DISSERTATION

SEBASTIAN WARNHOLZ

Freie Universität Berlin

Robust Methods in Small Area Estimation:
Numerical Solutions and Their Implementations

November 18, 2015 – version 0.1

CONTENTS

Part I

THEORY

This is the chapter where I want to present the theoretical concepts underpinning the development of software and application. Most notably is the robust version of a Fay-Herriot Type model with different variance-covariance structures.

1 METHODS IN SMALL AREA ESTIMATION

1.1 UNIT-LEVEL MODELS

1.2 AREA LEVEL MODELS

1.2.1 *The Fay Herriot Model*

The model is introduced by Fay and Herriot (1979) and is used in small area estimation for research on area-level. It is build on a sampling model:

$$y_i = \mu_i + e_i,$$

where y_i is a direct estimator of a statistic of interest μ_i for an area i with $i = 1, \ldots, D$ and D being the total 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. $e_i|\mu_i \sim N(0, \sigma^2_{e,i})$. The model is modified with the linking model by assuming a linear relationship between the true area statistic μ_i and some auxiliary variables x_i:

$$\mu_i = x_i^\top \beta + \nu_i, \; i = 1, \ldots, D.$$

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 ν_i are assumed to be independent and normally distributed, i.e. $\nu_i \sim N(0, \sigma^2_\nu)$ furthermore e_i and ν_i are assumed to be independent. Combining the sampling and linking model leads to:

$$y_i = x_i^\top \beta + \nu_i + e_i. \tag{1}$$

1.2.2 *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:

$$y_{ij} = x_{ij}^\top \beta + \nu_i + e_{ij}\ ,$$

where y_{ij} is the response in domain i of unit j with $j = 1, \ldots, n_i$, where n_i is the number of units in domain i. v_i is an area specific random effect following (i.i.d.) a normal distribution with zero mean and σ_v^2 as variance parameter. e_{ij} is the remaining deviation from the model, following (i.i.d.) a normal distribution with zero mean and σ_e^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 $y_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$, for which an area level model can be derived as:

$$y_i = x_i^\top \beta + v_i + e_i$$

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

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

$$\mu_{dt} = x_{dt}^\top \beta + u_{1d} + u_{2dt},$$

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 $\epsilon_{1d} \sim N(0, \sigma_1^2)$ are i.i.d. with $d = 1, \ldots, D$. $w_{d,l}$ 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 $d = 1, \ldots, D$ and $t = 1, \ldots, T$. 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 $x_{dt}^\top$ 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_{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. $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V}_e)$ where $\mathbf{V}_e$ is a diagonal matrix with the known σ_{dt}^2 on the main diagonal. $\mathbf{u} \sim N(\mathbf{0}, \mathbf{V}_u(\theta))$ 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 $\theta = (\sigma_1^2, \rho_1, \sigma_2^2, \rho_2)$.

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

and follows from the SAR(1) process in the model errors. $\Omega_2(\rho_2)$ 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):

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

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

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.

2 REVIEW OF ROBUST METHODS IN SMALL AREA ESTIMATION

2.1 ROBUST ML SCORE FUNCTIONS

In 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:

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

where $\mathbf{y}$ is the vector of response variables y_i, $\mathbf{X}$ the design matrix, $\mathbf{v}$ the vector of random effects and $\mathbf{e}$ the vector of sampling errors. Both error components are assumed to be normally distributed with $\mathbf{u} \sim N(0, \mathbf{G})$ and $\mathbf{e} \sim N(0, \mathbf{R})$ where $\mathbf{G}$ and $\mathbf{R}$ typically depend on some variance parameters θ. Thus the variance of y is given by $\mathbb{V}(\mathbf{y}) = \mathbf{V}(\theta) = \mathbf{Z}\mathbf{G}\mathbf{Z}^\top + \mathbf{R}$. Maximizing the likelihood of $\mathbf{y}$ with respect to β and θ leads to the following equations:

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

where q denotes the number of unknown variance parameters. 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 $\text{␣}(\mathbf{r}) = (\psi(r_1), \ldots, \psi(r_n))$. Solving the following robust ML-equations leads to robustified estimators for β and θ:

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

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

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.

2.2 MEAN SQUARED ERROR

2.2.1 *pseudo linear FH*

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

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

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

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

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:

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

we can restate it similarly to the above as:

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

with

$$w_{s,i}^{\top} = x_i^{\top}\mathbf{A} + \mathbf{B}$$

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:

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

2.3 NEW SECTION

- item
  - item 1.1
  - item 1.2
  - new item

$$x_i = y_i \tag{6}$$
$$x_i + y_i = y_i \tag{7}$$

cite me: Abberger (1997)
Reference math: (6)
Math without numbering:

$$x_i = y_i$$
$$x_i + y_i = y_i$$

```
x <- 1
<<<<<<< HEAD
plot(1:10)
=======
```

```
ggplot(data.frame(x = 1:10, y = 11:20)) +
  geom_point(aes(x, y)) +
  theme_thesis()
>>>>>>> 202c45f790ff3fb900468c7f4559f74b835f5fa8
```

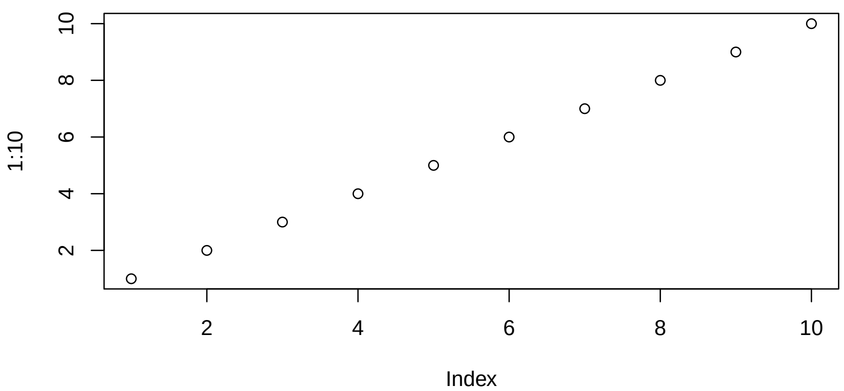

Figure 1: plot of chunk unnamed-chunk-2

3 ALGORITHMS

3.1 Review of numerical methods for non-linear optimizations in statistics and related fields

3.2 Algorithms for robust estimators in statistics

3.3 Propositions

3.3.1 *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

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

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:

$$\frac{\partial\Phi}{\partial\theta_l} = 2\frac{\partial}{\partial\theta_l}\left(\psi(\mathbf{r})^\top\mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\right)\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r}) + \operatorname{tr}\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_l}K\right) \tag{8}$$

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 8 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 8 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):

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

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:

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

3.3.2 *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:

$$\Phi_l(\theta) = \psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} \mathbf{V}^{-1} \mathbf{U}^{\frac{1}{2}} \psi(\mathbf{r}) - \operatorname{tr}\left(K\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_l} (\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top)^{-1} (\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top) \right) = 0 \tag{9}$$

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

$$\begin{aligned}
\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top &= \mathbf{Z} \begin{pmatrix} \sigma_1^2 \Omega_1 & \mathbf{0}_{D\times DT} \\ \mathbf{0}_{DT\times D} & \sigma_2^2 \Omega_2 \end{pmatrix} \mathbf{Z}^\top \\
&= \mathbf{Z} \left[\sigma_1^2 \begin{pmatrix} \Omega_1 & \mathbf{0}_{D\times DT} \\ \mathbf{0}_{DT\times D} & \mathbf{0}_{DT\times DT} \end{pmatrix} + \sigma_2^2 \begin{pmatrix} \mathbf{0}_{D\times D} & \mathbf{0}_{D\times DT} \\ \mathbf{0}_{DT\times D} & \Omega_2 \end{pmatrix} \right] \mathbf{Z}^\top \\
&= \begin{pmatrix} \mathbf{Z}\bar{\Omega}_1\mathbf{Z}^\top & \mathbf{Z}\bar{\Omega}_2\mathbf{Z}^\top \end{pmatrix} \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \end{pmatrix}
\end{aligned}$$

Thus equation 12 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.

3.3.2.1 *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:

$$\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top = \mathbf{Z}\begin{pmatrix} \sigma_1^2\Omega_1 & \mathbf{0}_{D\times DT} \\ \mathbf{0}_{DT\times D} & \sigma_2^2\Omega_2 \end{pmatrix}\mathbf{Z}^\top \tag{10}$$

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

$$\begin{aligned}
\sigma_1^2\Omega_1(\rho_1) &= \sigma_1^2\Omega_1\Omega_1(I-\rho_1\mathbf{W})^\top(I-\rho_1\mathbf{W}) \\
&= \sigma_1^2\left(\Omega_1\Omega_1 - \rho_1\Omega_1\Omega_1\mathbf{W}^\top - \rho_1\Omega_1\Omega_1\mathbf{W} + \rho_1^2\Omega_1\Omega_1\mathbf{W}^\top\mathbf{W}\right) \\
&= \sigma_1^2\left(\Omega_1\Omega_1 - \rho_1\Omega_1\Omega_1\mathbf{W}^\top\right) + \rho_1\left(-\sigma_1^2\Omega_1\Omega_1\mathbf{W} + \sigma_1^2\rho_1\Omega_1\Omega_1\mathbf{W}^\top\mathbf{W}\right)
\end{aligned} \tag{11}$$

Thus equation 10 can be rewritten as:

$$\begin{aligned}
\mathbf{Z}\mathbf{V}_u\mathbf{Z}^\top = \mathbf{Z}\Bigg[&\sigma_1^2\begin{pmatrix} \Omega_1\Omega_1 - \rho_1\Omega_1\Omega_1\mathbf{W}^\top & \mathbf{0}_{D\times DT} \\ \mathbf{0}_{DT\times D} & \mathbf{0}_{DT\times DT} \end{pmatrix} \\
&+ \rho_1\begin{pmatrix} -\sigma_1^2\Omega_1\Omega_1\mathbf{W} + \sigma_1^2\rho_1\Omega_1\Omega_1\mathbf{W}^\top\mathbf{W} & \mathbf{0}_{D\times DT} \\ \mathbf{0}_{DT\times D} & \mathbf{0}_{DT\times DT} \end{pmatrix} \\
&+ \sigma_2^2\begin{pmatrix} \mathbf{0}_{D\times D} & \mathbf{0}_{D\times DT} \\ \mathbf{0}_{DT\times D} & \Omega_2 \end{pmatrix}\Bigg]\mathbf{Z}^\top \\
= &\left(\mathbf{Z}\bar{\Omega}_{1,\sigma_1^2}\mathbf{Z}^\top \quad \mathbf{Z}\bar{\Omega}_{1,\rho_1}\mathbf{Z}^\top \quad \mathbf{Z}\bar{\Omega}_2\mathbf{Z}^\top\right)\begin{pmatrix} \sigma_1^2 \\ \rho_1 \\ \sigma_2^2 \end{pmatrix}
\end{aligned}$$

Thus equation 12 can be rewritten (analogously as above) to:

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

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

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

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

3.3.2.2 *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 $\theta = \sigma_v^2$ we can rewrite the equation such that:

$$\Phi(\theta) = \psi(\mathbf{r})^\top \mathbf{U}^{\frac{1}{2}}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta}\mathbf{V}^{-1}\mathbf{U}^{\frac{1}{2}}\psi(\mathbf{r}) - \mathrm{tr}\left(K\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta}(\mathbf{Z}\mathbf{G}\mathbf{Z}^\top)^{-1}(\mathbf{Z}\mathbf{G}\mathbf{Z}^\top)\right) = 0 \tag{12}$$

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 $\mathbf{Z}\mathbf{G}\mathbf{Z}^\top = \sigma_v^2\mathbf{I}$, where $\mathbf{I}$ is a $(D \times D)$ identity matrix. Furthermore $\frac{\partial \mathbf{V}}{\partial \theta} = \mathbf{I}$. Thus equation 12 can be rewritten to:

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

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

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

Part II

IMPLEMENTATION

This is the part where I want to introduce the software where the theoretical concepts find implementation.

4 VERIFICATION OF RESULTS

5 PERFORMANCE OF ALGORITHMS, A STATISTICIANS PERSPECTIVE

6 VALIDATION OF POINT ESTIMATES, A USERS PERSPECTIVE

7 SOFTWARE

Part III

RESULTS

This is the part where I will present all results. Most certainly they will contain a lot of model- and design-based simulation studies for various settings. Maybe there will be more data available and I can present some applications.

8 NUMERICAL ACCURACY

9 STABILITY

10 SPEED OF CONVERGENCE

11 COMPUTATIONAL COMPLEXITY

Part IV

APPENDIX

BIBLIOGRAPHY

Abberger, K. (1997). “Quantile smoothing in financial time series.” In: *Statistical Papers* 38 (2), pp. 125–148.

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.

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.