ROBUST EXTENSIONS OF THE FAY HERRIOT MODEL IN SMALL AREA ESTIMATION

SEBASTIAN WARNHOLZ

Theory, Implementation and Simulation Studies

December 11, 2015 – version 0.1

Sebastian Warnholz: *Robust Extensions of the Fay Herriot Model in Small Area Estimation,* Theory, Implementation and Simulation Studies, © December 11, 2015

SUPERVISORS:
Prof. Dr. Timo Schmid
Nikos Tzavidis

LOCATION:
Berlin

TIME FRAME:
December 11, 2015

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 (7 *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)

7 *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 7. 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 MIXED MODELS

3

ROBUST STATISTICS

Gervini / Yohai (2002): *A Class of Robust and Fully Efficient Regression Estima tors*

4

SMALL AREA ESTIMATION

4.1 UNIT LEVEL MODELS

4.2 AREA LEVEL MODELS

Area Level Models in Small Area Estimation play an important role in the production of reliable domain estimates.

- They can be used even if unit level observations are not accessible.
- In a model based estimation it is largely unsolved to incorporate design weights. Area level models can be used to start from a direct design based estimator.
- Unit level models often have problems with heterogeneity. An assumption, for example, for unit level data is that the error terms of a model are homescedastic given the random effects. This assumption is often not plausible and may call for more complex assumptions on the variance structure of the data. However such structures may or may not be known and cannot be modelled easily. This can also lead to computationally demanding procedures.

Given these considerations the most important factor to choose cadidate models is the availability of data. Very often there is not much of a choice but rather a decission given the available information. And given the availability of unit level data, the obvious choice is to consider a model which can use such information. Only if that fails for one of the above reasons can an area level model be of interest.

4.2.1 *The Fay Herriot Model*

A frequently used model in Small Area Estimation is a model introduced by Fay and Herriot (1979). It starts on the area level and is used in small area estimation for research on area-level. It is build on a sampling model:

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

where $\tilde{y}_i$ is a direct estimator of a statistic of interest θ_i for an area i with $i = 1, \ldots, D$ and D being the number of areas. The sampling error e_i is assumed to be independent and normally distributed with known variances $\sigma^2_{e,i}$, i.e. $e_i|\theta_i \sim \mathcal{N}(0, \sigma^2_{e,i})$. The model is modified

with the linking model by assuming a linear relationship between the true area statistic θ_i and some diterministic auxiliary variables x_i:

$$\theta_i = x_i^\top \beta + u_i$$

Note that x_i is a vector containing area-level (aggregated) information for P variables and β is a vector $(1 \times P)$ of regression coefficients describing the (linear) relationship. The model errors u are assumed to be independent and normally distributed, i.e. $u_i \sim \mathcal{N}(0, \sigma_u^2)$ furthermore e_i and u_i are assumed to be independent. Combining the sampling and linking model leads to:

$$\tilde{y}_i = x_i^\top \beta + u_i + e_i. \tag{1}$$

4.2.2 *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, $e_i \sim \mathcal{N}(0, \sigma_{e,i}^2 = \sigma_e^2/n_i)$. Under this unit level model a sufficient estimator for $\sigma_{e,i}^2$ can be derived from estimating $\sigma_{e,i}^2$, which can be done robust and non-robust in many ways.

4.2.3 *Spatio-Temporal Fay Harriot model*

The model stated in equation 1 has been modified for including historical information by modelling autocorrelated model errors and also by allowing for spatial correlation (in the model error). See the discussion in Marhuenda, Molina, and Morales (2013) for more details. Marhuenda, Molina, and Morales (2013) allow for both spatial and temporal correlation in the model errors. Hence the sampling model is (simply) extended to include historical information:

$$y_{dt} = \mu_{dt} + e_{dt},$$

with $d = 1, \dots, D$ and $t = 1, \dots, T$ where D and T are the total number of areas and time periods respectively. Here $e_{dt} \sim N(0, \sigma^2_{dt})$ are independent with known variances σ^2_{dt}. The model error is composed of a spatial autoregressive process of order 1 (SAR(1)) and an autoregressive process of order 1 (AR(1)):

$$\mu_{dt} = x^\top_{dt}\beta + u_{1d} + u_{2dt},$$

where u_{1d} and u_{2dt} follow a SAR(1) and AR(1) respectively:

$$u_{1d} = \rho_1 \sum_{l \neq d} w_{d,l} u_{1l} + \epsilon_{1d},$$

where $|\rho_1| < 1$ and $\epsilon_{1d} \sim N(0, \sigma^2_1)$ are i.i.d. with $d = 1, \dots, D$. $w_{d,l}$ are the elements of W which is the row standardized proximity matrix W^0. The elements in W^0 are equal to 1 if areas are neighboured and 0 otherwise (an area is not neighboured with itself) - thus the dimension of W^0 is $D \times D$. As stated above u_{2dt} follows an AR(1):

$$u_{2dt} = \rho_2 u_{2d,t-1} + \epsilon_{2dt},$$

where $|\rho_2| < 1$ and $\epsilon_{2dt} \sim N(0, \sigma^2_2)$ are i.i.d. with $d = 1, \dots, D$ and $t = 1, \dots, T$. Note that u_{1d} and u_{2dt} and e_{dt} 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_{dt} as elements, $\mathbf{X}$ is the $DT \times p$ design matrix containing the vectors $x^\top_{dt}$ 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_{dt}. Note that $\mathbf{u} = (u_1^\top, u_2^\top)$ where the $D \times 1$ vector u_1 and $DT \times 1$ vector u_2 have u_{1d} and u_{2dt} 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. $\mathbf{e} \sim N(\mathbf{o}, \mathbf{V}_e)$ where $\mathbf{V}_e$ is a diagonal matrix with the known σ^2_{dt} on the main diagonal. $\mathbf{u} \sim N(\mathbf{o}, \mathbf{V}_u(\theta))$ with the block diagonal covariance matrix $\mathbf{V}_u(\theta) = \text{diag}(\sigma_1^2\Omega_1(\rho_1), \sigma_2^2\Omega_2(\rho_2))$ where $\theta = (\sigma_1^2, \rho_1, \sigma_2^2, \rho_2)$.

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

and follows from the SAR(1) process in the model errors. $\Omega_2(\rho_2)$ has a block diagonal structure with $\Omega_{2d}(\rho_2)$ denoting the blocks where the definition of $\Omega_{2d}(\rho_2)$ follows from the AR(1) process:

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

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

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

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

$$\tilde{\beta}(\theta) = \Big(\mathbf{X}^\top\mathbf{V}^{-1}(\theta)\mathbf{X}\Big)^{-1}\mathbf{X}^\top\mathbf{V}^{-1}(\theta)\mathbf{y}$$

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

Hence the BLUP of u_1 and u_2 can be stated as:

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

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

Estimating θ leads to the EBLUE for β and EBLUPs for u_1 and u_2, hence an predictor for the area characteristic μ_{dt} is given by:

$$\hat{\mu}_{dt} = x_{dt}^\top\hat{\beta} + \hat{u}_{1d} + \hat{u}_{2dt}$$

Marhuenda, Molina, and Morales (2013) use a restricted maximum likelihood method to estimate θ 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.

4.3 ROBUST METHODS IN SMALL AREA ESTIMATION

4.3.1 *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:

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

$$\Phi_l(\theta) = \psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} \mathbf{V}^{-1} \mathbf{U}^{\frac{1}{2}} \psi(\mathbf{r}) - \operatorname{tr}\left(K \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l}\right) = 0 \quad (3)$$

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.

4.3.2 *Mean squared error*

4.3.2.1 *Bootstrap*

4.3.2.2 *Pseudo Linearization*

Chambers, J. Chandra, and Tzavidis (2011) and Chambers, H. Chandra, 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:

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

where

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

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

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

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 $\mathbf{w}_{is}^{\mathrm{RBLUP}}$ we have to use $\mathbf{w}_{is}^{\mathrm{REBLUP}}$. Then the estimator of the conditional MSE is given by:

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

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

with

$$a_{ij} = N_i w_{ij}^{\mathrm{REBLUP}} - I\,(j \in i)$$

and

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

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

5 ROBUST AREA LEVEL MODELS

5.1 FORMULATION

5.1.1 *Standard*

5.1.2 *Spatio Temporal*

5.2 MEAN SQUARED ERROR

5.2.0.1 *Pseudo Linearization*

This is the representation of the pseudo linear representation of the FH model. As it is introduced in Chambers, J. Chandra, and Tzavidis (2011) and Chambers, H. Chandra, 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

$$\theta_i = \gamma_i y_i + (1-\gamma_i) x_i^\top \beta \tag{4}$$

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

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

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:

$$\theta_i = x_i^{\top}\beta + u_i \tag{5}$$

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:

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

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

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

5.3 ALGORITHM

5.3.1 *Newton Raphson Algorithms*

Sinha and Rao (2009) propose a Newton-Raphson algorithm to solve equations 2 and 3 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:

$$\frac{\partial \Phi}{\partial \theta_l} = 2\frac{\partial}{\partial \theta_l}\left(\psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\right)\frac{\partial \mathbf{V}}{\partial \theta_l}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r}) + \operatorname{tr}\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_l}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_l}\mathbf{K}\right) \tag{6}$$

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 6 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 6 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):

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

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

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

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:

$$||\beta^{(m+1)} - \beta^{(m)}||^2 <$$

$$(\sigma_1^{2(m+1)} - \sigma_1^{2(m)})^2 + (\sigma_2^{2(m+1)} - \sigma_2^{2(m)})^2 + (\rho_1^{(m+1)} - \rho_1^{(m)})^2 + (\rho_2^{(m+1)} - \rho_2^{(m)})^2 <$$

5.3.2 *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 3 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:

$$\Phi_l(\theta) = \psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} \mathbf{V}^{-1} \mathbf{U}^{\frac{1}{2}} \psi(\mathbf{r}) - \operatorname{tr}\left(K \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} (\mathbf{Z} \mathbf{V}_u \mathbf{Z}^\top)^{-1} (\mathbf{Z} \mathbf{V}_u \mathbf{Z}^\top) \right) = 0 \quad (7)$$

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 10 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.

5.3.2.1 *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:

$$\mathbf{Z} \mathbf{V}_u \mathbf{Z}^\top = \mathbf{Z} \begin{pmatrix} \sigma_1^2 \Omega_1 & \mathbf{o}_{D \times DT} \\ \mathbf{o}_{DT \times D} & \sigma_2^2 \Omega_2 \end{pmatrix} \mathbf{Z}^\top \tag{8}$$

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

$$\begin{aligned} \sigma_1^2 \Omega_1(\rho_1) &= \sigma_1^2 \Omega_1 \Omega_1 (I - \rho_1 \mathbf{W})^\top (I - \rho_1 \mathbf{W}) \\ &= \sigma_1^2 \left(\Omega_1 \Omega_1 - \rho_1 \Omega_1 \Omega_1 \mathbf{W}^\top - \rho_1 \Omega_1 \Omega_1 \mathbf{W} + \rho_1^2 \Omega_1 \Omega_1 \mathbf{W}^\top \mathbf{W} \right) \\ &= \sigma_1^2 \left(\Omega_1 \Omega_1 - \rho_1 \Omega_1 \Omega_1 \mathbf{W}^\top \right) + \rho_1 \left(-\sigma_1^2 \Omega_1 \Omega_1 \mathbf{W} + \sigma_1^2 \rho_1 \Omega_1 \Omega_1 \mathbf{W}^\top \mathbf{W} \right) \end{aligned} \tag{9}$$

Thus equation 8 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 10 can be rewritten (analogously as above) to:

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

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

5.3.2.2 *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 3 where $\theta = \sigma_u^2$ we can rewrite the equation such that:

$$\Phi(\theta) = \psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta} \mathbf{V}^{-1} \mathbf{U}^{\frac{1}{2}} \psi(\mathbf{r}) - \operatorname{tr}\left(K \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta} (\mathbf{Z}\mathbf{G}\mathbf{Z}^\top)^{-1} (\mathbf{Z}\mathbf{G}\mathbf{Z}^\top) \right) = 0 \tag{10}$$

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 10 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

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

and

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

5.4 UNIT LEVEL OUTLIERS AND AREA LEVEL MODELS

Part II

IMPLEMENTATION

6 ALGORITHMS AND STATISTICS

6.1 NON-LINEAR OPTIMIZATIONS IN STATISTICS

6.2 ALGORITHMS FOR ROBUST ESTIMATORS IN STATISTICS

7 IMPLEMENTATION

7.1 SOFTWARE

7.2 VERIFICATION OF RESULTS

7.3 ACCURACY OF RESULTS

7.4 VALIDATION OF RESULTS

7.4.1 *Four Steps*

Section based on McCullough (2004)

Part III

RESULTS

This is the part where I will present all results.

8 NUMERICAL PROPERTIES

8.1 ACCURACY

8.2 STABILITY

8.3 SPEED OF CONVERGENCE

9

SIMULATION STUDIES

9.1 MODEL BASED SIMULATION STUDIES

9.1.1 *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 $e_i \sim \mathcal{N}(0, \sigma_{e,i}^2)$ where $\sigma_{e,i}^2$ 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 $i \in \{96, \ldots, 100\}$.

9.1.2 *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

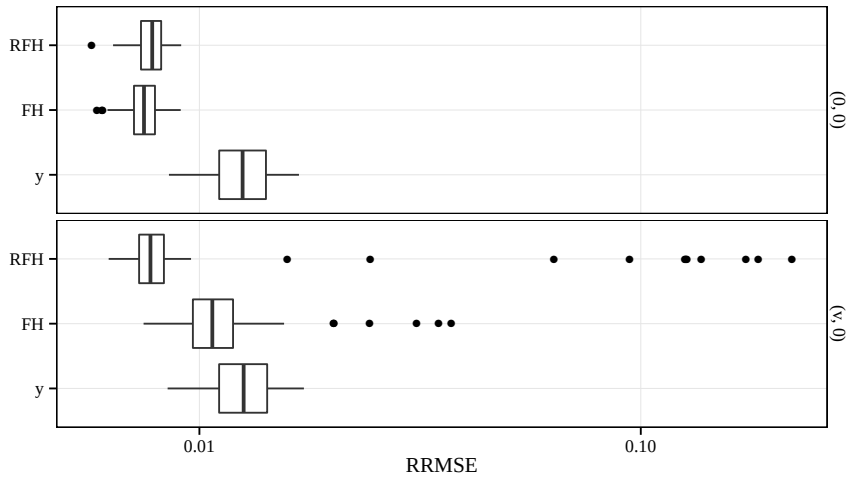

Figure 1: Boxplot with Relative Root Mean Squared Error (RRMSE)

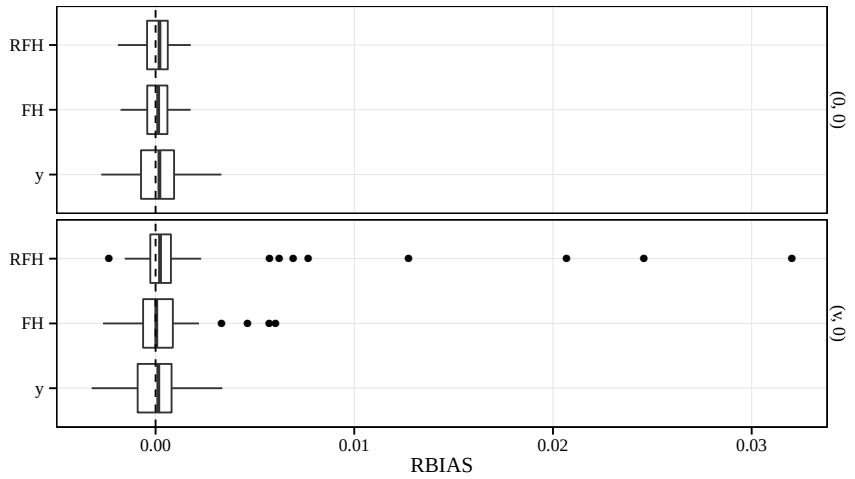

Figure 2: Boxplot with Relative Bias (RBIAS)

level population data. This allows for a number of interesting aspects which can be investigated and explained:

- 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 $e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$ where σ_e^2 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 ($\sigma_{e,i}^2 = \frac{\sigma_e^2}{n_i}$) 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.

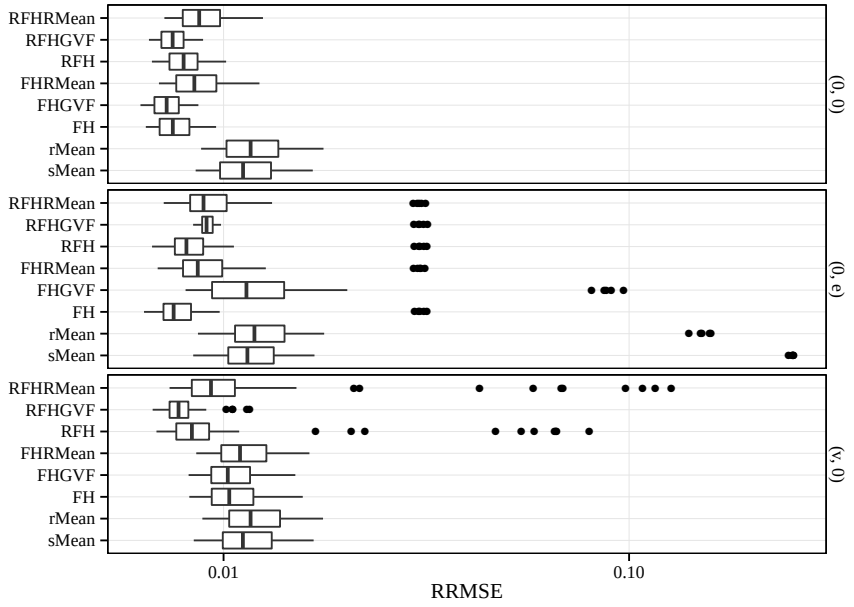

Figure 3: Boxplot with Relative Root Mean Squared Error (RRMSE)

9.2 DESIGN BASED SIMULATION STUDIES

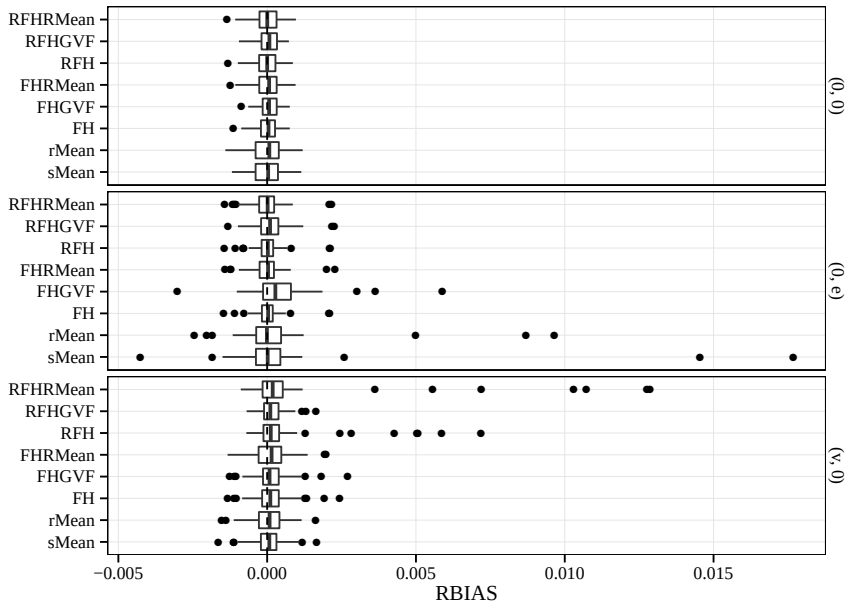

Figure 4: Boxplot with Relative Bias (RBIAS)

10 CASESTUDY

Part IV

APPENDIX

BIBLIOGRAPHY

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

Chambers, R., J. 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.

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.

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

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.

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

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