ghosh, M., **Natarajan, K.**, **Stroud, T. W. F.** and **Carlin, P.** (**1998**): *Generalized Linear Models for Small Area Estimation*. In: Journal of the American Statistical Association, 93 (441), p. 273–282.

Ghosh, M., **Natarajan, K.**, **Waller, L. A.** and **Kim, D.** (**1999**): *Hierarchical Bayes GLMs for the analysis of spatial data: An application to disease mapping - detection and inference*. In: Journal of Statistical Planning and Inference, 75 (2), p. 305–318.

Ghosh, M. and **Rao, J. N. K.** (**1994**): *Small area estimation: An appraisal (with discussion)*. In: Statistical Science, 9, p. 55–93.

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

Griffith, D. A. (**1996**): *Specifying the Geographic Weights Matrix*. In: **Arlinghaus, S. L.** (Ed.) Practical Handbook of Spatial Statistics, Boca Raton: CRC.

Haining, R. (**2003**): *Spatial Data Analysis Theory and Practice*. Cambridge University Press.

Hobert, J. P. and **Casella, G.** (**1996**): *The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models*. In: Journal of the American Statistical Association, 91, p. 1461–1473.

Hurn, M. A., **Husby, O. K.** and **Rue, H.** (**2003**): *A Tutorial on Image Analysis*. In: **Moller, J.** (Ed.) Spatial Statistics and Computational Methods, Springer.

Ibrahim, J. G. and **Laud, P. W.** (**1991**): *On Bayesian analysis of generalized linear models using Jeffreys' prior*. In: Journal of the American Statistical Association, 86, p. 981–986.

James, W. and **Stein, C.** (**1960**): *Estimation with Quadratic Loss*. In: Proceedings of the Fourth Berkley Symposium of Mathematical Statistics and Probability, 66, p. 361–379.

Jiang, J. and **Lahiri, P.** (**2006**): *Mixed Model Prediction and Small Area Estimation.* In: Test, 15 (1), p. 1–96.

Kass, R. E. and **Wassermann, L.** (**1996**): *The selection of prior distributions by formal rules.* In: Journal of the American Statistical Association, 91, p. 1343–1370.

Kelsall, J. E. and **Wakefield, J. C.** (**2008**): Discussion of Bayesian models for spatially correlated disease and exposure data by Best et al. In: **J. M. Bernardo, A. P. D., J. O. Berger** and **Smith, A. F. M.** (Eds.) Bayesian Statistics 6, Oxford: Oxford University Press.

Lahiri, P. and **Meza, J. L.** (**2002**): *Small-area estimation. In A.H. El-Shaarawi and W. W. Piegorsch (eds.), Encyclopedia of Environmetric*, Band 4. Chichester: John Wiley & Sons, 2010–2014 p.

Longford, N. T. (**2005**): *Missing Data and Small-Area Estimation.* Springer Science+Business Media, Inc.

Magg, K., **Münnich, R.** and **Schäfer, J.** (**2006**): *Small Area Estimation im Zensus 2011.* In: **Forschungsdatenzentrum der Statistischen Landesämter** (Ed.) Amtliche Mikrodaten für die Sozial- und Wirtschaftswissenschaften. Beiträge zu den Nutzerkonferenzen des FDZ der Statisischen Landesämter 2005, p. 93–105.

Marker, D. A. (**1999**): *Organization of small area estimators using generalized regression framework.* In: Journal of Official Statistics, 15, p. 1–24.

Münnich, R. and **Burgard, J. P.** (**2009**): Simulation der Strukturerhebung und Kleingebiet-Schätzungen. Phase 1 Endbericht; Internal Report, University of Trier.

Münnich, R., **Burgard, J. P.** and **Vogt, M.** (**2010**): *Small area estimation for population counts in the German census 2011.* In: Proceedings of the Survey Research Methods Section, ASA, accepted.

Münnich, R., **Gabler, S.** and **Ganninger, M.** (**2007**): *Some remarks on the register-based census 2010/2011 in Germany.* In: Proceedings of the workshop In-

novative Methodologies for Censuses in the New Millennium, Southampton, 2007; http://www.s3ri.soton.ac.uk/isi2007/papers/Paper03.pdf.

Münnich, R. and **Schmidt, K.** (**2002**): *Small Area Estimation in der Bevölkerungsstatistik*. In: Baden-Wuerttemberg in Wort und Zahl, 3, p. 139–145.

Moura, F. A. S. and **Migon, H. S.** (**2002**): *Bayesian Spatial Models for Small Area Estimation of Proportions*. In: Statistical Modelling, 2, p. 183–201.

Mukhopadhyay, P. (**1998**): *Small Area Estimation in Survey Sampling*. New Delhi: Narosa Publishing House.

Natarajan, R. and **McCulloch, C.** (**1995**): *A note on the existence of the posterior distribution for a class of mixed models for binomial responses*. In: Biometrika, 82, p. 639–643.

Ntzoufras, I. (**2009**): *Bayesian Modeling Using WinBUGS*. New Jersey: John Wiley & Sons.

Pfeffermann, D. (**2002**): *Small area estimation - new developments and directions*. In: International Statistical Review, 70, p. 125–143.

Rao, J. N. K. (**1986**): *Synthetic estimators, SPREE and the best model based predictors*. In: Proceedings of the Conference on Survey Methods in Agriculture, U.S. Dept. of Agriculture, Washington DC, p. 1–16.

Rao, J. N. K. (**1999**): *Some recent advances in model-based small area estimation*. In: Survey Methodology, 25, p. 175–186.

Rao, J. N. K. (**2001**): EB and EBLUP in small area estimation in Empirical Bayes and Likelihood inference. In: **Ahmed, S. E.** and **Reid, N.** (Eds.) Lecture Notes in Statistics, Band 148, p. 33–43, New York: Springer Verlag.

Rao, J. N. K. (**2003**): *Small Area Estimation*. New York: John Wiley & Sons.

Ripley, B. D. (**2004**): *Spatial Statistics (Wiley Series in Probability and Statistics)*. Wiley-Interscience.

Robert, C. P. and **Casella, G.** (**2010**): *Introducing Monte Carlo Methods with R*. New York: Springer.

Spiegel (**2003**): *Ostdeutschland: Alle Klugen weg?* In: Spiegel, 43, p. 24–24.

Sun, D., **Tsutakawa, K.** and **He, Z.** (**2001**): *Propriety of posteriors with improper priors in hierarchical linear mixed models*. In: Statistica Sinica, 11, p. 77–96.

Sun, D., **Tsutakawa, R. K.** and **Speckman, P. L.** (**1999**): *Posterior Distribution of hierarchical models using CAR(1) distributions*. In: Biometrika, 86, p. 341–350.

Vogt, M. (**2007**): Small Area Estimation: Die Schätzer von Fay-Herriot und Battese-Fuller-Harter. Unpublished diploma thesis, University of Trier.

Vogt, M. and **Münnich, R.** (**2009**): *On the existence of a posterior distribution for spatial mixed models with Binomial responses*. In: Metron - International Journal of Statistics, LXVII (2), p. 199–207.

Website: Census (**2010**a): Statistical Offices of the Länder and the Federal Statistical Office. *http://www.zensus2011.de/Statistik-Portal/en/Zensus/*, *Download: 29.01.2010*.

Website: Census (**2010**b): Statistical Offices of the Länder and the Federal Statistical Office. *http://www.zensus2011.de/Statistik-Portal/en/Zensus/en_nutzen.asp*, *Download: 29.01.2010*.

Website: Census (**2010**c): Statistical Offices of the Länder and the Federal Statistical Office. *http://www.zensus2011.de/Statistik-Portal/en/Zensus/en_smallArea.asp*, *Download: 29.01.2010*.

Website: Census (**2010**d): Statistical Offices of the Länder and the Federal Statistical Office. *http://www.zensus2011.de/Statistik-Portal/en/Zensus/en_stichprobe.asp*, *Download: 29.01.2010*.

Witting, H. (**1985**): *Mathematische Statistik 1 - Parametrische Verfahren bei festem Stichprobenumfang*. Stuttgart: Teubner.