Universität Trier

Area-Level Small Area Estimation with Missing Values

Jan Pablo Burgard
Domingo Morales
Anna-Lena Wölwer

Research Papers in Economics
No. 14/19

Area-Level Small Area Estimation with Missing Values

Jan Pablo Burgard*

Research Institute for Offcial and Survey Statistics (RIFOSS), Trier University, Germany

Domingo Morales

Operations Research Center, University Miguel Hernández de Elche, Spain

Anna-Lena Wölwer

Research Institute for Offcial and Survey Statistics (RIFOSS), Trier University, Germany

September 20, 2019

Abstract

Model-based small area predictors are derived under the assumption that data files are complete. In application to real data, files may contain missing values. We introduce a variant of the bivariate Fay-Herriot model that takes into account for missing values in one component of the target variable and give fitting algorithms to estimate the model parameters. Based on the new model, we introduce empirical best predictors of domain means and derive an approximation to the mean squared error.

Keywords: Multivariate models, Fay-Herriot model, small area estimation, missing values.

*The authors gratefully acknowledge the Spanish grant PGC2018-096840-B-I00 and by the grant "Algorithmic Optimization (ALOP) - graduate school 2126" funded by the German Research Foundation

1 Introduction

Fay and Herriot (1979) considered an area-level linear mixed model and derived an empirical best linear unbiased predictor (EBLUP) of a domain quantity that it is also called Fay-Herriot (FH) predictor, or FH-EBLUP. It is widely applied in the context of small area estimation (SAE). The idea of the FH predictor is to improve the precision of a direct estimator on a domain of interest by borrowing strength from other domains. This method, however, requires that for every statistic of interest there is a direct estimate available in all the domains of the study. In the common case of unplanned sample sizes within domain, this cannot be guaranteed. Therefore, it appears that for some domains direct estimates may be missing by chance. Furthermore, estimates in domains or table cells can be suppressed for confidentiality reasons, e.g. due to small cell sizes or high sampling errors as it was done in Zayatz (2007).

Since the publication of the FH predictor, many extensions were made to allow for different practical problems. Inter alia, Prasad and Rao (1990) and Datta and Lahiri (2000) proposed mean squared error (MSE) estimators for the FH predictor, Li and Lahiri (2010) and Yoshimori and Lahiri (2014) introduced new adjusted maximum likelihood fitting methods, Ybarra and Lohr (2008), Arima et al. (2017), Burgard et al. (2019a), and Burgard et al. (2019b) studied the effect of measurement errors in the covariates, Pratesi and Salvati (2008), González-Manteiga et al. (2010), Articus and Burgard (2014) and Morales et al. (2015) allow for a heterogeneous dependency structure in the FH model, Datta et al. (1996), González-Manteiga et al. (2008), Porter et al. (2015) and Benavent and Morales (2016) investigated and applied multivariate FH models, Esteban et al. (2012) and Marhuenda et al. (2013) estimated small area poverty proportions under temporal and spatio-temporal Fay-Herriot models respectively. Many other authors have studied further variants of the Fay-Herriot model adapted to different setups. However, for the problem of missing values in the dependent variable, to our knowledge, there is no approach giving an empirical best predictor (EBP) also for the domains with missings.

This manuscript introduces an EBP for settings where direct estimates are partially missing. Furthermore, it derives an approximation to the MSE of the EBP and the corresponding MSE estimator and it illustrates the potential benefits of the proposed approach under different simulated data scenarios. In addition, the manuscript presents an application of the developed methodology to a SAE problem with publicly available county-level data from the U.S. American Community Survey (ACS).

The manuscript is structured as follows: Section 2 introduces the bivariate Fay-Herriot model, which is the basis for the development of the EBP theory when part of the values of the target variables are missing. Section 3 divides the set of domains in three groups depending on the existence or not of missing values in each of the dependent variables and gives the corresponding EBPs. Section 4 gives algorithms for calculating the maximum likelihood and the residual maximum likelihood estimators of the model parameters. Section 5 derives an approximation to the mean squared error of the best predictor and proposes an explicit-formula estimator. Section 6 presents a parametric bootstrap procedure for estimating the mean squared error.

2 The bivariate Fay-Herriot model

Let U be a finite population partitioned into D domains $U_1, \dots, U_D$. Let $\mu_d = (\mu_{d1}, \mu_{d2})'$ be a vector of characteristics of interest in the domain d and let $y_d = (y_{d1}, y_{d2})'$ be a vector of direct estimates of μ_d calculated by using the data of the target survey sample.

The bivariate Fay-Herriot model is defined in two stages. The first stage indicates that direct estimators $\{y_d\}$ are unbiased and follow the *sampling model*

$$y_d = \mu_d + e_d, \quad \forall d \in \{1,\dots,D\}, \tag{1}$$

where the vectors $e_d = (e_{d1}, e_{d2})' \sim N_2(0, V_{ed})$ are independent and the 2×2 covariance matrices V_{ed} are known. In most cases, V_{ed} is taken to be the design-based covariance matrix of direct estimators y_d, $\forall d \in \{1,\dots,D\}$. The covariance matrices V_{ed} are

$$V_{ed} = \begin{pmatrix} \sigma^2_{ed1} & \sigma_{ed12} \\ \sigma_{ed12} & \sigma^2_{ed2} \end{pmatrix}, \quad \sigma_{ed12} = \rho_{ed12}\sigma_{ed1}\sigma_{ed2}, \quad \forall d \in \{1,\dots,D\}.$$

In the second stage the true area characteristic μ_{dk} is assumed to be linearly related to p_k explanatory variables, $k = 1, 2$, $d \in \{1,\dots,D\}$. Let $x'_{dk} = (x_{dk1},\dots,x_{dkp_k})$ be a row vector containing the true aggregated (population) values of p_k explanatory variables for μ_{dk} and let $X_d = \text{diag}(x'_{d1}, x'_{d2})$ be a $2 \times p$ block-diagonal matrix with $p = p_1 + p_2$. Let $\beta_k = (\beta_{k1},\dots,\beta_{kp_k})'$ be a column vector of size p_k containing the regression parameters β_{kj} for μ_{dk} and let $\beta = (\beta'_1, \beta'_2)'_{p\times 1}$. The *linking model* is

$$\mu_d = X_d\beta + u_d, \quad u_d = (u_{d1}, u_{d2})' \sim N_2(0, V_{ud}), \quad \forall d \in \{1,\dots,D\}, \tag{2}$$

where the vectors u_d's are independent of the vectors e_d's. The 2×2 covariance matrix V_{ud} depends on three unknown parameters, $\theta_1 = \sigma^2_{u1}$, $\theta_2 = \sigma^2_{u2}$ and $\theta_3 = \rho$, i.e.

$$V_{ud} = \begin{pmatrix} \sigma^2_{u1} & \rho\sigma_{u1}\sigma_{u2} \\ \rho\sigma_{u1}\sigma_{u2} & \sigma^2_{u2} \end{pmatrix}, \quad \forall d \in \{1,\dots,D\}.$$

The bivariate Fay-Herriot (BFH) model can be expressed as a single model in the form

$$y_d = X_d\beta + u_d + e_d, \quad \forall d \in \{1,\dots,D\}, \tag{3}$$

or in the matrix form

$$y = X\beta + u + e,$$

where

$$y = \underset{1\leq d\leq D}{\text{col}}(y_d),\ u = \underset{1\leq d\leq D}{\text{col}}(u_d),\ e = \underset{1\leq d\leq D}{\text{col}}(e_d),\ X = \underset{1\leq d\leq D}{\text{col}}(X_d).$$

We finally assume that u_d, e_d, $d \in \{1, \dots, D\}$, are independent. The BFH model (3) is a reparametrization of Model 3 introduced by Benavent and Morales (2016).

Let us define $V_d = V_{ud} + V_{ed}$, $\forall d \in \{1, \dots, D\}$. Under model (3), it holds that

$$E\,(y) = X\beta \quad \text{and} \quad V = \text{var}(y) = Z'V_uZ + V_e = V_u + V_e = \underset{1\leq d\leq D}{\text{diag}}\,(V_d).$$

3 Prediction with missing target values

Let us assume that some of the y_{dk} are missing. We define $y_{\bar{d}1} = (y_{d1}, 0)'$ and $y_{\bar{d}2} = (0, y_{d2})'$, and partition the domains into three groups:

$\mathbb{D}_1 = \{d \in \mathbb{N} : 1 \leq d \leq D_1\}$ contains the D_1 domains where only y_{d1} is observed.

$\mathbb{D}_2 = \{d \in \mathbb{N} : D_1 + 1 \leq d \leq D_1 + D_2\}$ contains the D_2 domains where only y_{d2} is observed.

$\mathbb{D}_3 = \{d \in \mathbb{N} : D_1 + D_2 + 1 \leq d \leq D\}$ contains the remaining domains where $y_d = (y_{d1}, y_{d2})'$ is fully observed.

If the BFH model (3) holds for $d \in \{1, \dots, D\}$ and the missing data obey scheme $\{1, \dots, D\} = \mathbb{D}_1 \cup \mathbb{D}_2 \cup \mathbb{D}_3$, we say that target vectors y_d obey a missing data BFH (MBFH) model. If the MBFH model holds, then

1. $y_{d1} \sim N_1\left(x'_{d1}\beta_1, \sigma^2_{u1} + \sigma^2_{ed1}\right)$ and $y_{d1}|_{u_d} \sim N_1\left(x'_{d1}\beta_1 + u_{d1}, \sigma^2_{ed1}\right)$ if $d \in \mathbb{D}_1$,
2. $y_{d2} \sim N_1\left(x'_{d2}\beta, \sigma^2_{u2} + \sigma^2_{ed2}\right)$ and $y_{d2}|_{u_d} \sim N_1\left(x'_{d2}\beta_2 + u_{d2}, \sigma^2_{ed2}\right)$ if $d \in \mathbb{D}_2$, and
3. $y_d \sim N_2\left(X_d\beta, V_{ud} + V_{ed}\right)$ and $y_d|_{u_d} \sim N_2\left(X_d\beta + u_d, V_{ed}\right)$ if $d \in \mathbb{D}_3$.

In a real situation where the target data follows a MBFH model, the BFH model is strictly applicable to $\mathbb{D}_3$, but not to $\mathbb{D}_1$ or $\mathbb{D}_2$. For example, under the BFH model we can only calculate EBLUPs of μ_d or u_d for $d \in \mathbb{D}_3$. However, in what follows we show that it is possible calculate EBPs for $d \in \mathbb{D}_1 \cup \mathbb{D}_2$ under the MBFH model.

As the kernel of the n-variate normal distribution is

$$f(y|\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right] \propto \exp\left[-\frac{1}{2}y'\Sigma^{-1}y + \mu'\Sigma^{-1}y\right],$$

we have the following three propositions.

Proposition 3.1. If $d \in \mathbb{D}_1$, then the best predictor (BP) of u_d under the MBFH model is

$$\hat{u}_d^{bp} = E[u_d|y_{d1}] = \Phi_{d1}\begin{pmatrix} \sigma^{-2}_{ed1} & 0 \\ 0 & 0 \end{pmatrix}(y_{\bar{d}1} - X_d\beta), \quad \Phi_{d1} = \left[\begin{pmatrix} \sigma^{-2}_{ed1} & 0 \\ 0 & 0 \end{pmatrix} + V^{-1}_{ud}\right]^{-1}.$$

Proof. The conditional distribution of u_d, given y_{d1}, is

$$\begin{aligned} f(u_d|y_{d1}) &\propto f(y_{d1}|u_d)f(u_d) \\ &\propto \exp\left\{-\frac{1}{2}u'_d\left[\begin{pmatrix} \sigma^{-2}_{ed1} & 0 \\ 0 & 0 \end{pmatrix} + V^{-1}_{ud}\right]u_d + u'_d\Phi^{-1}_{d1}\left[\Phi_{d1}\begin{pmatrix} \sigma^{-2}_{ed1} & 0 \\ 0 & 0 \end{pmatrix}(y_{\bar{d}1} - X_d\beta)\right]\right\}. \end{aligned}$$

Therefore, $f(u_d|y_{d_1})$ is a bivariate normal distribution with parameters

$$\text{var}(u_d|y_{d_1}) = \left[\begin{pmatrix} \sigma^{-2}_{ed1} & 0 \\ 0 & 0 \end{pmatrix} + V^{-1}_{ud}\right]^{-1} = \Phi_{d1}, \quad E[u_d|y_{d_1}] = \Phi_{d1}\begin{pmatrix} \sigma^{-2}_{ed1} & 0 \\ 0 & 0 \end{pmatrix}(y_{\bar{d}1} - X_d\beta).$$

$\square$

Proposition 3.2. If $d \in \mathbb{D}_2$, then the BP of u_d under the MBFH model is

$$\hat{u}_d^{bp} = E[u_d|y_{d2}] = \Phi_{d2} \begin{pmatrix} 0 & 0 \\ 0 & \sigma_{ed2}^{-2} \end{pmatrix} (y_{\bar{d}2} - X_d\beta), \quad \Phi_{d2} = \left(\begin{pmatrix} 0 & 0 \\ 0 & \sigma_{ed2}^{-2} \end{pmatrix} + V_{ud}^{-1} \right)^{-1}.$$

Proof. The proof is analogous as the one of Proposition 3.1.

Proposition 3.3. If $d \in \mathbb{D}_3$, then BP of u_d under the MBFH model is

$$\hat{u}_d^{bp} = E[u_d|y_d] = \Phi_d V_{ed}^{-1}(y_d - X_d\beta), \quad \Phi_d = \left(V_{ed}^{-1} + V_{ud}^{-1}\right)^{-1}.$$

Proof. The conditional distribution of u_d given y_d, is

$$\begin{aligned} f(u_d|y_d) &\propto f(y_d|u_d)f(u_d) \\ &\propto \exp\left\{-\frac{1}{2}u_d'\left(V_{ed}^{-1} + V_{ud}^{-1}\right)u_d + u_d'\Phi_d^{-1}\left[\Phi_d V_{ed}^{-1}(y_d - X_d\beta)\right]\right\}. \end{aligned}$$

Therefore, $f(u_d|y_d)$ is a bivariate normal distribution with parameters

$$\text{var}(u_d|y_d) = \left(V_{ed}^{-1} + V_{ud}^{-1}\right)^{-1} = \Phi_d, \quad E[u_d|y_d] = \Phi_d V_{ed}^{-1}(y_d - X_d\beta).$$

□

Corollary 3.3. The BP of μ_d, $d = 1, \ldots, D$, under the MBFH model is

$$\hat{\mu}_d^{bp} = X_d\beta + \hat{u}_d^{bp}. \tag{4}$$

Definition 3.1. The EBP of μ_d, $d = 1, \ldots, D$, under the MBFH model (MBFH-EBP) is obtained from formula (4) by plugging estimators $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}_{u1}^2$, $\hat{\sigma}_{u2}^2$ and $\hat{\rho}$ in the places of $\boldsymbol{\beta}$, σ_{u1}^2, σ_{u2}^2 and ρ respectively, i.e.

$$\hat{\mu}_d^{ebp} = X_d\hat{\beta} + \hat{u}_d^{ebp}. \tag{5}$$

4 Estimation of model parameters

This section presents the maximum likelihood (ML) and the residual maximum likelihood (REML) methods for estimating the model parameters.

4.1 Maximum likelihood method

The vector of model parameters is $\psi = (\beta', \theta')'$, where $\theta = (\theta_1, \theta_2, \theta_3)' = (\sigma_{u1}^2, \sigma_{u2}^2, \rho)'$. The log-likelihood is $l = \sum_{d=1}^D l_d$, where

$$
\begin{aligned}
l_d &= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log(\sigma_{u1}^2 + \sigma_{ed1}^2) - \frac{1}{2(\sigma_{u1}^2 + \sigma_{ed1}^2)}(y_{d1} - x_{d1}'\beta_1)^2, \quad \text{if } d \in \mathbb{D}_1,\\
l_d &= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log(\sigma_{u2}^2 + \sigma_{ed2}^2) - \frac{1}{2(\sigma_{u2}^2 + \sigma_{ed2}^2)}(y_{d2} - x_{d2}'\beta_2)^2, \quad \text{if } d \in \mathbb{D}_2,\\
l_d &= -\log 2\pi - \frac{1}{2}\log|V_d| - \frac{1}{2}(y_d - X_d\beta)'V_d^{-1}(y_d - X_d\beta), \qquad \text{if } d \in \mathbb{D}_3.
\end{aligned}
$$

The ML Fisher-scoring algorithm, with Fisher-information matrices $F_{\beta\beta}(\theta^{(r)}), F_{\theta\theta}(\theta^{(r)})$ and score vectors $U_\beta(\psi^{(r)}), U_\theta(\psi^{(r)})$, is

1. Set the initial values $\psi^{(0)} = (\beta_1^{(0)}, \beta_2^{(0)}, \theta^{(0)})$ and $\varepsilon > 0$.
2. Repeat the following steps until the tolerance or the boundary conditions are fulfilled.

   (a) Updating equations: Do

   $$\beta^{(r+1)} = \beta^{(r)} + F_{\beta\beta}^{-1}(\theta^{(r)})U_\beta(\psi^{(r)}), \quad \theta^{(r+1)} = \theta^{(r)} + F_{\theta\theta}^{-1}(\theta^{(r)})U_\theta(\psi^{(r)}).$$

   (b) Boundary condition: If $\theta_1^{(r+1)} > 0$, $\theta_2^{(r+1)} > 0$ and $\left|\theta_3^{(r+1)}\right| < 1$, continue. Otherwise, do $\hat{\psi} = \psi^{(r)}$ and stop.

   (c) Tolerance condition: If $\left|\psi_j^{(r+1)} - \psi_j^{(r)}\right| < \varepsilon$, $j = 1, \ldots, p+3$, do $\hat{\psi} = \psi^{(r+1)}$ and stop. Otherwise, continue.

3. Output: $\hat{\psi}$.

Remark. As starting values, we take $\beta_1^{(0)} = \hat{\beta}_1^{(0)}$, $\beta_2^{(0)} = \hat{\beta}_2^{(0)}$, $\hat{\theta}_{3,0} = 0$, $\hat{\theta}_{k,0} = \hat{\sigma}_{uk,0}^2$, $k = 1, 2$, where $\hat{\beta}_1^{(0)}$, $\hat{\beta}_2^{(0)}$ and $\hat{\sigma}_{uk,0}^2$ are the REML or the ML estimators of the corresponding bivariate Fay-Herriot that uses only the data of group $\mathbb{D}_3$.

4.2 Residual maximum likelihood method

In the case that there are no missing values, the residual maximum likelihood (REML) log-likelihood is

$$l_{reml}(\theta) = -\frac{2D-p}{2}\log 2\pi + \frac{1}{2}\log|X'X| - \frac{1}{2}\log|V| - \frac{1}{2}\log|X'V^{-1}X| - \frac{1}{2}y'Py, \qquad (6)$$

where $\theta = (\theta_1, \theta_2, \theta_3)$, $\theta_1 = \sigma_{u1}^2$, $\theta_2 = \sigma_{u2}^2$, $\theta_3 = \rho$, $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$, $PVP = P$ with $PX = 0$, y is the $2D$-vector $y = (y_1', \dots, y_D')'$, and X the corresponding $2D \times p$-matrix $X = (X_1, \dots, X_D)'$. P is defined as before on the reduced X and V.

In the case that there are missing values, the residual maximum likelihood (REML) log-likelihood is

$$\begin{aligned} l_{reml}(\theta) = & -\frac{D_1 + D_2 + 2(D - D_1 - D_2) - p}{2}\log 2\pi + \frac{1}{2}\log|X'X| \\ & -\frac{1}{2}\log|V| - \frac{1}{2}\log|X'V^{-1}X| - \frac{1}{2}y'Py, \end{aligned} \qquad (7)$$

where now, the vector y and the matrices X and V are reduced to those rows and columns that correspond to an observed value of y_{dk}, $\forall d \in \mathbb{D}$ and $k = 1, 2$. Let $\tilde{D} := D_1 + D_2 + 2(D - D_1 - D_2)$, then the length of y is $\tilde{D}$ and the dimensions of X and V are $\tilde{D} \times p$ and $\tilde{D} \times \tilde{D}$ respectively.

The REML Fisher-scoring algorithm, with Fisher-information matrix $F(\theta^{(k)})$ and score vector $S(\theta^{(k)})$, is

1. Set the initial values $\beta^{(0)}$, $\theta^{(0)}$, and $\varepsilon_k > 0$, $r = 1, \ldots, p+3$.
2. Repeat the following steps until the tolerance or the boundary conditions are fulfilled.
   (a) Updating equation for θ: Do $\theta^{(k+1)} = \theta^{(k)} + F^{-1}(\theta^{(k)})S(\theta^{(k)})$.
   (b) Boundary condition: If $\theta_1^{(k+1)} > 0$, $\theta_2^{(k+1)} > 0$ and $\left|\theta_3^{(k+1)}\right| < 1$, continue. Otherwise, do $\hat{\theta} = \theta^{(k)}$ and stop.
   (c) Updating equation for β: Do $\beta^{(k+1)} = \left(X'V^{-1}(\theta^{(k+1)})X\right)^{-1} X'V^{-1}(\theta^{(k+1)})y$.
   (d) Tolerance condition: If $\left|\theta_\ell^{(k+1)} - \theta_\ell^{(k)}\right| < \varepsilon_{p+\ell}$, $\left|\beta_k^{(k+1)} - \beta_k^{(k)}\right| < \varepsilon_k$, $r = 1, \ldots, p$, $\ell = 1, 2, 3$, do $\hat{\theta}_\ell = \theta_\ell^{(k+1)}$, $\hat{\beta} = \beta^{(k+1)}$ and stop. Otherwise, continue.
3. Output: $\hat{\theta}$, $\hat{\beta}$, $F^{-1}(\hat{\theta})$, $\left(X'V^{-1}(\hat{\theta})X\right)^{-1}$.

The asymptotic distributions of the REML estimators $\hat{\theta}$ and $\hat{\beta}$,

$$\hat{\theta} \sim N_3(\theta, F^{-1}(\theta)), \quad \hat{\beta} \sim N_k(\beta, (X'V^{-1}(\theta)X)^{-1}),$$

can be used to construct $(1-\alpha)$-level confidence intervals for the components θ_ℓ of θ and β_i of β. Let β_i denote the i-th component of the vector β, not the vector of regression parameters of the i-th component. The confidence intervals are given by

$$\hat{\theta}_\ell \pm z_{\alpha/2}\, \nu_{\ell\ell}^{1/2},\ \ell = 1, 2, 3, \quad \hat{\beta}_i \pm z_{\alpha/2}\, q_{ii}^{1/2},\ i = 1, \ldots, r, \tag{8}$$

where $F^{-1}(\hat{\theta}) = (\nu_{ab})_{a,b=1,2,3}$, $(X'V^{-1}(\hat{\theta})X)^{-1} = (q_{ij})_{i,j=1,\ldots,k}$ and z_α is the α-quantile of the $N(0,1)$ distribution. For $\hat{\beta}_i = \beta_0$, the p-value for testing the hypothesis $H_0:\ \beta_i = 0$ is

$$p\text{-value} = 2P_{H_0}(\hat{\beta}_i > |\beta_0|) = 2P(N(0,1) > |\beta_0|/\sqrt{q_{ii}}\,). \tag{9}$$

5 Analytic approximation of the mean squared errors

5.1 Best predictors

Let us first consider the group $\mathbb{D}_1$. We use the notation

$$\Phi_{d1} = \left(A_{d1} + V_{ud}^{-1}\right)^{-1} \triangleq \begin{pmatrix} \phi_{d1,11} & \phi_{d1,12} \\ \phi_{d1,12} & \phi_{d1,22} \end{pmatrix}, \quad A_{d1} = \begin{pmatrix} \sigma_{ed1}^{-2} & 0 \\ 0 & 0 \end{pmatrix}.$$

As $\hat{\mu}_d^{bp} - \mu_d = \hat{u}_d^{bp} - u_d$, the MSE matrix of $\hat{\mu}_d^{bp}$ is

$$\begin{aligned} \mathrm{MSE}(\hat{\mu}_d^{bp}) &= \mathrm{MSE}(\hat{u}_d^{bp}) = E\left[(\hat{u}_d^{bp} - u_d)(\hat{u}_d^{bp} - u_d)'\right] \\ &= \Phi_{d1} A_{d1} (V_{ud} + V_{ed}) A_{d1} \Phi_{d1} + V_{ud} - 2\Phi_{d1} A_{d1} V_{ud}. \end{aligned}$$

For the group $\mathbb{D}_2$, we have similar mathematical derivations. For the sake of brevity, we omit them.

Let us consider the group $\mathbb{D}_3$. As $\hat{\mu}_d^{bp} - \mu_d = \hat{u}_d^{bp} - u_d$, the MSE matrix of $\hat{\mu}_d^{bp}$ is

$$\begin{aligned} \mathrm{MSE}(\hat{\mu}_d^{bp}) &= \mathrm{MSE}(\hat{u}_d^{bp}) = E\left[(\hat{u}_d^{bp} - u_d)(\hat{u}_d^{bp} - u_d)'\right] \\ &= \Phi_d V_{ed}^{-1} (V_{ud} + V_{ed}) V_{ed}^{-1} \Phi_d + V_{ud} - 2\Phi_d V_{ed}^{-1} V_{ud}. \end{aligned}$$

5.2 Empirical best predictors

We sketch the derivation for the empirical best predictors for group $\mathbb{D}_1$. The mathematical derivations for group $\mathbb{D}_2$ and $\mathbb{D}_2$ are mostly analogous.

For $d \in \mathbb{D}_1$, we have the following approximation to $MSE(\hat{\mu}_d^{ebp})$.

$$MSE(\hat{\mu}_d^{ebp}) = G_{d1}(\theta) + G_{d2}(\theta) + G_{d3}(\theta) + O_{2\times 2}(D^{-1}),$$

where $G_{d2}(\theta) = G_{d2,11}(\theta) + G_{d2,22}(\theta) + G_{d2,12}(\theta) + G'_{d2,12}(\theta)$ and

$$\begin{aligned}
G_{d1}(\theta) &= \Phi_{d1}(\theta)A_{d1}(\theta)(V_{ud}(\theta)+V_{ed}(\theta))A_{d1}(\theta)\Phi_{d1}(\theta)+V_{ud}(\theta)-2\Phi_{d1}(\theta)A_{d1}(\theta)V_{ud}(\theta),\\
G_{d2,ab}(\theta) &= \begin{pmatrix} \mathrm{tr}\{H_{d\beta_b\beta_a,11}(\theta)\,\mathrm{cov}(\hat\beta_a,\hat\beta_b)\} & \mathrm{tr}\{H_{d\beta_b\beta_a,21}(\theta)\,\mathrm{cov}(\hat\beta_a,\hat\beta_b)\} \\ \mathrm{tr}\{H_{d\beta_b\beta_a,12}(\theta)\,\mathrm{cov}(\hat\beta_a,\hat\beta_b)\} & \mathrm{tr}\{H_{d\beta_b\beta_a,22}(\theta)\,\mathrm{cov}(\hat\beta_a,\hat\beta_b)\} \end{pmatrix},\ a,b=1,2,\\
G_{d3}(\theta) &= \frac{\sigma_{ud1}^2+\sigma_{ed1}^2}{\sigma_{ed1}^4}\begin{pmatrix} \mathrm{tr}\{G_{d\theta\theta,11}(\theta)\,\mathrm{var}(\hat\theta)\} & \mathrm{tr}\{G_{d\theta\theta,21}(\theta)\,\mathrm{var}(\hat\theta)\} \\ \mathrm{tr}\{G_{d\theta\theta,12}(\theta)\,\mathrm{var}(\hat\theta)\} & \mathrm{tr}\{G_{d\theta\theta,22}(\theta)\,\mathrm{var}(\hat\theta)\} \end{pmatrix}.
\end{aligned}$$

The remaining vectors and matrices are derived as follows. The EBP is a function of the estimators $(\hat\beta,\hat\theta)$ and of the target variable y_{d1}. For the sake of brevity, we write

$$h_d(\hat\beta,\hat\theta) \triangleq \hat\mu_d^{ebp} = X_d\hat\beta + \hat\Phi_{d1}A_{d1}(y_{\bar d1} - X_d\hat\beta),$$

$$\text{where}\quad \hat\Phi_{d1} = \Phi_{d1}(\hat\theta) = \left(A_{d1}+\hat V_{ud}^{-1}\right)^{-1},\quad \hat V_{ud} = V_{ud}(\hat\theta) = \begin{pmatrix} \hat\sigma_{u1}^2 & \hat\rho\hat\sigma_{u1}\hat\sigma_{u2} \\ \hat\rho\hat\sigma_{u1}\hat\sigma_{u2} & \hat\sigma_{u2}^2 \end{pmatrix}.$$

The derivatives of matrix $\Phi_{d1}(\theta)$ with respect to θ_ℓ, $\ell=1,2,3$, are

$$\frac{\partial\Phi_{d1}}{\partial\theta_\ell} = \left(A_{d1}+V_{ud}^{-1}\right)^{-1}V_{ud}^{-1}V_{ud\ell}V_{ud}^{-1}\left(A_{d1}+V_{ud}^{-1}\right)^{-1} = \begin{pmatrix} \phi_{d1\ell,11} & \phi_{d1\ell,12} \\ \phi_{d1\ell,12} & \phi_{d1\ell,22} \end{pmatrix}.$$

The derivatives of $h_d(\beta,\theta)$ with respect to β_{kj} and θ_ℓ, $k=1,2$, $j=1,\ldots,p_k$, $\ell=1,2,3$, are

$$\begin{aligned}
\frac{\partial h_d}{\partial\beta_{1j}} &= \begin{pmatrix} x_{d1j} \\ 0 \end{pmatrix} - \frac{x_{d1j}}{\sigma_{ed1}^2}\begin{pmatrix} \phi_{d1,11} \\ \phi_{d1,12} \end{pmatrix} \triangleq \begin{pmatrix} h_{d\beta_{1j},1} \\ h_{d\beta_{1j},2} \end{pmatrix},\quad \frac{\partial h_d}{\partial\beta_{2j}} = \begin{pmatrix} 0 \\ x_{d2j} \end{pmatrix} \triangleq \begin{pmatrix} h_{d\beta_{2j},1} \\ h_{d\beta_{2j},2} \end{pmatrix},\\
\frac{\partial h_d}{\partial\theta_\ell} &= \frac{\partial\Phi_{d1}}{\partial\theta_\ell}A_{d1}(y_{\bar d1}-X_d\beta) = \frac{y_{d1}-x'_{d1}\beta_1}{\sigma_{ed1}^2}\begin{pmatrix} \phi_{d1\ell,11} \\ \phi_{d1\ell,12} \end{pmatrix} \triangleq \frac{y_{d1}-x'_{d1}\beta_1}{\sigma_{ed1}^2}\begin{pmatrix} g_{d\theta_\ell,1} \\ g_{d\theta_\ell,2} \end{pmatrix}.
\end{aligned}$$

The 3×1 vectors containing the derivatives with respect to θ are $g_{d\theta,1} = \underset{1\le\ell\le3}{\mathrm{col}}\left(g_{d\theta_\ell,1}\right)$, $g_{d\theta,2} = \underset{1\le\ell\le3}{\mathrm{col}}\left(g_{d\theta_\ell,2}\right)$ and the corresponding 3×3 matrices are $G_{d\theta\theta,ab} = g_{d\theta,a}g'_{d\theta,b}$, $a,b=1,2$.

The $p_k \times 1$ vectors containing the derivatives with respect to β_k, $k = 1, 2$, are

$$h_{d\beta_k,1} = \underset{1\le j\le p_k}{\mathrm{col}}\big(h_{d\beta_{kj},1}\big), \quad h_{d\beta_k,2} = \underset{1\le j\le p_k}{\mathrm{col}}\big(h_{d\beta_{kj},2}\big),$$

and the corresponding $p_{k_1} \times p_{k_2}$ matrices are $H_{d\beta_{k_1}\beta_{k_2},ab} = h_{d\beta_{k_1},a} h'_{d\beta_{k_2},b}$, $k_1, k_2, a, b = 1, 2$.

An estimator of $MSE(\hat{\mu}_d^{ebp})$ is

$$mse(\hat{\mu}_d^{ebp}) = G_{d1}(\hat{\theta}) + G_{d2}(\hat{\theta}) + 2G_{d3}(\hat{\theta}).$$

6 Bootstrap approximations of the mean squared errors

This section introduces a parametric bootstrap procedure for approximating $MSE(\hat{\mu}_d^{ebp})$. Steps B1–B5 below describe the basic procedure for computing an approximation of $MSE(\hat{\mu}_d^{ebp})$ called *direct* parametric bootstrap estimator.

Parametric bootstrap procedure:

B1. Calculate the REML (or ML) estimates $\hat{\theta}$ and $\hat{\beta}$ of θ and β respectively by using the observable data, i.e. (y_{d1}, X_d) $\forall d \in \mathbb{D}_1$, (y_{d2}, X_d) $\forall d \in \mathbb{D}_2$, and (y_d, X_d) $\forall d \in \mathbb{D}_3$.

B2. $\forall d \in \{1, \dots, D\}$, generate independent and identically distributed vectors $u_d^* \sim N_2(0, V_{ud}(\hat{\theta}))$.

B3. $\forall d \in \{1, \dots, D\}$, generate independent vectors $e_d^* \sim N_2(0, V_{ed})$.

B4. Construct the bootstrap model

$$y_d^* = X_d\hat{\beta} + u_d^* + e_d^*, \quad \forall d \in \{1, \dots, D\}.$$

For step B5 we introduce further notation. Let E_* and MSE_* denote the expectation and MSE under the probability distribution induced by bootstrap model B4, given the initial target vector y. The bootstrap mean vectors are

$$\mu_d^* = X_d\hat{\beta} + u_d^*, \quad \forall d \in \{1,\dots,D\}.$$

Let $\hat{\beta}^*$ and $\hat{\theta}^*$ be the REML (or ML) estimators of the parameters $\hat{\beta}$ and $\hat{\theta}$ of bootstrap model B4. These estimators are calculated by using only the observable bootstrap data (y_{d1}^*, X_d) if $d \in \mathbb{D}_1$, (y_{d2}^*, X_d) if $d \in \mathbb{D}_2$, and (y_d^*, X_d) if $d \in \mathbb{D}_3$.

Let $\hat{\mu}_d^{*bp}$ and $\hat{\mu}_d^{*ebp}$ be the BP and EBP of μ_d^* under model B4 $\forall d \in \{1,\dots,D\}$. In the same way, the bootstrap MSE of $\hat{\mu}_d^{*ebp}$ is

$$\mathrm{MSE}_*^1(\hat{\mu}_d^{*ebp}) = E_*[(\hat{\mu}_d^{*ebp} - \mu_d^*)(\hat{\mu}_d^{*ebp} - \mu_d^*)'], \quad \forall d \in \{1,\dots,D\}.$$

These 2×2 matrices are called parametric bootstrap estimators. In practice, these estimators can be approximated via Monte Carlo as described in B5.

B5. Generate B bootstrap vectors $y^{*(b)} = (y_d^{*(b)} : d \in \{1,\dots,D\})$, $b = 1,\dots,B$, from model B4. From each vector $y^{*(b)}$, calculate the true means $\mu_d^{*(b)}$ and their EBPs $\hat{\mu}_d^{*ebp(b)}$ by using only the observable bootstrap data. Then compute the *direct* bootstrap estimators

$$\mathrm{mse}^1(\hat{\mu}_d^{ebp}) = B^{-1}\sum_{b=1}^{B}(\hat{\mu}_d^{*ebp(b)} - \mu_d^{*(b)})(\hat{\mu}_d^{*ebp(b)} - \mu_d^{*(b)})', \quad \forall d \in \{1,\dots,D\}. \tag{10}$$

Observe that $\mathrm{mse}^1(\hat{\mu}_d^{*ebp})$ is consistent for $\mathrm{MSE}_*^1(\hat{\mu}_d^{*ebp})$ as $B \to \infty$.

We can also apply the bootstrap technique to approximate the terms G_2 and G_3 of $MSE(\hat{\mu}_d^{ebp})$. Following this idea, we define the *term-to-term* bootstrap estimator as

$$\mathrm{MSE}_*^2(\hat{\mu}_d^{*ebp}) = G_{d1}(\hat{\theta}) + E_*[(\hat{\mu}_d^{*ebp} - \hat{\mu}_d^{*bp})(\hat{\mu}_d^{*ebp} - \hat{\mu}_d^{*bp})'].$$

As the plug-in estimator $G_{d1}(\hat{\theta})$ of $G_{d1}(\theta)$ is biased, we introduce the *bias-corrected* bootstrap estimator

$$\mathrm{MSE}_*^3(\hat{\mu}_d^{*ebp}) = 2G_{d1}(\hat{\theta}) - E_*[G_{d1}(\hat{\theta}^*)] + E_*[(\hat{\mu}_d^{*ebp} - \hat{\mu}_d^{*bp})(\hat{\mu}_d^{*ebp} - \hat{\mu}_d^{*bp})'].$$

A Monte Carlo approximation $\mathrm{mse}^a(\hat{\mu}_d^{*ebp})$ of the bootstrap matrix $\mathrm{MSE}_*^a(\hat{\mu}_d^{*ebp})$, for $a = 2, 3$, can be obtained similarly as (10).

References

Arima, S., Bell, W. R., Datta, G. S., Franco, C., and Liseo, B. (2017). Multivariate Fay–Herriot Bayesian estimation of small area means under functional measurement error. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(4):1191–1209.

Articus, C. and Burgard, J. P. (2014). A finite mixture Fay Herriot-type model for estimating regional rental prices in Germany. Technical report, University of Trier, Department of Economics. Research Papers in Economics. No. 14/14.

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

Burgard, J. P., Esteban, M. D., Morales, D., and Pérez, A. (2019a). A Fay–Herriot model when auxiliary variables are measured with error. *TEST*. https://doi.org/10.1007/s11749-019-00649-3.

Burgard, J. P., Krause, J., Kreber, D., et al. (2019b). Regularized area-level modelling for robust small area estimation in the presence of unknown covariate measurement errors. Technical report, University of Trier, Department of Economics. Research Papers in Economics. No. 4/19.

Datta, G. S., Ghosh, M., Nangia, N., and Natarajan, K. (1996). Estimation of median income of four-person families: a Bayesian approach. In Berry, D. A., Chaloner, K. M., and Geweke, J. K., editors, *Bayesian Analysis in Statistics and Econometrics*, chapter 11, pages 129–140. Wiley, New York.

Datta, G. S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 110:613–627.

Esteban, M. D., Morales, D., Pérez, A., and Santamaría, L. (2012). Small area estimation of poverty proportions under area-level time models. *Computational Statistics & Data Analysis*, 56(10):2840–2855.

Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74(366):269–277.

González-Manteiga, W., Lombarda, M. J., Molina, I., Morales, D., and Santamaría, L. (2010). Small area estimation under Fay–Herriot models with non-parametric estimation of heteroscedasticity. *Statistical Modelling*, 10(2):215–239.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., and Santamaría, L. (2008). Analytic and bootstrap approximations of prediction errors under a multivariate Fay–Herriot model. *Computational Statistics & Data Analysis*, 52(12):5242–5252.

Li, H. and Lahiri, P. (2010). An adjusted maximum likelihood method for solving small area estimation problems. *Journal of Multivariate Analysis*, 101(4):882–892.

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

Morales, D., Pagliarella, M. C., and Salvatore, R. (2015). Small area estimation of poverty indicators under partitioned area-level time models. *SORT: Statistics and Operations Research Transactions*, 39(1):19–34.

Porter, A. T., Wikle, C. K., and Holan, S. H. (2015). Small area estimation via multivariate Fay–Herriot models with latent spatial dependence. *Australian & New Zealand Journal of Statistics*, 57(1):15–29.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American statistical association*, 85(409):163–171.

Pratesi, M. and Salvati, N. (2008). Small area estimation: the EBLUP estimator based on spatially correlated random area effects. *Statistical methods and applications*, 17(1):113–141.

Ybarra, L. M. R. and Lohr, S. L. (2008). Small area estimation when auxiliary information is measured with error. *Biometrika*, 95(4):919–931.

Yoshimori, M. and Lahiri, P. (2014). A new adjusted maximum likelihood method for the Fay–Herriot small area model. *Journal of Multivariate Analysis*, 124:281–294.

Zayatz, L. (2007). Disclosure avoidance practices and research at the US Census Bureau: An update. *Journal of Official Statistics*, 23(2):253–265.