ROBUST EXTENSIONS OF THE FAY HERRIOT MODEL IN SMALL AREA ESTIMATION

SEBASTIAN WARNHOLZ

Theory, Implementation and Simulation Studies

February 1, 2016 – version 0.3

Sebastian Warnholz: *Robust Extensions of the Fay Herriot Model in Small Area Estimation,* Theory, Implementation and Simulation Studies, © February 1, 2016

SUPERVISORS:
Prof. Dr. Timo Schmid
Nikos Tzavidis

LOCATION:
Berlin

TIME FRAME:
February 1, 2016

CONTENTS

1 INTRODUCTION

This thesis introduces extensions to the Fay-Herriot Model, a frequentist approach to Empirical Bayes estimators (i.e. James-Stein estimators) frequently used in Small Area Estimation which is a sub-field within the field of Statistics. In the three parts of this thesis I provide methodological extensions to existing statistical models (i *Theory*), considerations for implementing the findings as software (4 *Implementation*) and results on numerical stability as well as statistical properties of the introduced estimators as well as a short case study (iii *Results*).

Small Area Estimation ...

Robust Methods in Small Area Estimation ...

Software in stats ...

The thesis is seperated into three main parts. i *Theory* introduces the underlying esimation methodology, i.e. linear mixed models, a review of model based methods in Small Area Estimation as well as outlier robust extensions within the field. Given these results extensions to existing methodology is introduced in the form of a robustified Fay-Herriot estimator with optional spatial and temporal correlated random effects. A special interest lies in the concrete implementation of such (robust) estimators, and to meet this focus several algorithms are proposed. (MSE, bias-correction)

4 *Implementation* introduces three main aspects: The verification that the implementation (in terms of software) is correct; How to evaluate the numerical accuracy and stability of the introduced algorithms; And which results to report to judge the quality of the numerical solution.

In iii *Results* the properties of the estimators are investigated in simulations and in the context of part 4. The numerical properties are devided into accuracy and stability. Statistical properties are shown for the most reliable implementations using model and design based simulation studies.

Part I

THEORY

This is the chapter where I want to present the theoretical concepts underpinning the development of software and application.

2 SMALL AREA ESTIMATION

2.1 OVERVIEW

In the following I give a general overview of the field of Small Area Estimation (SAE) necessary in order to accomodate this thesis within the SAE field. Rao (2003) as well as Rao and Molina (2015) give a comprehensive overview of the established methods and research research published in this field. Ghosh and Rao (1994), Rao (1999), Pfeffermann (2002) and Pfeffermann (2013) focus on the status quo and main lines of discussion within the field at their respective point in time.

The main endevour Small Area Estimation tries to solve is to produce *reliable predictions* of a *target statistic* for *small domains*. A *target statistic* can be a simple statistic such as a mean, a count, or quantiles; but it can take other forms: e.g. an inequality measure like the Gini coefficient for poverty mapping. Such statistics are produced for *small domains*, where domain refers to specific groups, e.g. a sector of industry, or groups defined by socio-economic characteristics. Because of its frequent application to administrative data, domains are often defined by areas as a geographical unit. They are small in the sense that they comprise few or no sampled units within these domains. This has the effect that a direct estimation, i.e. an estimation which only relies on the information available within domains, tends to be not reliable. *Reliability* is here measured either by the variance or mean squared error of the predictions.

Small Area Estimation tries to improve such domain predictions, often in terms of mean squared error, by *borrowing strength* from other domains. This can happen by taking additional information from other data sources into account, like census and register information. Also structures in the data, like spatial or temporal correlation, can be exploited to improve a prediction.

The importance of the field can be explained by the increasing demand for reliable estimates by policy makers and in official statistics. Results may be used for fund allocation, health programs, agriculture, or poverty mapping to name only a few of the fields of application. Traditionally such estimates rely on survey data; but as the target domains become more diverse, reliable estimates are connected to an increasing demand for sampled units within the domains. The conflict between the demand for more diverse domains as against the cost and feasibility for larger samples is the factor that stimulates the

progress within the field as this supplies the mechanism for optimising the ratio between sampled units and the reliability of estimates.

In general small area methods are divided into two categories: design based and model based methods. Design based methods can be considered the traditional methodology for analysing survey data; a comprehensive overview of these methods for SAE can be found in Lehtonen and Veijanen (2009). Design based methods comprise different direct and indirect techniques. The Horvitz-Thompson (HT) estimator (Horvitz and Thompson (1952)) which only uses sampled units within domains, synthetic regression estimates as well as model assisted methods like generalized regression (GREG) estimators (Särndal et al. (1992)) are examples of such estimators. What all these methods have in common is that they incorporate information of the sampling design into the estimation.

Conceptually design based and model based methods differ in that design based methods are used to optimally estimate a target parameter of a fixed and finite population. Model based methods rely instead on the idea that an observed sample is drawn from a population which is but one possible realisation of a *superpopulation* model, and it is the parameters of that superpopulation which are targeted. This difference leads to a trade-off when choosing between methods: model based methods can improve domain estimation in terms of variance even with small samples, however, they can not be considered design unbiased. Design based methods on the other hand are design unbiased but have larger and possibly unacceptably high variances for small samples (see Lehtonen and Veijanen (2009)).

Model based methods can be further separated into area and unit level models. Observations which can be associated with a specific domain are referred to as units. These can be companies within an industry sector or individuals within a municipality. The area level describes models which use information on area level, i.e. direct estimates for domains. A situation in which these models are considered is when data can only be provided as aggregates due to factors such as confidentiality.

One class of models in particular is favoured in different variations: mixed models. The Fay-Herriot (FH) model introduced by Fay and Herriot (1979) and the Battese-Harter-Fuller (BHF) model which was introduced by Battese et al. (1988) are the two basic models which are used respectively for area and unit level models. Underlying this is the idea to use auxiliary information in a regression to estimate a global conditional mean and add an extra component to capture the domain specific difference from that global mean. This general idea can be found in combination with different estimation methodologies, i.e. general linear mixed models which are typically associated with best linear unbiased predictors (BLUPs), empirical Bayes, and hierarchical Bayes. Although these different frameworks for estimation

differ with respect to optimality criteria, equivalence of the different derived estimators can be shown for special cases. A more general discussion of mixed models in SAE can be found in Jiang and Lahiri (2006). Rao (2003) and Rao and Molina (2015) provide a comprehensive overview and comparison of the different frameworks.

A general property of model based methods is that a lot of their benefits in terms of efficiency rely on strong distributional assumptions. Hence it is not only in the field of SAE that robust methods have been exploited to reduce the negative effect of a potential violation of these assumptions. The general problem here is that single observations can have unwanted and overly large impact on results. Such observations are typically called outliers. Chambers (1986) uses the term *representative* outliers to describe observations which are correctly recorded and can not be assumed to be unique. Non-representative outliers, on the other hand, may be best described as *not correctly recorded* and should be imputed or generally dealt with during the editing process of survey data.

To summarise robust methods in SAE it is necessary to distinguish between three different lines of discussion. First, if the distributional assumption - often a Gaussian distribution - appears to be implausible then intuition demands that it be replaced. This often leads to the use of non-symmetric or heavy-tailed distributions for the model error or the random effect. Due to their flexibility Bayesian modelling strategies are often used in this context; see for example Datta and Lahiri (1995) and Bell and Huang (2006). Secondly, methods are applied which are *naturally* more robust against outlying observations. Tzavidis et al. (2010) model a global conditional median, or more generally a quantile, instead of a mean. The third approach is to remain with the original model or method and *robustify* the estimation equations. In this context Sinha and Rao (2009) develop a robustified EBLUP; Beaumont and Rivest (2009) refer to a winzorisation of the Horvitz-Thompson estimator; and Beaumont (2004) introduces a robust extension to generalized regression estimation.

Given this background, in this thesis I introduce extensions to the Fay-Herriot area level model using a model based perspective. More precisely an EBLUP based approach is taken to derive predictions in a way that makes it possible to model spatial and temporal covariance structures in the random effects. These models are closely related to the results presented by Marhuenda et al. (2013), who introduce spatial and temporal extensions to the Fay-Herriot model. However the methods in this thesis are based on an estimation procedure which is robust against outliers following the methodology introduced by Sinha and Rao (2009). The extension introduced in the literature around robust EBLUPs (REBLUPs) have been focused on unit level models, thus results especially to MSE estimation and bias correction are extended to area level models.

In addition to the main focus of this Thesis reference is made to certain supplementary material on software having relevance to this thesis. Warnholz and Schmid (2015) introduces tools for simulation studies for the special case of small area estimation and (?) implements the methods introduced in this thesis. All the results in this thesis rely on these two packages; a more detailed discussion can be found in chapter (?).

2.2 SMALL AREA PREDICTION USING LINEAR MIXED MODELS

This section gives a general overview of linear mixed models and the best linear unbiased prediction (BLUP) and empirical BLUP (EBLUP). The unit and area level models introduced here are based on this class of models and the robust methodology is based on a robustified EBLUP. The original model dates back to Henderson (1950) and an early comprehensive overview can be found in Searle (1971). Jiang and Lahiri (2006) review liner mixed models in the context of SAE and the main results can also be found in Rao (2003) and Rao and Molina (2015).

2.2.1 *Linear Mixed Models*

A linear mixed model can be expressed by:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{2.1}$$

where $\mathbf{y}$ is the $(n \times 1)$ vector of response values; $\mathbf{X}$ is a $(n \times P)$ matrix containing auxiliary and deterministic information; β is the $(P \times 1)$ vector of regression coefficients; $\mathbf{Z}$ is a known matrix and $\mathbf{u}$ is a vector of random effects, such that $\mathbf{Z}\mathbf{u}$ is of dimension $(n \times 1)$; $\mathbf{e}$ is the $(n \times 1)$ vector of model errors. Note that $\mathbf{u}$ and $\mathbf{e}$ are both random variables where the basic assumption is that both have mean zero and finite variances. Furthermore, they are assumed to be independent.

If, in addition, $\mathbf{u}$ and $\mathbf{e}$ are assumed to follow a normal distribution, the model is called a Gaussian Linear Mixed Model. The distribution of $\mathbf{y}$ can then be derived as a multivariate normal of the form:

$$\begin{aligned} \mathbf{y} &\sim \mathcal{N}(\mathbf{X}\beta, \mathbf{V}) \\ \mathbf{y}|\mathbf{u} &\sim \mathcal{N}(\mathbf{X}\beta + \mathbf{Z}\mathbf{u}, \mathbf{V}_e) \end{aligned}$$

where $\mathbf{V} = \mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top + \mathbf{V}_e$ with $\mathbf{V}_u$ and $\mathbf{V}_e$ being the variance covariance matrices of $\mathbf{u}$ and $\mathbf{e}$, respectively. Such variance structure typically depend on some unknown dispersion parameters. To be more precise: $\mathbf{V}_u = \mathbf{V}_u(\delta_u)$ and $\mathbf{V}_e = \mathbf{V}_e(\delta_e)$ such that $\mathbf{V} = \mathbf{V}(\delta)$ with $\delta = (\delta_u, \delta_e)$.

2.2.2 *Best Linear Unbiased Prediction*

Given the model (2.1) above in SAE problems we are generally interested in estimating the expected value of $\mathbf{y}$ given $\mathbf{u}$:

$$\mu = \mathbf{l}^\top \beta + \mathbf{m}^\top \mathbf{u}$$

for specified values of $\mathbf{l}$ and $\mathbf{m}$. An estimator for μ can be obtained by replacing β and $\mathbf{u}$ with suitable estimators. For known variance components, δ, the best linear unbiased estimator (BLUE) is given by:

$$\tilde{\beta} = \tilde{\beta}(\delta) = \left(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y} \tag{2.2}$$

and the BLUP for $\mathbf{u}$ by:

$$\tilde{\mathbf{u}} = \tilde{\mathbf{u}}(\delta) = \mathbf{V}_u \mathbf{Z}^\top \mathbf{V}^{-1} \left(\mathbf{y} - \mathbf{X}\tilde{\beta}\right)$$

such that the BLUP estimator for μ can be stated as:

$$\tilde{\mu} = \tilde{\mu}(\delta) = \mathbf{l}^\top \tilde{\beta} + \mathbf{m}^\top \tilde{\mathbf{u}}. \tag{2.3}$$

The BLUP (2.3) of μ depends on known variance components δ. These values are typically unknown in applications and are themselves subject to estimation. If we replace δ with a suitable estimator, $\hat{\delta}$, the empirical BLUP (EBLUP) obtained is:

$$\hat{\mu} = \hat{\mu}(\hat{\delta}) = \mathbf{l}^\top \hat{\beta} + \mathbf{m}^\top \hat{\mathbf{u}} \tag{2.4}$$

where $\hat{\beta} = \tilde{\beta}(\hat{\delta})$ and $\hat{u} = \tilde{u}(\hat{\delta})$. To estimate δ a variety of estimators have been proposed. Commonly used estimators are based on maximum likelihood (ML) and restricted maximum likelihood (REML). The standard procedures in the literature are not directly feasible for the robust estimators introduced in chapter 3. Instead different algorithms are proposed. For a detailed discussion of the estimation of the variance parameters see Jiang and Lahiri (2006, pp. 9-11) and the literature quoted there.

2.2.3 *Mean Squared Prediction Error*

One of the main reasons for relying on small area methods is to reduce the mean squared error of domain predictions. Since domain predictions under a linear mixed model are derived as the EBLUP, we are generally interested in the mean squared prediction error (MSPE) of the EBLUP. Note that here I present results directly for the EBLUP instead of the BLUP since the latter has little practical relevance and the results are mainly needed to give a comprehensive context in which the existing literature can be extended. In general the estimation of the MSPE can be identified as one of the challenging problems in model based SAE (cf. Pfeffermann (2013)). Two approaches

are taken in the literature: The analytical identification of the MSPE and different resampling strategies. More detailed reviews of these strategies can be found in Rao (2003, p. 95 ff), Jiang and Lahiri (2006, p. 13 ff) and in Datta (2009).

Early results can be found in Kacker and Harville (1984) who propose an approximation to the MSPE of the EBLUP of a Gaussian linear mixed model. In particular they show that:

$$\mathrm{MSPE}(\hat{\mu}) = \mathrm{MSPE}(\mu) + \mathbb{E}\left(\hat{\mu} - \mu\right)^2. \tag{2.5}$$

The MSPE of μ can be decomposed such that

$$\mathrm{MSPE}(\mu) = g_1(\delta) + g_2(\delta)$$

(cf. Rao (2003, p. 98 ff)) where

$$g_1(\delta) = \mathbf{m}^\top \left(\mathbf{V}_u - \mathbf{V}_u \mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z} \mathbf{V}_u\right) \mathbf{m}$$
$$g_2(\delta) = \mathbf{d}^\top \left(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}\right) \mathbf{d}$$

with $\mathbf{d}^\top = \mathbf{l}^\top - \mathbf{b}^\top \mathbf{X}$ and $\mathbf{b}^\top = \mathbf{m}^\top \mathbf{V}_u \mathbf{Z}^\top \mathbf{V}^{-1}$. The second term in equation (2.5) was approximated by Kacker and Harville (1984) using a Taylor series approximation. Prasad and Rao (1990) have used instead the approximation by

$$\mathbb{E}\left(\hat{\mu} - \mu\right)^2 \approx g_3(\delta)$$

where

$$g_3(\delta) = \mathrm{tr}\left(\frac{\partial \mathbf{b}^\top}{\partial \delta} \mathbf{V} \left(\frac{\partial \mathbf{b}^\top}{\partial \delta}\right) \mathbf{V}_{\hat{\delta}}\right)$$

with $\mathbf{V}_{\hat{\delta}}$ being the asymptotic covariance matrix of $\hat{\delta}$. The derived MSPE depends on the unknown parameter vector δ. Prasad and Rao (1990) used a moment estimator for δ and replaced the above formulae so that an estimator of the MSPE can be defined as $\widehat{\mathrm{MSPE}}(\hat{\mu}) = g_1(\hat{\delta}) + g_2(\hat{\delta}) + g_3(\hat{\delta})$. Datta and Lahiri (2000) extended this approach for a wider range of models in SAE including ML and REML estimators. A good overview and comparison of the slightly different approaches can be found in Datta, Rao, et al. (2005) which is more specific in the sense that it focuses on area level models.

A different approach has been taken by Chambers, Chandra, and Tzavidis (2011) which is to define the EBLUP as a weighted sum of the sampled values and to derive an MSPE estimator under the assumption of independence between weights and sampled values. An advantage of this method is its wide applicability as the approach is not restricted to predictions under linear mixed models but extend to any predictor which can be represented as a weighted sum of the

sampled values. The approach has been extended in Chambers, Chandra, Salvati, et al. (2014) to robust methods in SAE and is of special interest in deriving an MSPE estimator for the methods proposed in this thesis. Hence these results are reviewed in more detail in Section 2.6.3.2.

As an alternative approach a wide range of different resampling strategies have been proposed. Jiang, Lahiri, and Wan (2002) introduce a jackknife method to estimate the MSPE in the context of longitudinal linear mixed models; this was modified by Lohr and Rao (2009) into a simpler form. Also important is the proposed double bootstrap method of Hall and Maiti (2006). However of special interest are the methods in the context of robust predictions under linear mixed models which is why the bootstrap methods of Sinha and Rao (2009) and Jiongo and Nguimkeu (2014) are reviewed in more detail in Section 2.6.3.1.

2.3 DIRECT ESTIMATORS

In this section I introduce a simple direct estimator and notation which is then used throughout the thesis. The estimator is of interest as it will be used in the simulation studies in Part iii. Lehtonen and Veijanen (2009) give a more comprehesive overview of direct estimators and design based methods.

Let a population be denoted by U, consisting of N units, which can be devided into D distinct domains or areas $U_1 \cup \cdots \cup U_D$. Each area is of size N_i, with $i = 1, \ldots, D$ denoting area i, such that $N = \sum_{i=1}^{D} N_i$. S denotes a sample from U which in turn can be devided into $S_1 \cup \cdots \cup S_D$, where n is the sample size of S and n_i of S_i such that $n = \sum_{i=1}^{D} n_i$.

Let y define the characteristic of interest and y_{ij} the response value for unit j with $j = 1, \ldots, N_i$ in area i. Furthermore let the target quantity be the population mean in area U_i defined by $\theta_i = \bar{Y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}$. Note that the target quantity is not necessarily a mean but can be any statistic in the population.

Given that the sample is drawn with simple random sampling without replacement (SRSWOR), a design unbiased estimator is given by

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \tag{2.6}$$

with its design variance $\mathbb{V}(\bar{y}_i | n_i) = (S_i^2/n_i)(1 - n_i/N_i)$ where $S_i^2 = 1/(N_i - 1) \sum_{j=1}^{N_i} \left(y_{ij} - \bar{Y}_i\right)^2$. To estimate the sampling variance, S_i^2 can be replaced by s_i^2 being an estimator using the sampled data, i.e. $s_i^2 = 1/(n_i - 1) \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y}_i\right)^2$. Note that n_i is random under SRSWOR such that specific areas can have very few or even no sampled units.

2.4 AREA LEVEL MODELS

In SAE mixed models are generally devided into area and unit level models. In this section a review of some of the more important results is given with respect to area level models. Section 2.5 then presents the basic unit level model.

2.4.1 *The Fay-Herriot Model*

The basic area level model was introduced by Fay and Herriot (1979) and has been used to predict the mean income of small areas using census data. The general backdrop is that only information on the area level is available, i.e. the direct estimates for the domains. The model is then built around two stages. The first stage is the sampling model:

$$\tilde{y}_i = \theta_i + e_i$$

where $\tilde{y}_i$ is a direct estimator for a statistic of interest, θ_i, for an area i with $i = 1, \ldots, D$. The sampling error e_i is assumed to be independent and normally distributed with known variances σ^2_{ei}, i.e. $e_i \sim \mathcal{N}(0, \sigma^2_{ei})$. The model is modified with a second stage, the linking model, by assuming a linear relationship between the true area statistic, θ_i, and some deterministic auxiliary variables $\mathbf{x}_i$:

$$\theta_i = \mathbf{x}_i^\top \beta + u_i$$

where $\mathbf{x}_i$ is a $(P \times 1)$ vector containing area-level information for P variables and β is a $(P \times 1)$ vector of regression coefficients. The model errors u_i are assumed to be independent and identically distributed following a normal distribution: $u_i \sim \mathcal{N}(0, \sigma^2_u)$. Furthermore e_i and u_i are assumed to be independent. Combining the sampling and the linking model thus leads to:

$$\tilde{y}_i = \mathbf{x}_i^\top \beta + u_i + e_i. \tag{2.7}$$

2.4.1.1 *Best Linear Unbiased Prediction*

Obtaining small area predictions under the model in (2.7) can be defined as a linear mixed model from which a BLUP and EBLUP can be derived. Basically tha model in (2.7) can be viewed directly as a linear mixed model as it was introduced in equation (2.1) where $\mathbf{Z} = \mathbf{I}_D$, $\mathbf{V}_u = \sigma^2_u \mathbf{I}_D$ and $\mathbf{V}_e = \text{diag}(\sigma^2_{e,1}, \ldots, \sigma^2_{e,D})$ with $\mathbf{I}_D$ being a $(D \times D)$ identity matrix. The vector of unknown variance parameters in this case is a scalar, such that $\delta = \sigma^2_u$ since σ^2_{ei} is assumed to be known.

The BLUP under the Fay-Herriot model, defined in equation (2.3), can then be obtained by setting $\tilde{\mu}_i = \tilde{\theta}_i^{FH}$, $\mathbf{l}_i^\top = \mathbf{x}_i^\top$, $\mathbf{m}_i^\top = 1$ and $y_i = \tilde{y}_i$:

$$\begin{aligned}\tilde{\theta}_i^{FH} = \tilde{\theta}_i^{FH}(\sigma_u^2) &= \mathbf{x}_i^\top \tilde{\beta} + \tilde{u}_i \\ &= \mathbf{x}_i^\top \tilde{\beta} + \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{ei}^2}\left(\tilde{y}_i - \mathbf{x}_i^\top \tilde{\beta}\right) \\ &= \gamma_i \tilde{y}_i + (1-\gamma_i)\mathbf{x}_i^\top \tilde{\beta} \qquad (2.8)\end{aligned}$$

with $\gamma_i = \sigma_u^2/(\sigma_u^2 + \sigma_{ei}^2)$. The BLUP depends on the variance parameter of the random effects, σ_u^2, which is unknown. To obtain the EBLUP under the Fay-Herriot model we can replace the unknown parameter with an estimate, leading to:

$$\hat{\theta}_i^{FH} = \hat{\gamma}_i \tilde{y}_i + (1-\hat{\gamma}_i)\,\mathbf{x}_i^\top \hat{\beta}$$

where $\hat{\gamma}_i = \hat{\sigma}_u^2/(\hat{\sigma}_u^2 + \sigma_{ei}^2)$. Note that the regression parameters are still estimated using the weighted least squares estimator of equation (2.2) with $\delta = \hat{\sigma}_u^2$.

For the estimation of σ_u^2 different approaches exist. Fay and Herriot (1979) propose a moment estimator from which they derive an algorithm to estimate σ_u^2. Rao (2003, pp. 118-119) reviews several other ideas. Also based on a moment estimator he derives the estimator also used in Prasad and Rao (1990). Both moment estimators have the property that they do not rely on a normal distribution, which is also true for the estimation of the regression coefficients. Alternatively σ_u^2 can be estimated using maximum likelihood or restricted maximum likelihood, which, in contrast, relies on the distributional assumptions. For details see also Datta and Lahiri (2000).

2.4.1.2 *Mean Squared Prediction Error*

The MSPE of the EBLUP under the Fay-Herriot model, $\hat{\theta}_i^{FH}$, is the subject of several studies. However it needs to be noted that interestingly Fay and Herriot (1979) did not assess the quantification of uncertainty associated with their predictions. In principle the MSPE can be defined for an EBLUP as it was discussed in section 2.2.3. Datta, Rao, et al. (2005) study the MSPE estimation using the results of Prasad and Rao (1990) for different estimations of the variance component σ_u^2. They compare the method of moment estimator of Prasad and Rao (1990) with the original estimator of Fay and Herriot (1979) and the maximum likelihood estimator proposed by Datta and Lahiri (2000). The main finding is that for the MSPE estimation the estimator by Fay and Herriot (1979) performs the best overall.

A jackknife MSPE estimator for linear mixed models was introduced by Jiang, Lahiri, and Wan (2002) and later subjected to several refinements; see for example Chen and Lahiri (2003). Chen and Lahiri (2008) then introduced a jackknife estimator based on the results of

Jiang, Lahiri, and Wan (2002) and explicitly targeted the MSPE estimation of the prediction under a Fay-Herriot model. They found satisfying results for the MSPE estimation using their method; however the conclusion with respect to the estimation of σ_u^2 is not as clear as in Datta, Rao, et al. (2005), where they compare a method of moment estimator to the method proposed by Fay and Herriot (1979).

A different line of discussion is stimulated by the fact that it is assumed that the sampling variances, σ_{ei}^2, are known parameters. In practice this is not the case and these parameters are estimated using the sample data. This can mean that they are themselves direct estimators; but if a direct mean is considered unreliable then its variance estimation cannot be considered reliable. Fay and Herriot (1979) use generalized variance functions - see Wolter (2007, pp. 272 ff) for a discussion of these methods - instead of direct estimators. Maiti et al. (2014) suggest instead to shrink both means and variances to account for the possibility of unstable direct variance estimates and also provide an estimator of the MSPE of the predictions. This approach is based on a Bayesian modelling strategy. You and Chapman (2006) provide results for the case that direct variance estimates are used in a hierarchical Bayes approach and can account for that extra variability in the MSPE estimation. MSPE estimators for an EBLUP based prediction using estimated sampling variances can be found in Wang and Fuller (2003) and in Rivest and Vandal (2003). Wang and Fuller (2003) derive an MSPE using asymptotic properties of the EBLUP. Rivest and Vandal (2003) on the other hand extend the results of Prasad and Rao (1990) and add an extra term to the MSPE estimator to account for the additional variability associated with the estimation of direct sampling variances.

2.4.1.3 *Discussion*

From a practical point of view the assumption of known sampling variances under the model is not plausible. Hence these variances, though subject to estimation, are treated as known constants. Some approaches to deal with this problem have been reviewed in section 2.4.1.2 because in principle this relates mostly to an underestimation of the true uncertainty of the predictions. Another dimension which was for example addressed by Maiti et al. (2014) is the instability of predictions when very heterogenous sampling variances are observed, and to deal with this by shrinking both means and variances. A further approach is to stabilise the sampling variances by using generalized variance functions or other smoothing techniques. Albeit these parameters are assumed to be known they can have a large impact on the validity of domain predictions. In section (?) I show how the robust FH model relates to this discussion in terms of stability and MSPE.

The response variable, $\tilde{y}_i$, denotes a direct estimator. This is, of course, not necessarily the sample mean but can be any other statistic. An important feature of this statistic is that it is design unbiased. So in principle it can be a direct design based estimator such as the HT estimator. However it is assumed that the sampling errors are independent. This can be a plausible assumption under simple random sampling but is not necessarily valid under an informative sampling design. With respect to the response variable it also needs to be noted that quite often it is not the direct estimator itself but a suitable transformation that is used. Fay and Herriot (1979) log-transform the direct estimator and suggest the use of a transformation such that a normal distribution is plausible, i.e. is supported by the observed data. Hence several suggestions have been made as to how to optimally transform the response variable and how to assess the estimation of the MSPE of the back transformed domain prediction; see for example Slud and Maiti (2006). Sugasawa and Kubokawa (2015) review and introduce several parametric transformations for the FH model and also discuss the possibility of MSPE estimation.

One of the main motivations to consider area level models is based on the availability of data. Especially with census or administrative data it may not be possible to give unit level information directly to the analyst due to reasons of confidentiality. Thus only aggregates, i.e. direct domain estimations, are available. Even with the availability of information there are other reasons to consider. One is the integration of sampling weights which in general is not directly feasible in model based methodology. Area level models present a way to at least incorporate design weights into the direct estimation and then have a design unbiased estimator on the area level. Other reasons can be practical considerations, e.g. the reduction of computational demands because area level data is often connected with a dramatic reduction in the number of observations hence more complex variance structures can be modelled with less computational effort. Other reasons are discussed in more detail in Namazi-Rad and Steel (2015) who consider different scenarios for the availability of auxiliary information, e.g. unit and area level variables and contextual variables. Their findings show that overall unit level models have more potential to reduce the MSPE of domain predictions; this is not surprising as parameter estimation under a unit level model uses more information and thus is more precise. This leaves the availability of data as the main reason to consider area level models.

Area level models in small area estimation have been used in many applications and the Fay-Herriot model is explicitly subject to numerous extensions. See for example Clement (2014) for a review in disease mapping; Guadarrama et al. (2015) for a review of methods used for poverty mapping; and Benavent and Morales (2016) for an EBLUP under a multivariate FH model. A comprehensive review of extension

can be found in Rao (2003, pp. 153 ff) and Rao and Molina (2015, pp ?). The following section reviews some advances for incorporating structures in space and time into the estimation process. Some of these results in addition to a robust extension is then the subject of chapter (?).

2.4.2 *Spatial and Temporal Fay-Harriot Models*

This section reviews spatial and temporal extensions to the FH model. These extensions are again subject of chapter (?) where they are then combined with robust estimation methodology which is reviewed in section (?). It is in principle intuitive that if historical data, e.g. yearly repeated surveys, are available it should be exploited; and the same is true for spatial structures. From a mixed model perspective two components can be modified in order to take such information into account: the random effect and the model error term. The use of correlated random effects to allow for spatial, temporal, or spatio-temporal effects may be beneficial with respect to domain predictions. In constrast to modifications to the model error this may add to the improvement of domain predictions in terms of MSPE and is favoured in the literature. Modifying the model error can be motivated to account for correlated sampling errors of the direct estimators. This may be of special interest if, for example, the area level information is based on panel surveyes.

2.4.2.1 *Spatial Extensions*

The standard FH model (2.7) uses a random effect to capture unobserved variation between areas. However it ignores unobserved spatial patterns which can be present when areas refer to geographical units. In the standard FH model the area specific random effects are independent; when spatial correlation is an issue then instead the assumption of correlated random effects appears to be more plausible because neighbouring areas may be similar. If no covariates are present to capture such effects Molina et al. (2009) showed that taking spatial correlation into account can be beneficial for domain predictions. B. B. Singh et al. (2005), Petrucci and Salvati (2006) and Pratesi and Salvati (2008) investigated the possibility to incorporate a simultaneous autoregressive process (SAR) into the domain predictions. In principle there are two options to model spatial correlation: conditional autoregressive models (CAR) and SAR. The main difference is that CAR models are based on a Markov field which implies that spatial correlation exists only locally between two or more neighbours but does not effect other non neighbouring units, i.e. such a process has no memory across space. SAR processes do not have this restriction and are more useful to describe a global correlation structure. For a comprehensive overview of different approaches see Cressie

(1993). To incorporate spatial correlation we can modify model (2.7) such that:

$$\tilde{y}_i = \mathbf{x}_i^\top \beta + u_{1i} + e_i \tag{2.9}$$

where the only difference to model (2.7) is to replace u_i with u_{1i} to represent the random effect. In contrast to the formulation in model (2.7) u_{1i} now follows a simultaneous autoregressive process of order one (SAR(1)) which is defined by:

$$u_{1i} = \rho_1 \sum_{l \neq i} w_{il} u_{1l} + \epsilon_{1i}$$

where $|\rho_1| < 1$ and $\epsilon_{1i} \sim \mathcal{N}(0, \sigma_1^2)$ are i.i.d. with $i = 1, \ldots, D$. w_{il} are the elements of $\mathbf{W}$ which is the row standardized proximity matrix $\mathbf{W}^0$. The elements in $\mathbf{W}^0$ are equal to 1 if areas are neighboured and 0 otherwise, it then follows that the dimension of $\mathbf{W}^0$ is $D \times D$. Using the methodology of section 2.2 the BLUP can be stated as:

$$\tilde{\theta}_i^{SFH} = \mathbf{x}_i^\top \tilde{\beta} + \tilde{u}_{1i}$$

where $\tilde{\theta}_i^{SFH}$ depends on the variance parameters $\delta = (\rho_1, \sigma_1^2)$. Pratesi and Salvati (2008) use a maximum likelihood estimator for both parameters. Replacing the unknown parameters with their respective estimates the EBLUP can be obtained by:

$$\hat{\theta}_i^{SFH} = \mathbf{x}_i^\top \hat{\beta} + \hat{u}_{1i}. \tag{2.10}$$

Note that this model is again subject of section (?) where the variance structure is reviewed in more detail. The MSPE for the EBLUP under model (2.9) is developed by Pratesi and Salvati (2008) who extend the results from Prasad and Rao (1990).

A different approach to model spatial correlation structures can be found in Chandra et al. (2012) who use a geographically weighted regression model and derive the EBLUP under this model. Salvati et al. (2012) use a geographically wighted M-quantile regression model for small area predictions. Porter et al. (2014) is an example how to use a CAR process in the context of the FH model, which is, additionally, extended to allow for functional covariates.

2.4.2.2 *Temporal Extensions*

An early extension by Choudhry and Rao (1989) allows for an autoregressive process in the error term, which, in the context of the FH model, is one way to account for correlated sampling errors. Rao and Yu (1994) then introduced a model to use auto correlated random effects to borrow strength for domain predictions:

$$\tilde{y}_{it} = \mathbf{x}_{it}^\top \beta + u_i + u_{2it} + e_{it} \tag{2.11}$$

where $i = 1, \dots, D$ and $t = 1, \dots, T$ where D and T are the total number of areas and time periods respectively. Here $e_{it} \sim \mathcal{N}(0, \sigma^2_{eit})$ are independent with known variances σ^2_{eit}. u_i corresponds to the random effect in model (2.7) and follows a normal distribution with zero mean and variance σ^2_u. Correlation over time is now incorporated by adding u_{2it} which is a correlated random effect following a AR(1), i.e. an autoregressive process of order one:

$$u_{2it} = \rho_2 u_{2i,t-1} + \epsilon_{2it}$$

where ρ_2 is the auto correlation coefficient with $|\rho_2| < 1$ and $\epsilon_{2it} \sim \mathcal{N}(0, \sigma^2_2)$ being i.i.d. with $i = 1, \dots, D$ and $t = 1, \dots, T$. The BLUP under model (2.11) can then be defined as:

$$\tilde{\theta}^{TFH}_{it} = \mathbf{x}^\top_{it}\tilde{\beta} + \tilde{u}_i + \tilde{u}_{2it}$$

where the variance parameters $\delta = (\sigma^2_u, \rho_2, \sigma^2_2)$ are assumed to be known. Replacing these parameters with their respective estimates the EBLUP can be obtained:

$$\hat{\theta}^{TFH}_{it} = \mathbf{x}^\top_{it}\hat{\beta} + \hat{u}_i + \hat{u}_{2it}. \tag{2.12}$$

Rao and Yu (1994) used method of moment estimators to estimate the elements in δ. Extensions to this model have been made by Datta, Lahiri, and Maiti (2002) by replacing the AR(1) process with a random walk. A. C. Singh et al. (1991) use a random slope model instead of correlated random effects, but also use an AR(1) to describe the variation of the random regression coefficients.

2.4.2.3 *Spatio Temporal Extensions*

Subject of this section is the combination of the spatial model (2.9) and temporal model (2.11); this combination was first introduced by Marhuenda et al. (2013). A similar approach but for a multinomial response can be found in López-Vizcaíno et al. (2015). B. B. Singh et al. (2005) use a similar approach with respect to spatial autocorrelation but they use a State Space Model with a Kalman filter to take advantage of time series data.

Following Marhuenda et al. (2013) a spatio temporal FH model can be formulated as:

$$\tilde{y}_{it} = \mathbf{x}^\top_{it}\beta + u_{1i} + u_{2it} + e_{it} \tag{2.13}$$

where in contrast to model (2.11) the first random effect component has been replaced with a correlated random effects component following a SAR(1). To summarise this model, we have u_{1i} following a SAR(1), u_{2it} following an AR(1), and e_{it} are i.i.d., furthermore u_{1i}, u_{2it} and e_{it} are assumed to be independent. Following the presentation of the previous sections, the spatio temporal BLUP under model (2.13) can be defined by:

$$\tilde{\theta}^{STFH}_{it} = \mathbf{x}^\top_{it}\tilde{\beta} + \tilde{u}_{1i} + \tilde{u}_{2it}.$$

Marhuenda et al. (2013) used an REML estimator for the unknown variance components $\delta = (\rho_1, \sigma_1^2, \rho_2, \sigma_2^2)$. Replacing the unknown components with their respective estimates leads then to the EBLUP under model (2.13):

$$\hat{\theta}_{it}^{STFH} = \mathbf{x}_{it}^{\top}\hat{\beta} + \hat{u}_{1i} + \hat{u}_{2it}. \tag{2.14}$$

In contrast to B. B. Singh et al. (2005) who derived an analytical MSPE for spatio temporal domain predictions and Pratesi and Salvati (2008) who derived an MSPE for spatial predictions, Marhuenda et al. (2013) propose to use a parametric bootstrap for the estimation of the MSPE.

2.4.2.4 *Discussion*

The use of correlation across space to improve domain predictions is in principle promising for applications. However, an important remark and result of the reviewed literature is, that modelling spatial autocorrelation as a random effect is beneficial if this structure can not be captured by auxiliary information, i.e. in the fixed effects part of a mixed model. This means such models can be used to capture *unobserved* spatial correlation. Also note that although the proximity matrix is introduced to represent neighbouring units in a geographical sense it can refer to more abstract relationships. Neighbouring units may also be defined, for example, by capturing the structure between industry sectors with strong inter connections. Such structures, in contrast to geographical units, may be defined with domain expertise. Thus this strategy can be generally useful for domain predictions instead of being restricted to the geographical sense of area predictions.

Different is the use of information over time. Marhuenda et al. (2013) note that in practice we may be interested to make predictions for the current time period, and not the past, and use historic information as additional information. The use of historic information may lead to an improvement of parameter estimates due to an increased sample size. The temporal random effect in model (2.11) can additionally be of use to improve domain predictions. B. B. Singh et al. (2005) note that this is especially the case if the historic information, often itself predictions, is more reliable than the current time period.

The spatial, temporal and spatio-temporal FH is again the subject in section (?) where they are combined with robust estimation methodology. Note that the review in this chapter may not be sufficient to deduce the representation of these models as linear mixed models, which will be stated more precisely in section (?).

2.5 UNIT LEVEL MODELS

The basic unit level model, which can also be viewed as a linear mixed model, was introduced by Battese et al. (1988) and inherits the authors

names as it is refered to as the Battese-Harter-Fuller (BHF) model. Note that the main focus of this thesis is the extension of area level models. However, some recent advances in the field, especially with respect to robust estimation methodology, have been made for unit level models. In chapter (?) these results are extended to the area level; here the basic unit level model and unit level notation is reviewed in order to give some more details in the following section (?). A review of unit level models in the context of linear mixed models and EBLUP based predictions can be found in Jiang and Lahiri (2006) and also in Rao (2003) and Rao and Molina (2015) who also review a variety of extensions.

The unit level model of Battese et al. (1988) can be expressed as linear mixed model by:

$$y_{ij} = \mathbf{x}_{ij}^{\top}\beta + u_i + e_{ij} \tag{2.15}$$

where y_{ij} denotes the response of unit $j = 1, \dots, n_i$ in area $i = 1, \dots, D$. The random effetc, u_i, are independent and identically distributed following a normal distribution such that $u_i \sim \mathcal{N}(0, \sigma_u^2)$. The error term e_{ij} are i.i.d. following a normal distribution with zero mean and variance σ_e^2. Furthermore the random effect and model error are independent. $\mathbf{x}_{ij}$ are the corresponding auxiliary variables for that unit.

Let the set of n sampled and $N - n$ non-sampled units be denoted by s and r respectivey.

Let $\hat{\beta}$ denote the best linear unbiased estimator (BLUE) of β and $\hat{u}_i$ the best linear unbiased predictor (BLUP) (cf. Henderson, 1950 or Searle, 1971). The empirical best linear unbiased predictor (EBLUP) for the mean in small area i under the Battese-Harter-Fuller model is then given by

$$\hat{\theta}_i^{BHF} = N_i^{-1}\left(\sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} (x_i^{\top}\hat{\beta} + \hat{u_i})\right). \tag{2.16}$$

2.6 ROBUST METHODS IN SMALL AREA ESTIMATION

- Sinha and Rao (2008): Robust Small Area Estimation Under Unit Level Models
- Richardson and Welsh (1995): Robust Restricted Maximum Likelihood in Mixed Linear Models
- Yau and Kuk (2002): Robust Estimation in Generalized Linear Mixed Models
- Hulliger (2010): Simple and Robust Estimators for Sampling
- Lange, Little and Taylor (1989): Robust Statistical Modeling Using the t Distribution

- Peel and McLachlan (2000): Robust mixture modelling using the t distribution
- Bell and Huang (2006): Using the t-distribution to Deal with Outliers in Small Area Estimation
- Huang and Bell (): Using the t-distribution in Small Area Estimation: An Application to SAIPE State Poverty Models
- Schoch (2012): Robust Unit-Level Small Area Estimation: A Fast Algorithm for Large Datasets
- Xie, Raghunathan and Lepkowski (2007): Estimation of the proportion of overweight individuals in small areas - a robust extension of the Fay-Herriot model
- Gershunskaya (2010): Robust Small Area Estimation Using a Mixture Model
- Datta and Lahiri (1995): Robust Hierarchical Bayes Estimation of Small Area Characteristics in the Presence of Covariates and Outliers
- Rao, Sinha and Dumitrescu (2014): Robust Small area estimation under semi-parametric mixed models
- Salvati et.al. (2012): Small area estimation via M-quantile geographically weighted regression
- Salvati et.al. (2009): Spatial M-Quantile Modles for Small Area Estimation
- Shmid and Münnich (2012): Spatial robust small area estimation
- Tzavidis et.al. (2010): Robust Prediction of Small Area Means and Distributions
- Gervini / Yohai (2002): A Class of Robust and Fully Efficient Regression Estimators
- Pra99

2.6.1 *Robust ML Score Functions*

Fellner (1986) studied the robust estimation of linear mixed model parameters. However, the proposed approach is based on given variance parameters θ which is why Sinha and Rao (2009) propose an estimation procedure in which robust estimators for β and θ are solved iteratively. With given robust estimates for β and θ the estimation of the random effects is straight forward, the main concern, however, lies with the estimation of robust variance parameters. Starting from a mixed model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}u + \mathbf{e}$$

where $\mathbf{y}$ is the response vector with elements y_i, $\mathbf{X}$ the design matrix, u the vector of random effects and $\mathbf{e}$ the vector of sampling errors. Both error components are assumed to be normally distributed with $\mathbf{u} \sim \mathcal{N}(0, \mathbf{G})$ and $\mathbf{e} \sim \mathcal{N}(0, \mathbf{R})$ where $\mathbf{G}$ and $\mathbf{R}$ typically depend on some

variance parameters θ. Thus the variance of y is given by $\mathbf{V}(\mathbf{y}) = \mathbf{V}(\theta) = \mathbf{Z}\mathbf{G}\mathbf{Z}^\top + \mathbf{R}$. Maximizing the likelihood of $\mathbf{y}$ with respect to β and θ leads to the following equations:

$$\begin{aligned} \mathbf{X}^\top \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) &= 0 \\ (\mathbf{y} - \mathbf{X}\beta)^\top \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) - \operatorname{tr}\left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} \right) &= 0 \end{aligned}$$

where q denotes the number of unknown variance parameters with $l = 1, \ldots, q$. Solving the above equations leads to the ML-estimates for β and θ. To robustify against outliers in the response variable, the residuals $(\mathbf{y} - \mathbf{X}\beta)$ are standardized and restricted by some influence function $\psi(\cdot)$. The standardized residuals are given by

$$\mathbf{r} = \mathbf{U}^{-\frac{1}{2}} (\mathbf{y} - \mathbf{X}\beta)$$

where $\mathbf{U}$ is the matrix of diagonal elements of $\mathbf{V}$ and thus also depends on the variance parameters θ. A typical choice for $\psi(\cdot)$ is Hubers influence function:

$$\psi(u) = u \min\left(1, \frac{b}{|u|}\right).$$

A typical choice for b is 1.345. The vector of robustified residuals is denoted by $\_(\mathbf{r}) = (\psi(r_1), \ldots, \psi(r_n))$. Solving the following robust ML-equations leads to robustified estimators for β and θ:

$$\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{U}^{\frac{1}{2}} \psi(\mathbf{r}) = 0 \tag{2.17}$$

$$\Phi_l(\theta) = \psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} \mathbf{V}^{-1} \mathbf{U}^{\frac{1}{2}} \psi(\mathbf{r}) - \operatorname{tr}\left(K \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} \right) = 0 \tag{2.18}$$

where K is a diagonal matrix of the same order as $\mathbf{V}$ with diagonal elements $c = \mathbb{E}[\psi^2(r)|b]$ where r follows a standard normal distribution.

2.6.2 *Bias Correction*

- Tzavidis and Chambers (2005): BIAS ADJUSTED SMALL AREA ESTIMATION WITH M-QUANTILE MODELS
- Jiongo, Haziza and Duchesne (2013): Controlling the bias of robust small-area estimators

2.6.3 *Mean Squared Error Estimation*

2.6.3.1 *Bootstrap*

- Jiiongo, Nguimekeu (2014): Bootstrapping Mean Squared Errors of Robust Small Area Estimators

2.6.3.2 *Pseudo Linearization*

Chambers, Chandra, and Tzavidis (2011) and Chambers, Chandra, Salvati, et al. (2014) deal with the estimation of the MSE of robust area predictors in the context of Small Area Estimation. In this section I review their results. Later in section ? I will, firt, adapt their findings to estimate the MSE of the robustified Fay Herriot model, and second use the linearization of robust mixed models to derive a fixed point algorithm to find solutions for the model parameters.

The central idea is to formulate the RBLUP as wigthed sum of the response vector:

$$\theta_i^{\mathrm{RBLUP}} = \sum_{j \in s} w_{ij}^{\mathrm{RBLUP}} y_{ij} = \left(\mathbf{w}_{is}^{\mathrm{RBLUP}}\right)^\top \mathbf{y}_s$$

where

$$\left(\mathbf{w}_{is}^{\mathrm{RBLUP}}\right)^\top = N_i^{-1}\left(\mathbf{1}_s^\top + (N_i - n_i)\left(\bar{x}_{ir}^\top \mathbf{A}_s + \bar{z}_{ir}^\top \mathbf{B}_s\left(\mathbf{I}_s - \mathbf{X}_s \mathbf{A}_s\right)\right)\right)$$

and

$$\mathbf{A}_s = \left(\mathbf{X}_s^\top \mathbf{V}_s^{-1} \mathbf{U}_s^{\frac{1}{2}} \mathbf{W}_{1s} \mathbf{U}_s^{-\frac{1}{2}} \mathbf{X}_s\right)^{-1} \mathbf{X}_s^\top \mathbf{V}_s^{-1} \mathbf{U}_s^{\frac{1}{2}} \mathbf{W}_{1s} \mathbf{U}_s^{-\frac{1}{2}}$$

with

$$\mathbf{W}_{1s} = \mathrm{diag}\left(w_j\right)_{n \times n}$$

and

$$w_{1j} = \frac{\psi\left(U_j^{-\frac{1}{2}}\left(y_j - x_j^\top \hat{\beta}^\psi\right)\right)}{U_j^{-\frac{1}{2}}\left(y_j - x_j^\top \hat{\beta}^\psi\right)}$$

$$\mathbf{B}_s = \left(\mathbf{Z}_s^\top \mathbf{V}_{es}^{-\frac{1}{2}} \mathbf{W}_{2s} \mathbf{V}_{es}^{-\frac{1}{2}} \mathbf{Z}_s + \mathbf{V}_u^{-\frac{1}{2}} \mathbf{W}_{3s} \mathbf{V}_u^{-\frac{1}{2}}\right)^{-1} \mathbf{Z}_s^\top \mathbf{V}_e^{-\frac{1}{2}} \mathbf{W}_{2s} \mathbf{V}_e^{-\frac{1}{2}}$$

with $\mathbf{W}_{2s}$ as diagonal matrix with ith component:

$$w_{2i} = \frac{\psi\left(\left(\sigma^{\psi}_{e,i}\right)^{-1}\left(y_i - x_i^{\top}\hat{\beta}^{\psi} - \hat{u}_i^{\psi}\right)\right)}{\left(\sigma^{\psi}_{e,i}\right)^{-1}\left(y_i - x_i^{\top}\hat{\beta}^{\psi} - \hat{u}_i^{\psi}\right)}$$

and with $\mathbf{W}_{3s}$ as $(m \times m)$ diagonal matrix with ith component:

$$w_{3i} = \frac{\psi\left(\left(\sigma^{\psi}_{u}\right)^{-1}\hat{u}_i^{\psi}\right)}{\left(\sigma^{\psi}_{u}\right)^{-1}\hat{u}_i^{\psi}}$$

This all assumes known variance parameters. When the variance parameters are unknown, they are estimated and instead of $\mathbf{w}_{is}^{\text{RBLUP}}$ we have to use $\mathbf{w}_{is}^{\text{REBLUP}}$. Then the estimator of the conditional MSE is given by:

$$\widehat{\text{MSE}}\left(\hat{\theta}_i^{\text{REBLUP}}\right) = \hat{\mathbb{V}}\left(\hat{\theta}_i^{\text{REBLUP}}\right) + \hat{\mathbb{B}}\left(\hat{\theta}_i^{\text{REBLUP}}\right)^2$$

$$\hat{\mathbb{V}}\left(\hat{\theta}_i^{\text{REBLUP}}\right) = N_i^{-2}\sum_{j\in s}\left(a_{ij}^2 + (N_i - n_i)\,n^{-1}\right)\lambda_j^{-1}\left(y_j - \hat{\mu}_j\right)^2$$

with

$$a_{ij} = N_i w_{ij}^{\text{REBLUP}} - I\left(j \in i\right)$$

and

$$\hat{\mathbb{B}}\left(\hat{\theta}_i^{\text{REBLUP}}\right) = \sum_{j\in s} w_{ij}^{\text{REBLUP}}\hat{\mu}_j - N_i^{-1}\sum_{j\in(r_i\cup s_i)}\hat{\mu}_j$$

Note that $\hat{\mu}_j$ is an unbiased estimator of the the conditional expectation $\mu_j = \mathbb{E}\left(y_j|\mathbf{x}_j, \mathbf{u}^{\psi}\right)$. $\lambda_j = 1 - 2\phi_{jj} + \sum_{k\in s}\phi_{kj}^2$.

3

ROBUST AREA LEVEL MODELS

3.1 THE PROBLEM OF OUTLYING OBSERVATIONS FOR AREA-LEVEL MODELS

> *"[Outliers:] observations obtained under seemingly normal circumstances, but that turn out to be extremely deviant from the main body of observations."* – Abelson (1995, p. 69)

- Hulliger (2010): *Simple and Robust Estimators for Sampling*, SFSO
  - Robuste direkte Schätzer mit Survey Gewichten (robuster HT)
- 2015-03-17: Bell, W. t-Distr. US-Census
  - Paper: *Using the t-distribution in Small Area Estimation: An Application to SAIPE State Poverty Models*
  - Paper: *Using the t-distribution to Deal with Outliers in Small Area Estimation*
- 2015-02-17: `Literatur` – Handbook of Statistics 29a – Dealing with Outliers in Survey Data, 247 ff.
  - Winzorization des HT-Schätzers
  - Buaumont
- Chakraborty, Datta and Mandal (2015): A two-component normal mixture alternative to the Fay-Herriot model
- Fabrizi, Trivisano (2010): Robust linear mixed modles for Small Area Estimation

This section provides some motivation for the study of robust area-level models. It mainly lays out under which scenarios a robust estimation may proove to be beneficial and how these points are addressed in later chapters. Most importantly you can find insights on what outlying observations are from an area-level perspective. And if such data points exist, what is the source of this abnormal behaviour?

In the following I want to distinguish between three types of outlying observations. Unit-level outliers and how they may effect the area-level analysis are discussed in 3.1.1. Area-level outliers can be described as outlying domains, which means that an entire domain - or all units in that domain - behaves differently than all others and are further discussed in section 3.1.2. A third kind can be best described as overly influential observations. The effect of such observations on the area-level is more subtle. A first intuition and possible sources of this type is given in section 3.1.3.

3.1.1 *Unit-Level Outliers*

Unit-level outliers can be representative or non-representative values but in most cases it is beyond our reach to judge which kind they are because we only see aggregates. They will influence the direct estimator in that this quntity may have unexpected high or low values. But they also influence the estimation of the standard error of the direct estimator which in turn can be used as the true or rather given variance parameters in a Fay-Herriot type model.

In the situation where the direct variance estimators are used, two conflicting effects need to be considered. First, we use an obviously unreliable estimator for the sampling error and assume that such values can be used as the *true* variance parameters in the Fay-Herriot model. Such estimators will have poor properties when the target is the *true* sampling error especially in the context of outliers. This problem stimulates the recent discussion around smoothing the variance estimates prior to using them in a Fay-Herriot model (see ? for reference of this discussion and section 3.1.3 for how the robust estimator relates to this discussion).

Second, we may consider the direct variance estimators not as an informative quantity for the *true* sampling error, but as informative about the unit-level sample. This is an important aspect because the variance becomes a reliablility index. The estimated sampling variance is large if outliers are present or in general the sample is heterogenous; and it is low for areas in which we have a reliable direct estimator. In this case the direct variance estimation will weight down unreliable direct estimators, which is why the Fay-Herriot model can be considered to be robust against unit-level outliers.

When we consider unit-level outliers it is relevant to ask why we would not use a robust direct estimator. From a practical point of view we may only have access to non-robust estimators. On the other hand it may be crucial feedback to a data provider to instead report robust estimates, e.g. a robust mean or median. Also it is unclear how this relates to the *self-adjusting* effect of the Fay-Herriot model. This aspect is addressed by model-based simulations in section (?).

3.1.2 *Area-Level Outliers*

Area-level outliers are domains or areas which are far away from the bulk of the observations. The source of such observations are not single units influencing the direct estimator, but the fact that there exist domains which are substantially different from the majority.

If we take the standpoint of a mixed-model perspective this means that there are outliers in the random effect. From a model perspective where we treat the random effect as i.i.d. random variable following a normal distribution such values may arise by chance alone. This

would mean that in truth the distribution is indeed normal but we are in one of those rare cases in which we observe abnormal behaviour of single domains.

This argumentation seems to be unrealistic because under repeated sampling we would expect the same domains to be abnormal which stimulates a more fundamental discussion of mixed-models in Small Area Estimation (see ?); and it gives reason to frame a random effect as fixed but unknown quntity which is approximated by a normal distribution. In this setting the existance of outliers means that the normal distribution may well be a good approximation for the majority of areas but not for all them. This will raise the question if a different distribution should be used to approximate the random effect, e.g. a mixture distribution of normals or a scewed distribution.

Regardless of the perspective, in the application in which such values are observed a robust estimation technique can be beneficial. The main effect of a misspecified distributional assumption is that the variance parameter of the random effect will be overestimated - at least this is what we should expect in the presence of outliers. Also the parameter estimates of the fixed effects part can be influenced by such observations, allthough it is unclear how they are unfluenced specifically (under- or over-estimated).

Looking at the shrinkage estimator it can be derived how an overestimation of the estimated variance parameter of the random effect influences area-level predictions. The larger the estimated variance compared to the area-specific sampling variance the stronger the prediction relies on the direct estimator. However, the robust prediction will turn out not to be the shrinkage estimator but instead a robustified best linear unbiased predictor for which the influence is more subtle. The intuition of the effect, however, remains the same.

3.1.3 *The Creation of Overly Influencial Observations*

In this section I want to introduce a third kind of outlying observation. Such values are better understood as overly influencial observations as they do not necessarily are far away from the center of the data. Furthermore they can not be understood by looking at the direct estimator alone but must be seen as pairs of direct estimator and sampling variance. Together they determine the impact of a single observation on the overall predictions.

To distinguish them from area- and unit-level outliers consider two hypothetical scenarios from which these values can arise:

- There are no outliers present and the unit-level population model is that of (?). However, the sample sizes for most areas are very small - say between 5 and 50 - and only for very few domains we have sufficient sample size of say 500. This will inevitably result in a very heterogenous picture on the area-level where

domains with sample size 500 will appear to be extremely reliable. The effect can be that such observations will dominate the global mean and thus effectively become the value we are shrinking against.
- We have unit-level outliers influencing domain specific direct estimators. However, instead of using the estimated sampling variance we smooth the variances against a global parameter - essentially shrinking both means and variances as suggested in (?). This results in outlying domains for which, after smoothing, the direct estimators appear to be more reliable than they truly are. Effectively this reduces the self adjusting effect described in section 3.1.1.

Especially the second scenario should not be misunderstood as an argument against smoothing, or that we should not care about unstable variance estimation. Allthogh the variances are not directly smoothed the influence of observations is bounded in a robust estimation procedure. In fact the introduced robust area-level model can be considered to have a positive effect with respect to both scenarios because single observations are restricted in their influence. Furthermore the robust model may present an alternative to smoothing strategies, although that has not been the initial intention, without the danger of creating overly influencial observations.

3.2 THE ROBUST FAY-HERRIOT MODEL AND EXTENSIONS

- Chakraborty, Datta and Mandal (2015): A two-component normal mixture alternative to the Fay-Herriot model

The variance structure under model (2.11) can again be derived as a special mixed linear model. Hence the concrete variance components are stated with respect to model (2.1). Here $V_e = \text{diag}\left(\{\{\sigma^2_{eit}\}_{t=1}^{T}\}_{i=1}^{D}\right)$ and is similar to the structure under model (2.7). $V_u(\delta) = \text{diag}\left(\sigma^2_u \mathbf{I}_D, \sigma^2_2 \Omega(\rho_2)\right)$ which depends on the vector of unknown variance parameters $\delta = (\sigma^2_1, \rho_2, \sigma^2_2)$. $\Omega_2(\rho_2)$ has a block diagonal structure with $\Omega_{2i}(\rho_2)$ denoting the block for observation i with:

$$\Omega_{2i}(\rho_2) = \frac{1}{1-\rho_2^2} \begin{pmatrix} 1 & \rho_2 & \cdots & \rho_2^{T-2} & \rho_2^{T-1} \\ \rho_2 & 1 & & & \rho_2^{T-2} \\ \vdots & & \ddots & & \vdots \\ \rho_2^{T-2} & & & 1 & \rho_2 \\ \rho_2^{T-1} & \rho_2^{T-2} & \cdots & \rho_2 & 1 \end{pmatrix}_{T \times T}$$

Furthermore $\mathbf{uZ} = (\mathbf{I}_D, \mathbf{Z}_2)_{DT \times (D+DT)}$

$$\tilde{y}_{it} = \theta_{it} + e_{it}$$

The model error is composed of a spatial autoregressive process of order 1 (SAR(1)) and an autoregressive process of order 1 (AR(1)):

$$\theta_{it} = \mathbf{x}_{it}^\top \beta + u_{1i} + u_{2it}$$

where u_{1i} and u_{2it} follow a SAR(1) and AR(1) respectively:

$$u_{1i} = \rho_1 \sum_{l \neq i} w_{il} u_{1l} + \epsilon_{1i}$$

where $|\rho_1| < 1$ and $\epsilon_{1i} \sim \mathcal{N}(0, \sigma_1^2)$ are i.i.d. with $i = 1, \ldots, D$. w_{il} are the elements of $\mathbf{W}$ which is the row standardized proximity matrix $\mathbf{W}^0$. The elements in $\mathbf{W}^0$ are equal to 1 if areas are neighboured and 0 otherwise, thus the dimension of $\mathbf{W}^0$ is $D \times D$. As stated above u_{2it} follows an AR(1): Note that u_{1i} and u_{2it} and e_{it} are independent and the sampling error variance parameters are assumed to be known. The model can then be stated as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where $\mathbf{y}$ is the $DT \times 1$ vector containing y_{it} as elements, $\mathbf{X}$ is the $DT \times P$ design matrix containing the vectors $\mathbf{x}_{it}^\top$ as rows, $\mathbf{u}$ is the $(D + DT) \times 1$ vector of model errors and $\mathbf{e}$ the $DT \times 1$ vector of sampling errors e_{it}. Note that $\mathbf{u} = (\mathbf{u}_1^\top, \mathbf{u}_2^\top)$ where the $D \times 1$ vector $\mathbf{u}_1$ and $DT \times 1$ vector $\mathbf{u}_2$ have u_{1i} and u_{2it} as elements respectively. Furthermore $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$ has dimension $DT \times (D + DT)$, where $\mathbf{Z}_1 = \mathbf{I}_D \otimes \mathbf{1}_T$ ($\mathbf{I}_D$ denotes a $D \times D$ identity matrix and $\mathbf{1}_T$ a $1 \times T$ vector of ones) has dimension $DT \times D$ and $\mathbf{Z}_2$ is a $DT \times DT$ identity matrix.

Concerning the variance of $\mathbf{y}$ first consider the distributions of all error components. $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_e)$ where $\mathbf{V}_e$ is a diagonal matrix with the known σ_{eit}^2 on the main diagonal. $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_u(\delta))$ with the block diagonal covariance matrix $\mathbf{V}_u(\delta) = \operatorname{diag}(\sigma_1^2 \Omega_1(\rho_1), \sigma_2^2 \Omega_2(\rho_2))$ where $\delta = (\sigma_1^2, \rho_1, \sigma_2^2, \rho_2)$.

$$\Omega_1(\rho_1) = \left((\mathbf{I}_D - \rho_1 \mathbf{W})^\top (\mathbf{I}_D - \rho_1 \mathbf{W}) \right)^{-1}$$

and follows from the SAR(1) process in the model errors.

The variance of $\mathbf{y}$ can thus be stated as:

$$\mathbb{V}(\mathbf{y}) = \mathbf{V}(\delta) = \mathbf{Z}\mathbf{V}_u(\delta)\mathbf{Z}^\top + \mathbf{V}_e$$

The BLUE of β and BLUP of δ can be stated as (see Henderson, 1975):

$$\tilde{\beta}(\delta) = \left(\mathbf{X}^\top \mathbf{V}^{-1}(\delta) \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{V}^{-1}(\delta) \mathbf{y}$$

$$\tilde{\mathbf{u}}(\delta) = \mathbf{V}_u(\delta) \mathbf{Z}^\top \mathbf{V}^{-1}(\delta) \left(\mathbf{y} - \mathbf{X}\tilde{\beta}(\delta) \right)$$

Hence the BLUP of $\mathbf{u}_1$ and $\mathbf{u}_2$ can be stated as:

$$\tilde{\mathbf{u}}_1(\delta) = \sigma_1^2 \Omega_1(\rho_1) \mathbf{Z}^\top \mathbf{V}^{-1}(\delta) \left(\mathbf{y} - \mathbf{X}\tilde{\beta}(\delta) \right)$$

$$\tilde{\mathbf{u}}_2(\delta) = \sigma_2^2 \Omega_2(\rho_2) \mathbf{Z}^\top \mathbf{V}^{-1}(\delta) \left(\mathbf{y} - \mathbf{X}\tilde{\beta}(\delta)\right)$$

Estimating δ leads to the EBLUE for β and EBLUPs for $\mathbf{u}_1$ and $\mathbf{u}_2$, hence an predictor for the area characteristic θ_{it} is given by:

$$\hat{\theta}_{it}^{STFH} = \mathbf{x}_{it}^\top \hat{\beta} + \hat{\mathbf{u}}_{1i} + \hat{\mathbf{u}}_{2it}$$

Marhuenda et al. (2013) use a restricted maximum likelihood method to estimate δ independently of β. An open question is if this approach can be applied for the robust spatio-temporal model. Thus we will continue with the discussion of robust small area methods.

3.3 BIAS CORRECTION

The two existing methodologies which are available to do bias correction for robust models in SAE are not easily adapted to the general context of the discussed area level models. Both approaches are based on a robust scale estimate of the residuals within in a domain. In the standard RFH and SRFH there is only one observation per domain, such that these estimates can not be directly transfered.

However, the general problem remains. In the setting of the robustified score functions and a misspecified distribution in the sense that outlying observations have a different mean, the model can introduce a severe bias to those outlying estimations.

Independet of this question Fay and Herriot (1979) refer to Efron and Morris (1971) and Efron and Morris (1972) who argue that the Bayes Estimator and and the Emperical Bayes Estimator may improve the overall prediction performance but can be ill suited for specific domains. In essence this describes also the situation with the RFH models which introduce a bias to the prediction of outlying domains.

Efron and Morris also suggest a simple correction to the prediction which can be directly applied to the robust models under study. They suggest to restrict the prediction by an interval around the direct estimator. The width of this intervall can be constructed as a multiple of the known standard errors under the model:

$$\theta_i^{BC} = \begin{cases} \tilde{y}_i - c & \text{if } \theta_i^{RFH} < \tilde{y}_i - c \\ \theta_i^{RFH} & \text{if } \tilde{y}_i - c \leqslant \theta_i^{RFH} \leqslant \tilde{y}_i + c \\ \tilde{y}_i + c & \text{if } \theta_i^{RFH} > \tilde{y}_i + c \end{cases}$$

3.4 MEAN SQUARED ERROR ESTIMATION

3.4.1 *Bootstrap*

3.4.2 *Pseudo Linearization*

This is the representation of the pseudo linear representation of the FH model. As it is introduced in Chambers, Chandra, and Tzavidis (2011) and Chambers, Chandra, Salvati, et al. (2014).

Presenting the FH in pseudo linear form means to present the area means as a weighted sum of the response vector y. The FH model is given by

$$\theta_i = \gamma_i y_i + (1-\gamma_i) x_i^\top \beta \tag{3.1}$$

where $\gamma_i = \frac{\sigma_u^2}{\sigma_u^2+\sigma_e^2}$, so it can be represented as

$$\theta_i = w_i^\top y$$

where

$$w_i^\top = \gamma_i \mathbf{I}_i^\top + (1-\gamma_i) x_i^\top \mathbf{A}$$

and

$$\mathbf{A} = \left(\mathbf{X}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\mathbf{W}\mathbf{U}^{-\frac{1}{2}}\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\mathbf{W}\mathbf{U}^{-\frac{1}{2}}$$

with

$$\mathbf{W} = \mathrm{Diag}(w_j), \text{ with } j = 1,\dots,n$$

and

$$w_j = \frac{\psi\left(u_j^{-\frac{1}{2}}(y_j - x_j^\top\beta)\right)}{u_j^{-\frac{1}{2}}(y_j - x_j^\top\beta}$$

Note that if ψ is the identity or equally a huber influence function with a large smoothing constant, i. e.inf:

$$\mathbf{A} = \left(\mathbf{X}\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{V}^{-1}$$

The fixed point function derived from these formulas are the following:

$$\beta = \mathbf{A}(\beta)y$$

This whole thing can also be addapted to define the random effects. If we define the model in an alternative way:

$$\theta_i = x_i^\top \beta + u_i \tag{3.2}$$

we can restate it similarly to the above as:

$$\theta_i = w_{s,i}^\top \mathbf{y}$$

$$\theta = \mathbf{W}\mathbf{y}$$

where $\mathbf{W}$ is the matrix containing the weights, i.e.

$$\mathbf{W} = \begin{pmatrix} w_{s,1}^\top \\ \vdots \\ w_{s,D}^\top \end{pmatrix} = \mathbf{X}\mathbf{A} + \mathbf{B}\left(\mathbf{I} - \mathbf{X}\mathbf{A}\right)$$

with

$$w_{s,i}^\top = x_i^\top \mathbf{A} + z_i^\top \mathbf{B}\left(\mathbf{I} - x_i^\top \mathbf{A}\right)$$

where $\mathbf{A}$ is defined as above and

$$\mathbf{B} = \left(\mathbf{V}_e^{-\frac{1}{2}}\mathbf{W}_2\mathbf{V}_e^{-\frac{1}{2}} + \mathbf{V}_u^{-\frac{1}{2}}\mathbf{W}_3\mathbf{V}_u^{-\frac{1}{2}}\right)^{-1}\mathbf{V}_e^{-\frac{1}{2}}\mathbf{W}_2\mathbf{V}_e^{-\frac{1}{2}}$$

with $\mathbf{W}_2$ as diagonal matrix with ith component:

$$w_{2i} = \frac{\psi\{\sigma_{e,i}^{-1}(y_i - x_i^\top \beta - u_i)\}}{\sigma_{e,i}^{-1}(y_i - x_i^\top \beta - u_i)}$$

and with $\mathbf{W}_3$ as diagonal matrix with ith component:

$$w_{3i} = \frac{\psi\{\sigma_u^{-1}u_i\}}{\sigma_u^{-1}u_i}$$

The fixed point function derived from these formulas are the following:

$$\begin{aligned} u &= \mathbf{B}(u)\,(\mathbf{I} - \mathbf{X}\mathbf{A})\,y \\ &= \mathbf{B}(u)\,(y - \mathbf{X}\beta) \end{aligned}$$

Given the weights we have a weighted mean, for which we need the MSE:

$$\mathbb{MSE}\left(\hat{\theta}|\mathbf{X}, \beta, \mathbf{u}\right) = \mathbb{E}\left(\left(\hat{\theta} - \theta\right)^2\right) = \mathbb{V}\left(\hat{\theta}\right) + \mathbb{E}\left(\hat{\theta} - \theta\right)^2$$

$$\mathbb{V}\left(\hat{\theta}\right) = \mathbb{V}\left(\mathbf{W}y\right) = \mathbf{W}^2\mathbb{V}\left(\mathbf{y}\right) = \mathbf{W}^2\left(\sigma^2_{e,1}, \ldots, \sigma^2_{e,D}\right)^\top$$

$$\mathbb{E}\left(\hat{\theta} - \theta\right) = \mathbb{E}\left(\mathbf{W}y\right) - \mathbb{E}\left(\theta\right) = \mathbf{W}\mathbb{E}\left(\mathbf{y}\right) - \mathbb{E}\left(\theta\right) = \mathbf{W}\theta - \theta$$

3.5 ALGORITHM

3.5.1 *Newton Raphson Algorithms*

Sinha and Rao (2009) propose a Newton-Raphson algorithm to solve equations 2.17 and 2.18 iteratively. The iterative equation for β is given by:

$$\beta^{(m+1)} = \beta^{(m)} + \left(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{D}(\beta^{(m)})\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r}(\beta^{(m)}))$$

where $\mathbf{D}(\beta) = \frac{\partial\psi(\mathbf{r})}{\partial\mathbf{r}}$ is a diagonal matrix of the same order as $\mathbf{V}$ with elements

$$D_{jj} = \begin{cases} 1 \text{ for } |r_j| \leqslant b \\ 0 \text{ else} \end{cases}, j = 1, \ldots, n$$

The iterative equation for θ can be stated as:

$$\theta^{(m+1)} = \theta^{(m)} - \left(\Phi'(\theta^{(m)})\right)^{-1}\Phi(\theta^{(m)})$$

where $\Phi'(\theta^m)$ is the derivative of $\Phi(\theta)$ evaluated at $\theta^{(m)}$. The derivative of Φ is given by Schmid, 2012, p.53:

$$\frac{\partial\Phi}{\partial\theta_l} = 2\frac{\partial}{\partial\theta_l}\left(\psi(\mathbf{r})^\top\mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\right)\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r}) + \operatorname{tr}\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}K\right) \tag{3.3}$$

where

$$\frac{\partial}{\partial\theta_l}\left(\psi(\mathbf{r})^\top\mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\right) = \frac{\partial}{\partial\theta_l}(\psi(\mathbf{r})^\top)\mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1} + \psi(\mathbf{r})^\top\frac{\partial}{\partial\theta_l}(\mathbf{U}^{\frac{1}{2}})\mathbf{V}^{-1} - \psi(\mathbf{r})^\top\mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{V}^{-1}.$$

In Schmid (2012) adopted this procedure for the Spatial Robust EBLUP and essentially we will follow the same procedure Schmid (2012, p.74ff.). Thus we will directly consider the algorithm for the Spatio Temporal model introduced earlier. Since the model considered by Sinha and Rao (2009) contained a block diagonal variance structure where all off-diagonals are zero, equation 3.3 is valid with respect to the earlier specified variance parameters σ_1^2 and σ_2^2 from the spatio temporal Fay Herriot model. The derivative of Φ with respect to ρ_1 and ρ_2, however, is different. To adapt the notation, let $\theta = (\sigma_1^2, \sigma_2^2)$ for which equation 3.3 holds. Let $\rho = (\rho_1, \rho_2)$ denote the vector of correlation parameters as they already have been defined above. Then the iterative equation for ρ is can be stated as:

$$\rho^{(m+1)} = \rho^{(m)} + \left(\Phi'(\rho^{(m)})\right)^{-1} \Phi(\rho^{(m)})$$

where the derivative of Φ with respect to ρ is given by Schmid (2012, p.76):

$$\begin{aligned}\frac{\partial \Phi}{\partial \rho_l} =& 2\frac{\partial}{\partial \rho_l}\left(\psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\right)\frac{\partial \mathbf{V}}{\partial \rho_l}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r}) \\ &+ \psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \rho_l \partial \rho_l}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r}) \\ &+ \operatorname{tr}\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \rho_l \partial \rho_l}\mathrm{K} - \mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_l}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_l}\mathrm{K}\right)\end{aligned}$$

The partial derivatives of $\mathbf{V}$ with respect to θ and ρ are given by:

$$\begin{aligned}\frac{\partial \mathbf{V}}{\partial \sigma_1^2} =& \mathbf{Z}_1 \Omega_1(\rho_1)\mathbf{Z}_1^\top \\ \frac{\partial \mathbf{V}}{\partial \sigma_2^2} =& \Omega_2(\rho_2) \\ \frac{\partial \mathbf{V}}{\partial \rho_1} =& -\sigma_1^2 \mathbf{Z}_1 \Omega_1(\rho_1)\frac{\partial \Omega_1^{-1}(\rho_1)}{\partial \rho_1}\Omega_1(\rho_1)\mathbf{Z}_1^\top \\ \frac{\partial \mathbf{V}}{\partial \rho_2} =& \sigma_2^2 \operatorname{diag}\left(\frac{\partial \Omega_{2d}(\rho_2)}{\partial \rho_2}\right) \\ \frac{\partial \mathbf{V}}{\partial \rho_1 \partial \rho_1} =& -\sigma_1^2 \mathbf{Z}_1 \frac{\partial \Omega_1(\rho_1)}{\partial \rho_1}\frac{\partial \Omega_1^{-1}(\rho_1)}{\partial \rho_1}\Omega_1(\rho_1)\mathbf{Z}_1^\top \\ &- \sigma_1^2 \mathbf{Z}_1 \Omega_1(\rho_1)\frac{\partial \Omega_1^{-1}(\rho_1)}{\partial \rho_1 \partial \rho_1}\Omega_1(\rho_1)\mathbf{Z}_1^\top \\ &- \sigma_1^2 \mathbf{Z}_1 \Omega_1(\rho_1)\frac{\partial \Omega_1^{-1}(\rho_1)}{\partial \rho_1}\frac{\partial \Omega_1(\rho_1)}{\partial \rho_1}\mathbf{Z}_1^\top \\ \frac{\partial \mathbf{V}}{\partial \rho_2 \partial \rho_2} =& \text{Needs to be TEXed}\end{aligned}$$

where

$$
\begin{aligned}
\frac{\Omega_1(\rho_1)}{\partial\rho_1} &= -\Omega_1(\rho_1)\frac{\partial\Omega_1^{-1}(\rho_1)}{\partial\rho_1}\Omega_1(\rho_1)\,, \\
\frac{\partial\Omega_1^{-1}(\rho_1)}{\partial\rho_1} &= -\mathbf{W}-\mathbf{W}^\top+2\rho_1\mathbf{W}^\top\mathbf{W}\,, \\
\frac{\partial\Omega_1^{-1}(\rho_1)}{\partial\rho_1\partial\rho_1} &= 2\mathbf{W}^\top\mathbf{W} \\
\frac{\partial\Omega_{2d}(\rho_2)}{\partial\rho_2} &= \frac{1}{1-\rho_2^2}\begin{pmatrix} 0 & 1 & \cdots & \cdots & (T-1)\rho_2^{T-2} \\ 1 & 0 & & & (T-2)\rho_2^{T-3} \\ \vdots & & \ddots & & \vdots \\ (T-2)\rho_2^{T-3} & & & 0 & 1 \\ (T-1)\rho_2^{T-2} & \cdots & \cdots & 1 & 0 \end{pmatrix} + \frac{2\rho_2\Omega_{2d}(\rho_2)}{1-\rho_2^2}
\end{aligned}
$$

Having identified all iterative equations the adapted algorithm from Schmid (2012) is as follows:

- Choose initial values for β^0, θ^0 and ρ^0.
- Compute $\beta^{(m+1)}$, with given variance parameters and correlation parameters
  - Compute $\theta^{(m+1)}$, with given regression and correlation parameters
  - Compute $\rho^{(m+1)}$, with given variance and regression parameters
- Continue step 2 until the following stopping rule holds:

$$
\begin{aligned}
||\beta^{(m+1)}-\beta^{(m)}||^2 &< \text{const} \\
(\sigma_1^{2(m+1)}-\sigma_1^{2(m)})^2+(\sigma_2^{2(m+1)}-\sigma_2^{2(m)})^2+(\rho_1^{(m+1)}-\rho_1^{(m)})^2+(\rho_2^{(m+1)}-\rho_2^{(m)})^2 &< \text{const}
\end{aligned}
$$

3.5.2 *Fixed Point Algorithms*

Inspired by: Chatrchi Golshid (2012): Robust Estimation of Variance Components in Small Area Estimation, Master-Thesis, Ottawa, Ontario, Canada: p. 16ff.:

> The fixed-point iterative method relies on the fixed-point theorem: "If g(x) is a continuous function for all $x \in [a;b]$, then g has a fixed point in $[a;b]$." This can be proven by assuming that $g(a) \geqslant a$ and $g(b) \leqslant b$. Since g is continuous the intermediate value theorem guarantees that there exists a c such that $g(c) = c$.

Starting from equation 2.18 where $\theta = (\sigma_1^2, \sigma_2^2)$ and (ρ_1, ρ_2) are assumed to be known, we can rewrite the equation such that:

$$\Phi_l(\theta) = \psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} \mathbf{V}^{-1} \mathbf{U}^{\frac{1}{2}} \psi(\mathbf{r}) - \operatorname{tr}\left(K\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} (\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top)^{-1} (\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top) \right) = 0 \quad (3.4)$$

Note that because the matrix $\mathbf{V}_e$ is assumed to be known for the FH model, it can be omitted. Let $\mathbf{o}_{r\times c}$ define a matrix filled with 0 of dimension $(r \times c)$ then:

$$\begin{aligned}
\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top =& \mathbf{Z} \begin{pmatrix} \sigma_1^2\Omega_1 & \mathbf{o}_{D\times DT} \\ \mathbf{o}_{DT\times D} & \sigma_2^2\Omega_2 \end{pmatrix} \mathbf{Z}^\top \\
=& \mathbf{Z} \left[\sigma_1^2 \begin{pmatrix} \Omega_1 & \mathbf{o}_{D\times DT} \\ \mathbf{o}_{DT\times D} & \mathbf{o}_{DT\times DT} \end{pmatrix} + \sigma_2^2 \begin{pmatrix} \mathbf{o}_{D\times D} & \mathbf{o}_{D\times DT} \\ \mathbf{o}_{DT\times D} & \Omega_2 \end{pmatrix} \right] \mathbf{Z}^\top \\
=& \begin{pmatrix} \mathbf{Z}\bar{\Omega}_1\mathbf{Z}^\top & \mathbf{Z}\bar{\Omega}_2\mathbf{Z}^\top \end{pmatrix} \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \end{pmatrix}
\end{aligned}$$

Thus equation 3.7 can be rewritten to:

$$\psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} \mathbf{V}^{-1} \mathbf{U}^{\frac{1}{2}} \psi(\mathbf{r}) = \operatorname{tr}\left(K\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} (\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top)^{-1} \begin{pmatrix} \mathbf{Z}\bar{\Omega}_1\mathbf{Z}^\top & \mathbf{Z}\bar{\Omega}_2\mathbf{Z}^\top \end{pmatrix} \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \end{pmatrix} \right)$$

Let

$$\begin{pmatrix} \psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_1^2} \mathbf{V}^{-1} \mathbf{U}^{\frac{1}{2}} \psi(\mathbf{r}) \\ \psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_2^2} \mathbf{V}^{-1} \mathbf{U}^{\frac{1}{2}} \psi(\mathbf{r}) \end{pmatrix} = a(\theta) \,,$$

then

$$\theta = \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \end{pmatrix} = A(\theta)^{-1} a(\theta) \,,$$

where

$$A(\theta) = \begin{pmatrix} \operatorname{tr}\left(K\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_1^2} (\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top)^{-1} \mathbf{Z}\bar{\Omega}_1\mathbf{Z}^\top \right) & \operatorname{tr}\left(K\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_1^2} (\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top)^{-1} \mathbf{Z}\bar{\Omega}_2\mathbf{Z}^\top \right) \\ \operatorname{tr}\left(K\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_2^2} (\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top)^{-1} \mathbf{Z}\bar{\Omega}_1\mathbf{Z}^\top \right) & \operatorname{tr}\left(K\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_2^2} (\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top)^{-1} \mathbf{Z}\bar{\Omega}_2\mathbf{Z}^\top \right) \end{pmatrix} .$$

So, the fixed point algorithm can be presented as follows:

$$\theta^{m+1} = A(\theta^{(m)})^{-1} a(\theta^{(m)})$$

At this time the fixed-point algorithm for $\theta = (\sigma_1^2, \sigma_2^2)$ will replace the corresponding step in Issue 1.

3.5.2.1 *N-S: Fixed-Point-Algorithm - Spatial Correlation*

To extend the above algorithm to not only being used for the estimation of $\theta = (\sigma_1^2, \sigma_2^2)$ but also for the spatial correlation parameter ρ_1 reconsider:

$$\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top = \mathbf{Z}\begin{pmatrix}\sigma_1^2\Omega_1 & \mathbf{o}_{D\times DT}\\ \mathbf{o}_{DT\times D} & \sigma_2^2\Omega_2\end{pmatrix}\mathbf{Z}^\top \tag{3.5}$$

and the specification of $\Omega_1(\rho_1) = \left((I-\rho_1\mathbf{W})^\top(I-\rho_1\mathbf{W})\right)^{-1}$:

$$\begin{aligned}\sigma_1^2\Omega_1(\rho_1) &= \sigma_1^2\Omega_1\Omega_1(I-\rho_1\mathbf{W})^\top(I-\rho_1\mathbf{W})\\ &= \sigma_1^2\left(\Omega_1\Omega_1 - \rho_1\Omega_1\Omega_1\mathbf{W}^\top - \rho_1\Omega_1\Omega_1\mathbf{W} + \rho_1^2\Omega_1\Omega_1\mathbf{W}^\top\mathbf{W}\right)\\ &= \sigma_1^2\left(\Omega_1\Omega_1 - \rho_1\Omega_1\Omega_1\mathbf{W}^\top\right) + \rho_1\left(-\sigma_1^2\Omega_1\Omega_1\mathbf{W} + \sigma_1^2\rho_1\Omega_1\Omega_1\mathbf{W}^\top\mathbf{W}\right)\end{aligned} \tag{3.6}$$

Thus equation 3.5 can be rewritten as:

$$\begin{aligned}\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top = \mathbf{Z}\Bigg[&\sigma_1^2\begin{pmatrix}\Omega_1\Omega_1 - \rho_1\Omega_1\Omega_1\mathbf{W}^\top & \mathbf{o}_{D\times DT}\\ \mathbf{o}_{DT\times D} & \mathbf{o}_{DT\times DT}\end{pmatrix}\\ &+\rho_1\begin{pmatrix}-\sigma_1^2\Omega_1\Omega_1\mathbf{W} + \sigma_1^2\rho_1\Omega_1\Omega_1\mathbf{W}^\top\mathbf{W} & \mathbf{o}_{D\times DT}\\ \mathbf{o}_{DT\times D} & \mathbf{o}_{DT\times DT}\end{pmatrix}\\ &+\sigma_2^2\begin{pmatrix}\mathbf{o}_{D\times D} & \mathbf{o}_{D\times DT}\\ \mathbf{o}_{DT\times D} & \Omega_2\end{pmatrix}\Bigg]\mathbf{Z}^\top\\ = &\begin{pmatrix}\mathbf{Z}\bar{\Omega}_{1,\sigma_1^2}\mathbf{Z}^\top & \mathbf{Z}\bar{\Omega}_{1,\rho_1}\mathbf{Z}^\top & \mathbf{Z}\bar{\Omega}_2\mathbf{Z}^\top\end{pmatrix}\begin{pmatrix}\sigma_1^2\\ \rho_1\\ \sigma_2^2\end{pmatrix}\end{aligned}$$

Thus equation 3.7 can be rewritten (analogously as above) to:

$$\psi(\mathbf{r})^\top\mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r}) = \operatorname{tr}\left(K\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}(\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top)^{-1}\begin{pmatrix}\mathbf{Z}\bar{\Omega}_{1,\sigma_1^2}\mathbf{Z}^\top & \mathbf{Z}\bar{\Omega}_{1,\rho_1}\mathbf{Z}^\top & \mathbf{Z}\bar{\Omega}_2\mathbf{Z}^\top\end{pmatrix}\begin{pmatrix}\sigma_1^2\\ \rho_1\\ \sigma_2^2\end{pmatrix}\right)$$

Let

$$\begin{pmatrix}\psi(\mathbf{r})^\top\mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\sigma_1^2}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r})\\ \psi(\mathbf{r})^\top\mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\rho_1}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r})\\ \psi(\mathbf{r})^\top\mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\sigma_2^2}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r})\end{pmatrix} = a(\theta)\,,$$

then

$$\theta = \begin{pmatrix} \sigma_1^2 \\ \rho_1 \\ \sigma_2^2 \end{pmatrix} = A(\theta)^{-1} a(\theta) \, ,$$

where

$$A(\theta) = \begin{pmatrix} \operatorname{tr}\left(\gamma(\sigma_1^2)\mathbf{Z}\bar{\Omega}_{1,\sigma_1^2}\mathbf{Z}^\top\right) & \operatorname{tr}\left(\gamma(\sigma_1^2)\mathbf{Z}\bar{\Omega}_{1,\rho_1}\mathbf{Z}^\top\right) & \operatorname{tr}\left(\gamma(\sigma_1^2)\mathbf{Z}\bar{\Omega}_2\mathbf{Z}^\top\right) \\ \operatorname{tr}\left(\gamma(\rho_1)\mathbf{Z}\bar{\Omega}_{1,\sigma_1^2}\mathbf{Z}^\top\right) & \operatorname{tr}\left(\gamma(\rho_1)\mathbf{Z}\bar{\Omega}_{1,\rho_1}\mathbf{Z}^\top\right) & \operatorname{tr}\left(\gamma(\rho_1)\mathbf{Z}\bar{\Omega}_2\mathbf{Z}^\top\right) \\ \operatorname{tr}\left(\gamma(\sigma_2^2)\mathbf{Z}\bar{\Omega}_{1,\sigma_1^2}\mathbf{Z}^\top\right) & \operatorname{tr}\left(\gamma(\sigma_2^2)\mathbf{Z}\bar{\Omega}_{1,\rho_1}\mathbf{Z}^\top\right) & \operatorname{tr}\left(\gamma(\sigma_2^2)\mathbf{Z}\bar{\Omega}_2\mathbf{Z}^\top\right) \end{pmatrix}$$

and $\gamma(\theta_l) = K\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_l}(\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top)^{-1}$

3.5.2.2 *More on the Fixed Point*

Inspired by: Chatrchi Golshid (2012): Robust Estimation of Variance Components in Small Area Estimation, Master-Thesis, Ottawa, Ontario, Canada: p. 16ff.:

> The fixed-point iterative method relies on the fixed-point theorem: "If g(x) is a continuous function for all $x \in [a; b]$, then g has a fixed point in $[a; b]$." This can be proven by assuming that $g(a) \geqslant a$ and $g(b) \leqslant b$. Since g is continuous the intermediate value theorem guarantees that there exists a c such that $g(c) = c$.

Starting from equation 2.18 where $\theta = \sigma_u^2$ we can rewrite the equation such that:

$$\Phi(\theta) = \psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r}) - \operatorname{tr}\left(K\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta}(\mathbf{Z}\mathbf{G}\mathbf{Z}^\top)^{-1}(\mathbf{Z}\mathbf{G}\mathbf{Z}^\top)\right) = 0 \tag{3.7}$$

Note that because the matrix $\mathbf{R}$ is assumed to be known for the FH model, it can be omitted. Note that under the simple Fay-Herriot Model $\mathbf{Z}\mathbf{G}\mathbf{Z}^\top = \sigma_u^2\mathbf{I}$, where $\mathbf{I}$ is a $(D \times D)$ identity matrix. Furthermore $\frac{\partial \mathbf{V}}{\partial \theta} = \mathbf{I}$. Thus equation 3.7 can be rewritten to:

$$\psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r}) = \operatorname{tr}\left(K\mathbf{V}^{-1}\mathbf{G}^{-1}\sigma_u^2\right)$$

This can be solved for the fixed Point and is directly presented in algorithmic notation, such that:

$$\theta^{m+1} = A(\theta^{(m)})^{-1} a(\theta^{(m)}) \, ,$$

where

$$A(\theta) = \mathrm{tr}\left(K\mathbf{V}^{-1}\mathbf{G}^{-1}\right)$$

and

$$a(\theta) = \psi(\mathbf{r})^{\top}\mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r})$$

Part II

IMPLEMENTATION

4

IMPLEMENTATION

- Zielke (1973): *Testmatrizen mit maximaler Konditionszahl*, Computing
  - Testmatrizen für LLS Probleme.
- Weihs / Mersmann / Ligges (2014): *Foundations of Statistical Algorithms*, CRC
  - Chapter 3 (Varification): How to test the results of a LLS solver. Generation of Test Data
- Kennedy / Gentle (1980): *Statistical Computing*
  - Chapter 10 (Unconstrained Optimization and Nonlinear Regression): Basic algorithms for solving nonlinear regression problem (e.g. Newton-Raphson)
  - Chapter 11.3 (Robust Estimators): Algorithms to solve robust estimators. Only a brief description of the status quo.
- Thisted (1988): *Elements of Statistical Computing*
  - Chapter 4 (Nonlinear Statistical Methods): All about solving ML problems
- Altman / Gill / McDonald (2004): *Numerical Issues in Statistical Computing for the Social Scientist*
  - Chapter 8 (Some Details of Nonlinear Estimation): How to evaluate the numerical results of a solution (Very important!)
  - Chapter 3 (Evaluating Statistical Software): How to verify the correctness of a statistical algorithm. Not as promising as it sounds.
- Sugasawa / Kubokawa (2015): *Parametric transformed Fay–Herriot model for small area estimation*, Journal of Multivariate Analysis
  - Box-Cox und vergleichbare Transformationen für Fay-Herriot Modelle
- Alfons, Templ and Filzmoser (2010): An Object-Oriented Framework for Statistical Simulation: The R Package simFrame
- Alfons et.al. (2010): Simulation of synthetic population data for houshold surveyes with application to EU-SILC

4.1 SOFTWARE

4.2 VERIFICATION OF RESULTS

4.3 ACCURACY OF RESULTS

4.4 VALIDATION OF RESULTS

4.4.1 *Four Steps*

Section based on McCullough (2004)

Part III

RESULTS

This is the part where I will present all results.

5 NUMERICAL PROPERTIES

5.1 ACCURACY

5.2 STABILITY

5.3 SPEED OF CONVERGENCE

6

SIMULATION STUDIES

6.1 MODEL BASED SIMULATION STUDIES

6.1.1 *The Area-Level Perspective*

In this section we present some results of a simulation study. To make the results comparable to other model based simulations on area level models we begin in this section with a simulation study on area level. Thus we can discuss area level outliers which is what all the others do. Section? will then introduce then a simulation study in which we start with a unit-level population and can thus introduce both, outlying observations and areas.

To begin with we define the area level model from which we draw the data:

$$\bar{y}_i = 100 + 1 \cdot x_i + v_i + \bar{e}_i$$

- The single regressor, x, is a deterministic sequence defined as $x_i = \frac{i}{2D} + 1$ where D is the number of domains (taken from spatio temporal FH).
- The random effect, v, is drawn from a normal distribution, i.e. $v_i \sim \mathcal{N}(0, \sigma_u^2)$ where σ_u^2 is defined with respect to the scenario.
- The sampling error, e, is drawn from $e_i \sim \mathcal{N}(0, \sigma_{ei}^2)$ where σ_{ei}^2 is an equidistant sequence from 0.8 to 1.2 with D elements.
- General characteristics: $D = 100$ and $R = 500$ being the sumber of Monte Carlo repetitions.

To illustrate the greatness of the model we investigate two different scenarios:

1. *(o, o)* This is the scenario where the Fay Herriot model holds. In this scenario $\sigma_u^2 = 1$.
2. *(v, o)* This is the scenario with area level outliers. $\sigma_u^2 = 1$ for 95 % of the areas, i.e. for $i \in \{1, \ldots, 95\}$, and $\sigma_u^2 = 20^2$ for $i \in \{96, \ldots, 100\}$.

6.1.1.1 *Estimation of the MSE*

To begin with we define the area level model from which we draw the data:

$$\tilde{y}_i = 100 + 2 \cdot x_i + u_i + e_i$$

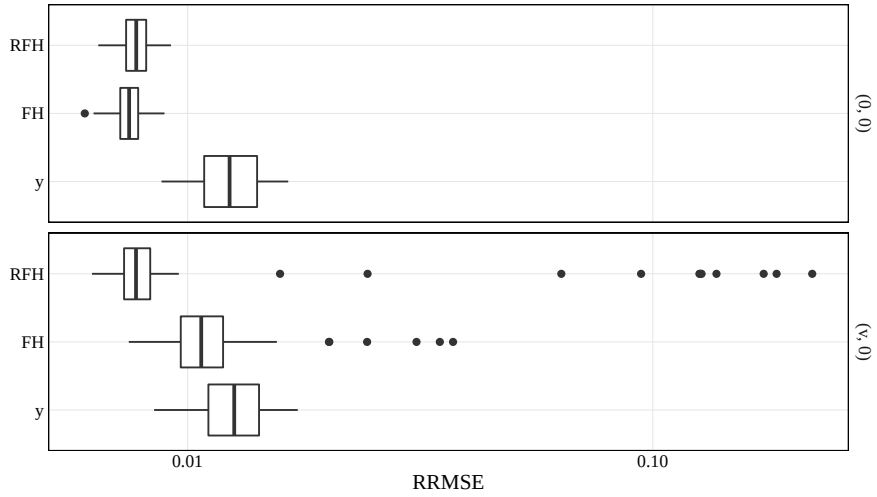

Figure 6.1: Boxplot with Relative Root Mean Squared Error (RRMSE)

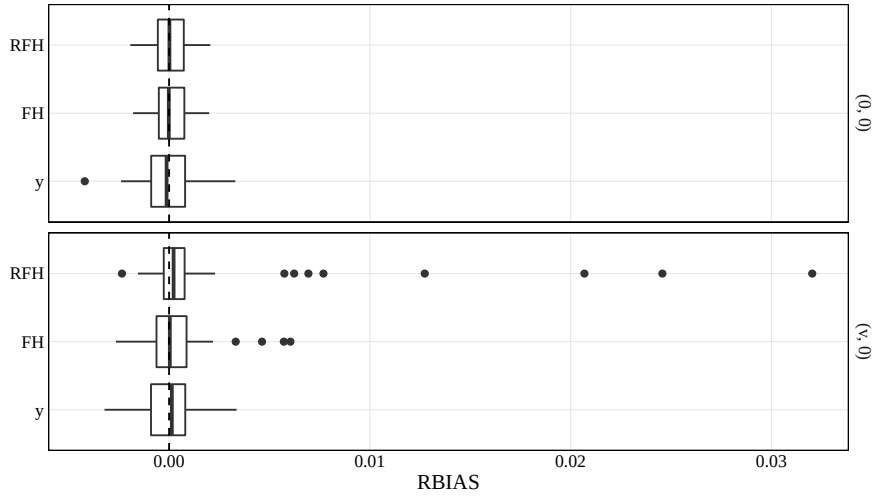

Figure 6.2: Boxplot with Relative Bias (RBIAS)

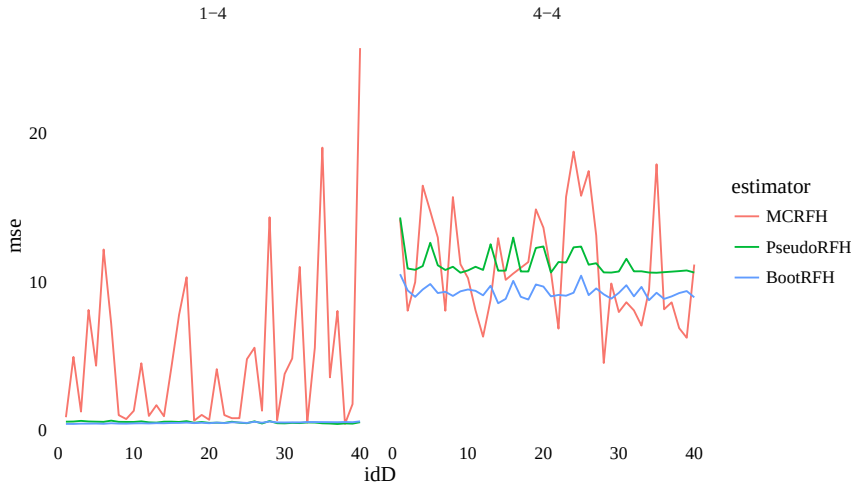

Figure 6.3: Absolute values of the estimated Mean Squared Error using the pseudo linearizatin compared to the Monte Carlo MSE

6.1.2 *From Unit to Area Level Data*

In this section we want to present a different perspective for simulation studies on area level models. Namely by starting from the unit level population data. This allows for a number of interesting aspects which can be investigated and explained:

6.1.3 *From Unit to Area Level Models*

In later simulation studies we will consider data in which area level statistics are computed from individual information. From a contextual point of view, starting from individual information is advantageous in the sense that outlying areas can be motivated more easily. Also the question for a good estimator for the sampling variances can be motivated when knowing the underlying individual model. Hence, I will derive the Fay-Herriot model starting from unit-level. Consider the following model:

$$y_{ij} = x_i^\top \beta + u + e_i \,,$$

where y_{ij} is the response in domain i of unit j with $i = 1, \ldots, n_i$, where n_i is the number of units in domain i. u is an area specific random effect following (i.i.d.) a normal distribution with zero mean and σ_u^2 as variance parameter. e_{ij} is the remaining deviation from the model, following (i.i.d.) a normal distribution with zero mean and $\sigma_{e_i}^2$ as variance parameter. This unit level model is defined under strong assumptions, still, assumptions most practitioner are willing to make which could simplify the identification of the sampling variances under the area level model.

From this model consider the area statistics $\tilde{y}_i = \frac{1}{n_j} \sum_{j=1}^{n_i} y_{ij}$, for which an area level model can be derived as:

$$\tilde{y}_i = x_i^\top \beta + u_i + e_i$$

Considering the mean in a linear model, it can be expressed as $\bar{y} = \bar{x}\beta$; the random effect was defined for each area, hence it remains unaltered for the area level model. The error term in this model can be expressed as the sampling error and its standard deviation as the (conditional) standard deviation of the aggregated area statistic, which in this case is a mean. Hence, $e_i \sim \mathcal{N}(0, \sigma^2_{ei} = \sigma^2_e/n_i)$. Under this unit level model a sufficient estimator for σ^2_{ei} can be derived from estimating σ^2_{ei}, which can be done robust and non-robust in many ways.

- Under the Fay Herriot model we specify an area level model with heteroscedastic sampling distribution. The model does not explain what the source of heteroscedasticity is. We induce two different sources, one is the sampling size varying accros domains, and the other one are unit level outliers.
- When we begin from the unit level we may ask what the true sampling variance is; which is assumed to be known under the model. In practice it is an estimated quantity which itself, like a direct estimator of the mean, is unreliable. So it will be relevant to discuss how the Fay Herriot estimator is performing when the direct variance estimator is an additional source of uncertainty. Recent research suggests that smoothed variances should be plugged into the FH estimator. In contrast to this discussion we will argue that direct variance estimators are a viable source of information when unit level outliers are an issue.
- In the context of unit level outliers it can be intuitive to suggest a robust direct estimator instead of the sample mean. Hence we compare how the use of a median and a huber type estimator of the mean compete against the sampling mean.

In general our simulation setup borrows from unit level scenarios from the literature to make this exercise as convenient as possible. The basis is again a linear mixed model, this time defined on the unit population level:

$$y_{ij} = 100 + 1 \cdot x_i + v_i + e_{ij}$$

- The regressor, x_i, and random effect, v_i, are defined in the same way as for the area level scenario.
- The error term, e_{ij}, is defined as $e_{ij} \sim \mathcal{N}(0, \sigma^2_e)$ where σ^2_e is varies across simulation scenarios.
- From this populatin model we draw samples with simple random sampling without replacement. The sample sizes are $n_i \in \{5, \ldots, 15\}$ and $N_i = 1000$; D is again 100.

- The sample is then aggregated using different direct estimators. The sample mean, the sample median and a robust direct estimator (huber m-type). Note that x_i is constant within domains. For the variance estimation we use the sample variance, a generalized variance function which can be considered optimal under the population model and the median absolute deviance from the median within domains.
- On area level the standard Fay Herriot model is used with different variance estimators and corresponding direct estimators and compared with the RFH.

With these settings we are interested in several different choices varying across simulation scenarios. We want to emphasise how unit and area level outliers can influence area level predictions.

1. *(o, o)* This is the scenario in which the area level model as described in section? holds, i.e. there are no outliers. Here $\sigma_e^2 = 4^2$ and $\sigma_u^2 = 1$ for all domains. However, the sampling variances ($\sigma_{ei}^2 = \frac{\sigma_e^2}{n_i}$) derived under the unit level population model range from 1 to 3.2 for their respective sample size. $\sigma_u^2 = 1$.
2. *(v, o)* This scenario is close to the area level data generation in the previous section where we induced area level outliers. For that purpose we choose $\sigma_u^2 = 1$ for the areas where $i \in \{1, \ldots, 95\}$ and $\sigma_u^2 = 20^2$ for $i \in \{96, \ldots, 100\}$
3. *(o, e)* In this scenario unit level outliers do exist. To make the magnitude comparable to simulation studies in the literature we choose $\sigma_e^2 = 150^2$ for $i \in \{90, \ldots, 95\}$.
4. *(v, e)* This scenario is the combination of 2 and 3 where we have area level and unit level outliers, however not in the same domains.

6.2 DESIGN BASED SIMULATION STUDIES

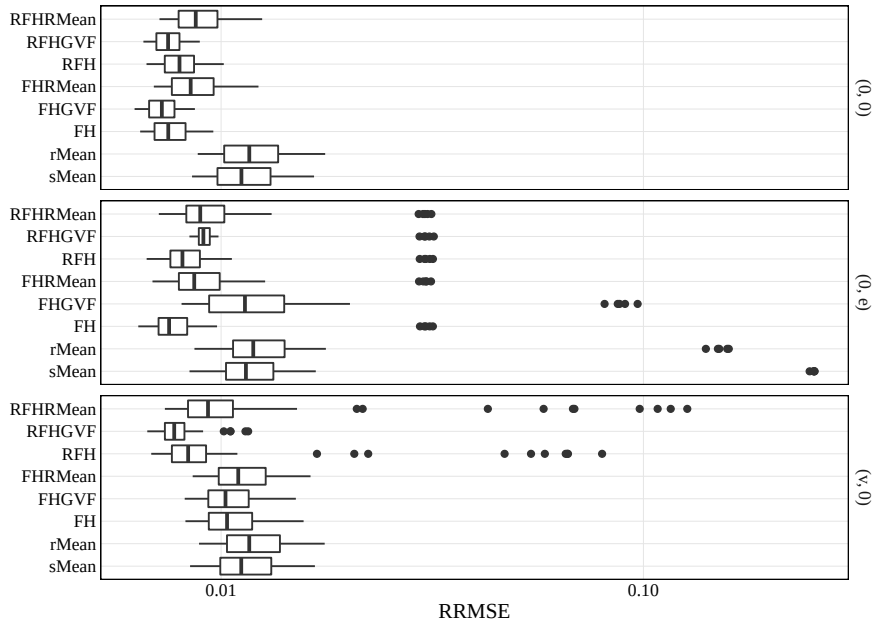

Figure 6.4: Boxplot with Relative Root Mean Squared Error (RRMSE)

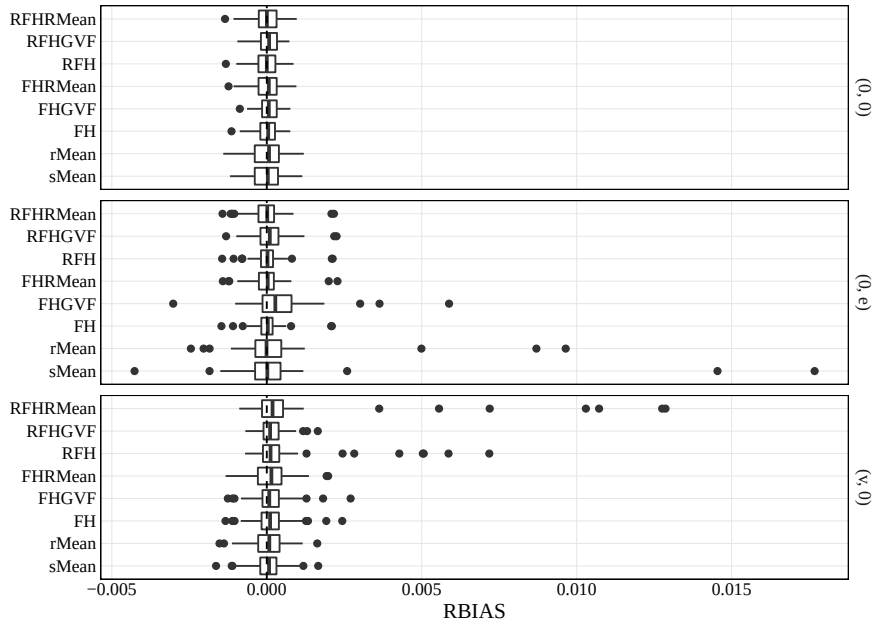

Figure 6.5: Boxplot with Relative Bias (RBIAS)

Part IV

APPENDIX

BIBLIOGRAPHY

Abelson, R. P. (1995). *Statistics as Principled Argument*. New York: Taylor & Francis Group.

Battese, G. E., R. M. Harter, and W. A. Fuller (1988). "An error component model for prediction of county crop areas using survey and satellite data." In: *Journal of the American Statistical Association* 83 (401), pp. 28–36.

Beaumont, J.-F. (2004). "Robust generalized regression estimation." In: *Survey Methodology* 30.2, pp. 195–208.

Beaumont, J.-F. and L.-P. Rivest (2009). "Sample Surveys: Design, Methods and Applications." In: ed. by D. Pfeffermann and C. R. Rao. Vol. 29 A. Handbook of Statistics. Elsevier. Chap. Dealing with Outliers in Survey Data, pp. 247–280.

Bell, W. R. and E. T. Huang (2006). "Using the t-distribution to Deal with Outliers in Small Area Estimation." In: *Proceedings of Statistics Canada Symposium 2006: Methodological Issues in Measuring Population Health*.

Benavent, R. and D. Morales (2016). "Multivariate Fay-Herriot models for small area estimation." In: *Computational Statistics & Data Analysis* 94, pp. 372–390.

Chambers, R. (1986). "Outlier robust finite population estimation." In: *Journal of the American Statistical Association* 81, pp. 1063–1069.

Chambers, R., H. Chandra, N. Salvati, and N. Tzavidis (2014). "Outlier Robust Small Area Estimation." In: *Journal of the Royal Statistical Society: Series B* 76 (1), pp. 47–69.

Chambers, R., H. Chandra, and N. Tzavidis (2011). "On bias-robust mean squared error estimation for pseudo-linear small area estimators." In: *Survey Methodology* 37 (2), pp. 153–170.

Chandra, H., N. Salvati, R. Chambers, and N. Tzavidis (2012). "Small Area Estimation under Spatial Nonstationarity." In: *Computational Statistics & Data Analysis* 56.10, pp. 2875–2888.

Chen, S. and P. Lahiri (2003). "A comparison of different MSPE estimators of EBLUP for the Fay-Herriot model." In: *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 905–911.

— (2008). "On Mean Squared Prediction Error Estimation in Small Area Estimation Problems." In: *Communications in Statistics-Theory and Methods* 37.11, pp. 1792–1798.

Choudhry, G. H. and J. N. K. Rao (1989). "Small area estimation using models that combine time series and cross-sectional data." In: *Proceeding of the Statistics Canada Symposium on Analysis of Data in Time*, pp. 67–74.

Clement, E. P. (2014). "Small Area Estimation with Application to Disease Mapping." In: *International Journal of Probability and Statistics* 3.1, pp. 15–22.

Cressie, N. (1993). *Statistics for spatial data*. New York: Wiley.

Datta, G. S. (2009). "Handbook of Statistics: Inference and Analysis." In: ed. by D. Pfeffermann and C.R. Rao. Vol. 29 B. Elsevier. Chap. Model-Based Approach to Small Area Estimation, pp. 251–288.

Datta, G. S. and P. Lahiri (1995). "Robust Hierarchical Bayes Estimation of Small Area Characteristics in the Presence of Covariates and Outliers." In: *Journal of Multivariate Analysis* 54, pp. 310–328.

— (2000). "A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems." In: *Statistica Sinica* 10, pp. 613–627.

Datta, G. S., P. Lahiri, and T. Maiti (2002). "Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data." In: *Journal of Statistical Planning and Inference* 102.1, pp. 83–97.

Datta, G. S., J. N. K. Rao, and D. D. Smith (2005). "On measuring the variability of small area estimators under a basic area level model." In: *Biometrika* 92 (1), pp. 183–196.

Efron, B. and C. Morris (1971). "Limiting the Risk of Bayes and Empirical Bayes Estimators - Part I: The Bayes Case." In: *Journal of the American Statistical Association* 66.336, pp. 807–815.

— (1972). "Limiting the Risk of Bayes and Empirical Bayes Estimators - Part II: The Empirical Bayes Case." In: *Journal of the American Statistical Association* 67.337, pp. 130–139.

Fay, R. E. and R. A. Herriot (1979). "Estimation of Income for Small Places: An Application of James-Stein Procedures to Census Data." In: *Journal of the American Statistical Association* 74 (366), pp. 269–277.

Fellner, W. H. (1986). "Robust Estimation of Variance Components." In: *Technometrics* 28 (1), pp. 51–60.

Ghosh, M. and J. N. K. Rao (1994). "Small Area Estimation: An Appraisal." In: *Statistical Science* 9 (1), pp. 55–93.

Guadarrama, M., I. Molina, and J. N. K. Rao (2015). *A Comparison of Small Area Estimation Methods for Poverty Mapping*. UC3M Working Papers: Statistics and Econometrics 15-05. Universidad Carlos III de Madrid, Departamento de Estadistica.

Hall, P. and T. Maiti (2006). "On Parametric Bootstrap Methods for Small Area Prediction." In: *Journal of the Royal Statistical Society: Series B* 68 (2), pp. 221–238.

Henderson, C. R. (1950). "Estimation of genetic parameters." In: *Annals of Mathematical Statistics* 21 (2), pp. 309–310.

— (1975). "Best Linear Unbiased Estimation and Prediction under a Selection Model." In: *Biometrics* 31 (2), pp. 423–447.

Horvitz, D.G. and D.J. Thompson (1952). "A generalization of sampling without replacement from a finite universe." In: *Journal of the American Statistical Association* 47, pp. 663–685.

Jiang, J. and P. Lahiri (2006). "Mixed Model Prediction and Small Area Estimation." In: *Test* 15 (1), pp. 1–96.

Jiang, J., P. Lahiri, and S.-M. Wan (2002). "A Unified Jackknife Theory for Empirical Best Prediction with M-Estimation." In: *The Annals of Statistics* 30 (6), pp. 1782–1810.

Jiongo, V. D. and P. Nguimkeu (2014). *Bootstrapping Mean Squared Errors of Robust Small Area Estimators*. Tech. rep. URL: http://www2.gsu.edu/~econpn/Pierre_Nguimkeu/RESEARCH_files/bootstrap_mse_24122014u.pdf.

Kacker, R. N. and D. A. Harville (1984). "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models." In: *Journal of the American Statistical Association* 79, pp. 853–862.

Lehtonen, R. and A. Veijanen (2009). "Sample Surveys: Inference and Analysis." In: ed. by D. Pfeffermann and C.R. Rao. Vol. 29 B. Handbook of Statistics. Elsevier. Chap. Design-based Methods of Estimation for Domains and Small Areas, pp. 219–249.

Lohr, S. L. and J. N. K. Rao (2009). "Jackknife estimation of mean squared error of small area predictors in nonlinear mixed models." In: *Biometrika*, pp. 1–12.

López-Vizcaíno, E., M. J. Lombardía, and D. Morales (2015). "Small area estimation of labour force indicators under a multinomial model with correlated time and area effects." In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178.3, pp. 535–565.

Maiti, T., H. Ren, and S. Sinha (2014). "Prediction Error of Small Area Predictors Shrinking Both Means and Variances." In: *Scandinavian Journal of Statistics* 41.3, pp. 775–790.

Marhuenda, Y., I. Molina, and D. Morales (2013). "Small area estimation with spatio-temporal Fay-Herriot models." In: *Computational Statistics and Data Analysis* 58, pp. 308–325.

McCullough, B. D. (2004). "Numerical Issues in Statistical Computing for the Social Scientist." In: ed. by D. J. Balding, N. A. C. Cressie, and N. I. Fisher. John Wiley & Sons, Inc. Chap. Some Details of Nonlinear Estimation, pp. 199–218.

Molina, I., N. Salvati, and M. Pratesi (2009). "Bootstrap for estimating the mean squared error of the Spatial EBLUP." In: *Computational Statistics* 24 (3), pp. 441–458.

Namazi-Rad, M.-R. and D. Steel (2015). "What Level of Statistical Model Should We Use in Small Area Estimation?" In: *Australian & New Zealand Journal of Statistics* 57.2, pp. 275–298.

Petrucci, A. and N. Salvati (2006). "Small Area Estimation for Spatial Correlation in Watershed Erosion Assessment." In: *Journal of agricultural, biological, and environmental statistics* 11.2, pp. 169–182.

Pfeffermann, D. (2002). "Small Area Estimation - New Developments and Directions." In: *International Statistical Review* 70 (1), pp. 125–143.

— (2013). "New Important Developments in Small Area Estimation." In: *Statistical Science* 28 (1), pp. 40–68.

Porter, A. T., S. H. Holan, C. K. Wikle, and N. Cressie (2014). "Spatial Fay-Herriot models for small area estimation with functional covariates." In: *Spatial Statistics* 10, pp. 27–42.

Prasad, N. G. N. and J. N. K. Rao (1990). "The Estimation of the Mean Squared Error of Small Area Estimators." In: *Journal of the American Statistical Association* 85 (409), pp. 163–171.

Pratesi, M. and N. Salvati (2008). "Small area estimation: the EBLUP estimator based on spatially correlated random area effects." In: *Statistical Methods & Applications* 17, pp. 113–141.

Rao, J. N. K. (1999). "Some recent advances in model-based small area estimation." In: *Survey Methodology* 25, pp. 175–186.

— (2003). *Small Area Estimation*. New York: Wiley.

Rao, J. N. K. and I. Molina (2015). *Small Area Estimation*. New York: Wiley.

Rao, J. N. K. and M. Yu (1994). "Small-Area Estimation by Combining Time-Series and Cross-Sectional Data." In: *Canadian Journal of Statistics* 22.4, pp. 511–528.

Rivest, L.-P. and N. Vandal (2003). "Mean Squared Error Estimation for Small Areas when the Small Area Variances are Estimated." In: *Proceedings of the International Conference on Recent Advances in Survey Sampling*. Laboratory for Research in Statistics and Probability, Carleton University Ottawa, ON, Canada.

Salvati, N., N. Tzavidis, M. Pratesi, and R. Chambers (2012). "Small area estimation via M-quantile geographically weighted regression." In: *Test* 21.1, pp. 1–28.

Särndal, C.-E., B. Swensson, and J.H. Wretman (1992). *Model Assisted Survey Sampling*. New York: Springer.

Schmid, T. (2012). "Spatial Robust Small Area Estimation applied on Business Data." PhD thesis. University of Trier.

Searle, S. R. (1971). *Linear Models*. New York: Wiley.

Singh, A. C., H. J. Mantel, and B. W. Thomas (1991). *Time Series Generalization of Fay-Herriot Estimation for Small Areas*. Methodology Branch, Business Survey Methods Division, Statistics Canada.

Singh, B. B., G. K. Shukla, and D. Kundu (2005). "Spatio-temporal models in small area estimation." In: *Survey Methodology* 31.2, p. 183.

Sinha, S. K. and J. N. K. Rao (2009). "Robust Small Area Estimation." In: *The Canadian Journal of Statistics* 37 (3), pp. 381–399.

Slud, E. V. and T. Maiti (2006). "Mean-squared error estimation in transformed Fay-Herriot models." In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.2, pp. 239–257.

Sugasawa, S. and T. Kubokawa (2015). "Parametric transformed Fay-Herriot model for small area estimation." In: *Journal of Multivariate Analysis* 139, pp. 295–311.

Tzavidis, N., S. Marchetti, and R. Chambers (2010). "Robust estimation of small area means and quantiles." In: *Australian and New Zealand Journal of Statistics* 52 (2), pp. 167–186.

Wang, J. and W. A. Fuller (2003). "The Mean Squared Error of Small Area Predictors Constructed With Estimated Area Variances." In: *Journal of the American Statistical Association* 98 (463), pp. 716–723.

Warnholz, S. and T. Schmid (2015). *saeSim: Simulation Tools for Small Area Estimation*. R package version 0.7.0. URL: http://CRAN.R-project.org/package=saeSim.

Wolter, K. (2007). *Introduction to Variance Estimation*. Springer Science & Business Media.

You, Y. and B. Chapman (2006). "Small Area Estimation Using Area Level Models and Estimated Sampling Variances." In: *Survey Methodology* 32.1, pp. 97–103.