Supplementary material for manuscript "Small Area Estimation of Socioeconomic Indicators for Sampled and Unsampled Domains"

1 Fitting algorithms

1.1 *Maximum Likelihood Method*

The first derivatives of V_{ud} with respect to $\theta_1 = \sigma_{u1}^2$, $\theta_2 = \sigma_{u2}^2$ and $\theta_3 = \rho$ are

$$\begin{aligned} V_{ud1} &= \frac{\partial V_{ud}}{\partial \sigma_{u1}^2} = \begin{pmatrix} 1 & \frac{\rho\sigma_{u2}}{2\sigma_{u1}} \\ \frac{\rho\sigma_{u2}}{2\sigma_{u1}} & 0 \end{pmatrix}, \quad V_{ud2} = \frac{\partial V_{ud}}{\partial \sigma_{u2}^2} = \begin{pmatrix} 0 & \frac{\rho\sigma_{u1}}{2\sigma_{u2}} \\ \frac{\rho\sigma_{u1}}{2\sigma_{u2}} & 1 \end{pmatrix}, \\ V_{ud3} &= \frac{\partial V_{ud}}{\partial \rho} = \begin{pmatrix} 0 & \sigma_{u1}\sigma_{u2} \\ \sigma_{u1}\sigma_{u2} & 0 \end{pmatrix}. \end{aligned}$$

The second derivatives of V_{ud} with respect to $\theta_1 = \sigma_{u1}^2$, $\theta_2 = \sigma_{u2}^2$ and $\theta_3 = \rho$ are

$$\begin{aligned} V_{ud11} &= \frac{\partial^2 V_{ud}}{\partial \sigma_{u1}^2 \partial \sigma_{u1}^2} = \begin{pmatrix} 0 & \frac{-\rho\sigma_{u2}}{4\sigma_{u1}^3} \\ \frac{-\rho\sigma_{u2}}{4\sigma_{u1}^3} & 0 \end{pmatrix}, \quad V_{ud22} = \frac{\partial^2 V_{ud}}{\partial \sigma_{u2}^2 \partial \sigma_{u2}^2} = \begin{pmatrix} 0 & \frac{-\rho\sigma_{u1}}{4\sigma_{u2}^3} \\ \frac{-\rho\sigma_{u1}}{4\sigma_{u2}^3} & 0 \end{pmatrix}, \\ V_{ud33} &= \frac{\partial^2 V_{ud}}{\partial \rho \partial \rho} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad V_{ud12} = \frac{\partial^2 V_{ud}}{\partial \sigma_{u1}^2 \partial \sigma_{u2}^2} = \begin{pmatrix} 0 & \frac{\rho}{4\sigma_{u1}\sigma_{u2}} \\ \frac{\rho}{4\sigma_{u1}\sigma_{u2}} & 0 \end{pmatrix}, \\ V_{ud13} &= \frac{\partial^2 V_{ud}}{\partial \sigma_{u1}^2 \partial \rho} = \begin{pmatrix} 0 & \frac{\sigma_{u2}}{2\sigma_{u1}} \\ \frac{\sigma_{u2}}{2\sigma_{u1}} & 0 \end{pmatrix}, \quad V_{ud23} = \frac{\partial^2 V_{ud}}{\partial \sigma_{u2}^2 \partial \rho} = \begin{pmatrix} 0 & \frac{\sigma_{u1}}{2\sigma_{u2}} \\ \frac{\sigma_{u1}}{2\sigma_{u2}} & 0 \end{pmatrix}. \end{aligned}$$

It holds that $y \sim N_2(X\beta, V)$, with covariance matrix $V = \underset{1\leq d\leq D}{\text{diag}}(V_d)$, $V_d = V_{ud} + V_{ed}$. 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

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

Let us first consider group $\mathbb{D}_1$. The first partial derivatives of l_d are

$$\begin{aligned}
\frac{\partial l_d}{\partial \beta_{1j}} &= \frac{1}{\sigma_{u1}^2+\sigma_{ed1}^2}(y_{d1}-x_{d1}'\beta_1)x_{d1j}, \quad \frac{\partial l_d}{\partial \beta_{2j}}=0, \quad \frac{\partial l_d}{\partial \theta_2}=\frac{\partial l_d}{\partial \sigma_{u2}^2}=0,\\
\frac{\partial l_d}{\partial \theta_1} &= \frac{\partial l_d}{\partial \sigma_{u1}^2}=-\frac{1}{2}\frac{1}{\sigma_{u1}^2+\sigma_{ed1}^2}+\frac{1}{2}\frac{1}{(\sigma_{u1}^2+\sigma_{ed1}^2)^2}(y_{d1}-x_{d1}'\beta_1)^2, \quad \frac{\partial l_d}{\partial \theta_3}=\frac{\partial l_d}{\partial \rho}=0.
\end{aligned}$$

The second partial derivatives of l_d are

$$\begin{aligned}
\frac{\partial^2 l_d}{\partial \beta_{1j_1}\partial \beta_{1j_2}} &= -\frac{x_{d1j_1}x_{d1j_2}}{\sigma_{u1}^2+\sigma_{ed1}^2}, \quad \frac{\partial^2 l_d}{\partial \beta_{1j_1}\partial \beta_{2j_2}}=0, \quad \frac{\partial^2 l_d}{\partial \beta_{2j_1}\partial \beta_{2j_2}}=0, \quad \frac{\partial^2 l_d}{\partial \beta_{2j}\partial \theta_\ell}=0,\ \ell=1,2,3,\\
\frac{\partial^2 l_d}{\partial \beta_{1j}\partial \theta_1} &= -\frac{1}{(\sigma_{u1}^2+\sigma_{ed1}^2)^2}(y_{d1}-x_{d1}'\beta_1)x_{d1j}, \quad \frac{\partial^2 l_d}{\partial \beta_{1j}\partial \theta_\ell}=0,\ \ell=2,3,\\
\frac{\partial^2 l_d}{\partial \theta_1^2} &= \frac{1}{2}\frac{1}{(\sigma_{u1}^2+\sigma_{ed1}^2)^2}-\frac{1}{(\sigma_{u1}^2+\sigma_{ed1}^2)^3}(y_{d1}-x_{d1}'\beta_1)^2,\\
\frac{\partial^2 l_d}{\partial \theta_1\partial \theta_2} &= \frac{\partial^2 l_d}{\partial \theta_1\partial \theta_3}=0, \quad \frac{\partial^2 l_d}{\partial \theta_2\partial \theta_\ell}=\frac{\partial^2 l_d}{\partial \theta_3\partial \theta_\ell}=0,\ \ell=1,2,3.
\end{aligned}$$

By changing the sign and taking expectations, we have

$$\begin{aligned}
F_{d\beta_{1j_1}\beta_{1j_2}} &= -E\left[\frac{\partial^2 l_d}{\partial \beta_{1j_1}\partial \beta_{1j_2}}\right]=\frac{x_{d1j_1}x_{d1j_2}}{\sigma_{u1}^2+\sigma_{ed1}^2}, \quad F_{d\beta_{2j_1}\beta_{2j_2}}=F_{d\beta_{1j_1}\beta_{2j_2}}=0,\\
F_{d\beta_{1j}\theta_\ell} &= F_{d\beta_{2j}\theta_\ell}=0, \quad \ell=1,2,3,\\
F_{d\theta_1\theta_1} &= \frac{1}{2}\frac{1}{(\sigma_{u1}^2+\sigma_{ed1}^2)^2}, \quad F_{d\theta_2\theta_2}=F_{d\theta_3\theta_3}=F_{d\theta_2\theta_\ell}=F_{d\theta_3\theta_\ell}=0,\ \ell=1,2,3.
\end{aligned}$$

For group $\mathbb{D}_2$, the procedure is accordingly. For the sake of brevity, we do not repeat the mathematical derivations.

Let us consider group $\mathbb{D}_3$. The first partial derivatives of l_d are

$$\begin{aligned}
\frac{\partial l_d}{\partial \beta_{1j}} &= (x_{d1j},0)V_d^{-1}(y_d-X_d\beta), \quad \frac{\partial l_d}{\partial \beta_{2j}}=(0,x_{d2j})V_d^{-1}(y_d-X_d\beta),\\
\frac{\partial l_d}{\partial \theta_\ell} &= -\frac{1}{2}\mathrm{tr}\left(V_d^{-1}V_{ud\ell}\right)+\frac{1}{2}(y_d-X_d\beta)'V_d^{-1}V_{ud\ell}V_d^{-1}(y_d-X_d\beta).
\end{aligned}$$

The second partial derivatives of l_d are

$$\begin{aligned}
\frac{\partial^2 l_d}{\partial \beta_{1j_1}\partial \beta_{1j_2}} &= -(x_{d1j_1},0)V_d^{-1}(x_{d1j_2},0)', \quad \frac{\partial^2 l_d}{\partial \beta_{2j_1}\partial \beta_{2j_2}}=-(0,x_{d2j_1})V_d^{-1}(0,x_{d2j_2})',\\
\frac{\partial^2 l_d}{\partial \beta_{1j_1}\partial \beta_{2j_2}} &= -(x_{d1j_1},0)V_d^{-1}(0,x_{d2j_2})', \quad \frac{\partial^2 l_d}{\partial \beta_{1j}\partial \theta_\ell}=-(x_{d1j},0)V_d^{-1}V_{ud\ell}V_d^{-1}(y_d-X_d\beta),\\
\frac{\partial^2 l_d}{\partial \beta_{2j}\partial \theta_\ell} &= -(0,x_{d2j})V_d^{-1}V_{ud\ell}V_d^{-1}(y_d-X_d\beta),
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 l_d}{\partial\theta_{\ell_1}\partial\theta_{\ell_2}} &= \frac{1}{2}\mathrm{tr}\left(V_d^{-1}V_{ud\ell_2}V_d^{-1}V_{ud\ell_1}\right) - \frac{1}{2}\mathrm{tr}\left(V_d^{-1}V_{ud\ell_1\ell_2}\right) \\
&\quad -(y_d - X_d\beta)'V_d^{-1}V_{ud\ell_2}V_d^{-1}V_{ud\ell_1}V_d^{-1}(y_d - X_d\beta) \\
&\quad +\frac{1}{2}(y_d - X_d\beta)'V_d^{-1}V_{ud\ell_1\ell_2}V_d^{-1}(y_d - X_d\beta).
\end{aligned}
$$

By changing the sign and taking expectations, we have

$$
F_{d\beta_{1j_1}\beta_{1j_2}} = -E\left[\frac{\partial^2 l_d}{\partial\beta_{1j_1}\partial\beta_{1j_2}}\right] = (x_{d1j_1}, 0)V_d^{-1}(x_{d1j_2}, 0)', \quad F_{d\beta_{2j_1}\beta_{2j_2}} = (0, x_{d2j_1})V_d^{-1}(0, x_{d2j_2})',
$$

$$
F_{d\beta_{1j_1}\beta_{2j_2}} = (x_{d1j_1}, 0)V_d^{-1}(0, x_{d2j_2})', \;\; F_{d\beta_{1j}\theta_\ell} = F_{d\beta_{2j}\theta_\ell} = 0, \;\; F_{d\theta_{\ell_1}\theta_{\ell_2}} = \frac{1}{2}\mathrm{tr}\left(V_d^{-1}V_{ud\ell_2}V_d^{-1}V_{ud\ell_1}\right).
$$

The components of the score vector are

$$
S_{\beta_1 j} = \sum_{d=1}^D \frac{\partial l_d}{\partial\beta_{1j}}, \; 1 \le j \le p_1; \quad S_{\beta_2 j} = \sum_{d=1}^D \frac{\partial l_d}{\partial\beta_{2j}}, \; 1 \le j \le p_2; \quad S_{\theta_\ell} = \sum_{d=1}^D \frac{\partial l_d}{\partial\theta_\ell}, \; \ell = 1, 2, 3.
$$

The $(p+3)\times 1$ score vector is $S(\psi) = (S_\beta'(\psi), S_\theta'(\psi)'$, where

$$
S_\beta(\psi) = \left(S_{\beta_{11}}, \ldots, S_{\beta_{1p_1}}, S_{\beta_{21}}, \ldots, S_{\beta_{2p_2}}\right)', \quad S_\theta(\psi)' = (S_{\theta_1}, S_{\theta_2}, S_{\theta_3})'.
$$

The components of the Fisher information matrix are

$$
F_{\beta_{1j_1}\beta_{1j_2}} = \sum_{d=1}^D F_{d\beta_{1j_1}\beta_{1j_2}}, \; F_{\beta_{2j_1}\beta_{2j_2}} = \sum_{d=1}^D F_{d\beta_{2j_1}\beta_{2j_2}}, \; F_{\beta_{1j_1}\beta_{2j_2}} = \sum_{d=1}^D F_{d\beta_{1j_1}\beta_{2j_2}},
$$

$$
F_{\beta_{1j}\theta_\ell} = \sum_{d=1}^D F_{d\beta_{1j}\theta_\ell}, \; F_{\beta_{2j}\theta_\ell} = \sum_{d=1}^D F_{d\beta_{2j}\theta_\ell}, \; F_{\theta_{\ell_1}\theta_{\ell_2}} = \sum_{d=1}^D F_{d\theta_{\ell_1}\theta_{\ell_2}}.
$$

The blocks of the Fisher information matrix are

$$
F_{\beta_1\beta_1} = \left(F_{\beta_{1j_1}\beta_{1j_2}}\right)_{j_1,j_2=1,\ldots,p_1}, F_{\beta_2\beta_2} = \left(F_{\beta_{2j_1}\beta_{2j_2}}\right)_{j_1,j_2=1,\ldots,p_2}, F_{\beta_1\beta_2} = \left(F_{\beta_{1j_1}\beta_{2j_2}}\right)_{j_1=1,\ldots,p_1; j_2=1,\ldots,p_2},
$$

$$
F_{\beta_1\theta} = \left(F_{\beta_{1j}\theta_\ell}\right)_{j=1,\ldots,p_1;\ell=1,2,3}, \; F_{\beta_2\theta} = \left(F_{\beta_{2j}\theta_\ell}\right)_{j=1,\ldots,p_2;\ell=1,2,3}, \; F_{\theta\theta} = \left(F_{\theta_{\ell_1}\theta_{\ell_2}}\right)_{\ell_1=1,2,3;\ell_2=1,2,3}.
$$

The $(p+3)\times(p+3)$ Fisher information matrix is

$$
F(\psi) = \begin{pmatrix} F_{\beta_1\beta_1} & F_{\beta_1\beta_2} & F_{\beta_1\theta} \\ F_{\beta_2\beta_1} & F_{\beta_2\beta_2} & F_{\beta_2\theta} \\ F_{\theta\beta_1} & F_{\theta\beta_2} & F_{\theta\theta} \end{pmatrix} = \begin{pmatrix} F_{\beta\beta} & 0 \\ 0 & F_{\theta\theta} \end{pmatrix}, \quad F_{\beta\beta}(\psi) = \begin{pmatrix} F_{\beta_1\beta_1} & F_{\beta_1\beta_2} \\ F_{\beta_2\beta_1} & F_{\beta_2\beta_2} \end{pmatrix}_{p\times p}.
$$

The ML Fisher-scoring algorithm 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

$$\begin{aligned}\beta^{(r+1)} &= \beta^{(r)} + F_{\beta\beta}^{-1}(\beta^{(r)}, \theta^{(r)}) S_\beta(\beta^{(r)}, \theta^{(r)}),\\ \theta^{(r+1)} &= \theta^{(r)} + F_{\theta\theta}^{-1}(\beta^{(r+1)}, \theta^{(r)}) S_\theta(\beta^{(r+1)}, \theta^{(r)}).\end{aligned}$$

(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 $\left|\psi_m^{(r+1)} - \psi_m^{(r)}\right| < \varepsilon$, $m = 1, \ldots, p+3$, do $\hat{\psi} = \psi^{(r+1)}$ and stop. Otherwise, continue.

3. Output: $\hat{\psi}$, $F^{-1}(\hat{\psi})$.

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 ML estimates of the corresponding bivariate Fay-Herriot that uses only the data of group $\mathbb{D}_3$.

1.2 *Residual Maximum Likelihood Method*

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

$$l_{reml}(\theta) = -\frac{2D-p}{2}\log 2\pi + \frac{1}{2}\log|X'X| - \frac{1}{2}\log|V| - \frac{1}{2}\log|X'V^{-1}X| - \frac{1}{2}\,y'Py, \qquad (1.1)$$

where $\theta = (\theta_1, \theta_2, \theta_3)$, $\theta_1 = \sigma_{u1}^2$, $\theta_2 = \sigma_{u2}^2$, $\theta_3 = \rho$, y is the $2D$-vector $y = (y_1', \ldots, y_D')'$, X is the corresponding $2D \times p$-matrix $X = (X_1, \ldots, X_D)'$, and $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$, $PVP = P$ with $PX = 0$.
In the case that there are missing values, the residual maximum likelihood (REML) log-likelihood is

$$\begin{aligned}l_{reml}(\theta) = \; & -\frac{D_1 + D_2 + 2(D - D_1 - D_2) - p}{2}\log 2\pi + \frac{1}{2}\log|X'X| \\ & -\frac{1}{2}\log|V| - \frac{1}{2}\log|X'V^{-1}X| - \frac{1}{2}\,y'Py,\end{aligned} \qquad (1.2)$$

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 $k = 1, 2$. P is defined as before on the reduced X and V. 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.
By applying the formulas

$$\frac{\partial \log|V|}{\partial \theta} = \operatorname{tr}\left(V^{-1}\frac{\partial V}{\partial \theta}\right), \quad \frac{\partial V^{-1}}{\partial \theta} = -V^{-1}\frac{\partial V}{\partial \theta}V^{-1},$$

we calculate the first partial derivatives of l_{reml} with respect to θ_ℓ, i.e.

$$\frac{\partial l_{reml}(\theta)}{\partial \theta_\ell} = -\frac{1}{2}\operatorname{tr}\left(P\frac{\partial V}{\partial \theta_\ell}\right) - \frac{1}{2}\,y'\frac{\partial P}{\partial \theta_\ell}y, \quad \ell = 1, 2, 3.$$

Let us define $G = V^{-1}X(X'V^{-1}X)^{-1}$, so that $P = (I - GX')V^{-1} = V^{-1}(I - XG')$. The first partial derivatives of $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$ with respect to θ_ℓ are

$$\frac{\partial P}{\partial \theta_\ell} \;=\; -(I - GX')V^{-1}\frac{\partial V}{\partial \theta_\ell}V^{-1}(I - GX')' = -P\frac{\partial V}{\partial \theta_\ell}P, \quad \ell = 1,2,3\,.$$

Therefore, the score vector is

$$S(\theta) = (S_1, S_2, S_3)',\ S_\ell = S_\ell(\theta) = \frac{\partial l_{reml}}{\partial \theta_\ell} = -\frac{1}{2}\,\mathrm{tr}(PV_\ell) + \frac{1}{2}\,y'PV_\ell Py, \quad \ell = 1,2,3,$$

where $V_\ell = \frac{\partial V}{\partial \theta_\ell} = \underset{1\leq d\leq D}{\mathrm{diag}}\,(V_{ud\ell})$, $Q = (X'V^{-1}X)^{-1}$ and

$$\begin{aligned}
V_{ud1} &= \frac{\partial V_{ud}}{\partial \sigma_{u1}^2} = \begin{pmatrix} 1 & \frac{\rho\sigma_{u2}}{2\sigma_{u1}} \\ \frac{\rho\sigma_{u2}}{2\sigma_{u1}} & 0 \end{pmatrix}, \quad V_{ud2} = \frac{\partial V_{ud}}{\partial \sigma_{u2}^2} = \begin{pmatrix} 0 & \frac{\rho\sigma_{u1}}{2\sigma_{u2}} \\ \frac{\rho\sigma_{u1}}{2\sigma_{u2}} & 1 \end{pmatrix}, \\
V_{ud3} &= \frac{\partial V_{ud}}{\partial \rho} = \begin{pmatrix} 0 & \sigma_{u1}\sigma_{u2} \\ \sigma_{u1}\sigma_{u2} & 0 \end{pmatrix}.
\end{aligned}$$

For $a, b = 1,2,3$, the second partial derivatives of the REML log-likelihood function are

$$\frac{\partial l_{reml}^2(\theta)}{\partial \theta_a \partial \theta_b} = \frac{1}{2}\,\mathrm{tr}\,(PV_aPV_b) - y'PV_aPV_bPy,$$

where the last equality follows from the fact that V_ℓ is symmetric, $\ell = 1,2,3$. By changing the sign, taking expectations and applying $PX = 0$, $PV = I - V^{-1}XQX'$ and the formula

$$E\,[y'A\;y] = \mathrm{tr}\,(A\;\mathrm{var}(y)) + E[y]'A\;E[y],$$

we get the components of the Fisher information matrix, i.e.

$$F_{ab} \;=\; F_{ab}(\theta) = \frac{1}{2}\,\mathrm{tr}\,(PV_aPV_b)\,, \quad a, b = 1,2,3.$$

Therefore, the Fisher information matrix is

$$F(\theta) = \left(F_{a,b}\right)_{a,b=1,2,3}, \quad F_{ab} = F_{ab}(\theta) = \frac{1}{2}\mathrm{tr}(PV_aPV_b), \quad a, b = 1,2,3.$$

The REML Fisher-scoring algorithm is

1. Set the initial values $\beta^{(0)}$, $\theta^{(0)}$, and $\varepsilon_{p+\ell} > 0$, $\varepsilon_i > 0$, $\ell = 1,2,3$, $i = 1,\ldots,p$. Let β_i denote the i-th element of the vector β, not the vector of regression parameters of the i-th variable.

2. Repeat the following steps until the tolerance or the boundary conditions are fulfilled.

   (a) Updating equation for θ: Do $\theta^{(r+1)} = \theta^{(r)} + F^{-1}(\theta^{(r)})S(\theta^{(r)})$.

   (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 $\hat{\theta} = \theta^{(r)}$ and stop.

(c) Updating equation for β: Do $\beta^{(r+1)} = \left(X'V^{-1}(\theta^{(r+1)})X\right)^{-1} X'V^{-1}(\theta^{(r+1)})y$.

(d) Tolerance condition: If $\left|\theta_\ell^{(r+1)} - \theta_\ell^{(r)}\right| < \varepsilon_{p+\ell}$, $\left|\beta_i^{(r+1)} - \beta_i^{(r)}\right| < \varepsilon_i$, $\ell = 1,2,3$, $i = 1,\dots,p$, do $\hat{\theta}_\ell = \theta_\ell^{(r+1)}$, $\hat{\beta}_i = \beta_i^{(r+1)}$ 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. The confidence intervals are given by

$$\hat{\theta}_\ell \pm z_{\alpha/2}\, \nu_{\ell\ell}^{1/2},\ \ell = 1,2,3, \quad \hat{\beta}_i \pm z_{\alpha/2}\, q_{ii}^{1/2},\ i = 1,\dots,p, \tag{1.3}$$

where $F^{-1}(\hat{\theta}) = (\nu_{ab})_{a,b=1,2,3}$, $(X'V^{-1}(\hat{\theta})X)^{-1} = (q_{ij})_{i,j=1,\dots,p}$ 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

$$p\text{-value} = 2P_{H_0}(\hat{\beta}_i > |\beta_0|) = 2P(N(0,1) > |\beta_0|/\sqrt{q_{ii}}\,). \tag{1.4}$$

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

2 Best prediction of random effects

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 results R1-R3.

R1. If $d\in\mathbb{D}_1$, then the BP of u_d under the MBFH model is

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

where $y_{\bar{d}1}=(y_{d1},0)'$.

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

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

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

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

□

R2. If $d\in\mathbb{D}_2$, then the BP of u_d under the MBFH model is

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

where $y_{\bar{d}2}=(0,y_{d2})'$.

Proof. The proof is analogous as the one of Result1.

R3. If $d\in\mathbb{D}_3$, then BP of u_d under the MBFH model is

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

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

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

□

3 Analytic Approximation of the Mean Squared Error of the MBFH Predictor

3.1 *Best Predictors*

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

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

The BP of μ_d is $\hat{\mu}_d^{bp} = X_d\beta + \hat{u}_d^{bp}$, where

$$\hat{u}_d^{bp} = \Phi_{d1}A_{d1}(y_{\bar{d}1} - X_d\beta) = \Phi_{d1}A_{d1}\begin{pmatrix} u_{d1} + e_{d1} \\ -x'_{d2}\beta_2 \end{pmatrix}.$$

The variance matrix of $\hat{\mu}_d^{bp}$ is

$$\begin{aligned}\mathrm{var}(\hat{\mu}_d^{bp}) &= \mathrm{var}(\hat{u}_d^{bp}) = E\big[\hat{u}_d^{bp}\hat{u}_d^{bp\prime}\big] = \Phi_{d1}A_{d1}E\left[\begin{pmatrix} u_{d1} + e_{d1} \\ -x'_{d2}\beta_2 \end{pmatrix}(u_{d1} + e_{d1}, -x'_{d2}\beta_2)\right]A_{d1}\Phi_{d1} \\ &= \Phi_{d1}A_{d1}\begin{pmatrix} \sigma_{u1}^2 + \sigma_{ed1}^2 & 0 \\ 0 & 0 \end{pmatrix}A_{d1}\Phi_{d1} = \Phi_{d1}A_{d1}(V_{ud} + V_{ed})A_{d1}\Phi_{d1}.\end{aligned}$$

Further, the expectation matrix $E\big[\hat{u}_d^{bp}u'_d\big]$ is

$$E\big[\hat{u}_d^{bp}u'_d\big] = \Phi_{d1}A_{d1}E\left[\begin{pmatrix} u_{d1} + e_{d1} \\ -x'_{d2}\beta_2 \end{pmatrix}(u_{d1}, u_{d2})\right] = \Phi_{d1}A_{d1}\begin{pmatrix} \sigma_{u1}^2 & \rho\sigma_{u1}\sigma_{u2} \\ 0 & 0 \end{pmatrix} = \Phi_{d1}A_{d1}V_{ud}.$$

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

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

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

Let us consider the group $\mathbb{D}_3$. The BP of μ_d is $\hat{\mu}_d^{bp} = X_d\beta + \hat{u}_d^{bp}$, where

$$\hat{u}_d^{bp} = \Phi_d V_{ed}^{-1}(y_d - X_d\beta), \quad \Phi_d = \left(V_{ed}^{-1} + V_{ud}^{-1}\right)^{-1} \triangleq \begin{pmatrix} \phi_{d,11} & \phi_{d,12} \\ \phi_{d,12} & \phi_{d,22} \end{pmatrix}.$$

The variance matrix of $\hat{\mu}_d^{bp}$ is

$$\mathrm{var}(\hat{\mu}_d^{bp}) = \mathrm{var}(\hat{u}_d^{bp}) = E\big[\hat{u}_d^{bp}\hat{u}_d^{bp\prime}\big] = \Phi_d V_{ed}^{-1}\mathrm{var}(u_d + e_d)V_{ed}^{-1}\Phi_d = \Phi_d V_{ed}^{-1}(V_{ud} + V_{ed})V_{ed}^{-1}\Phi_d.$$

Further, the expectation matrix $E\big[\hat{u}_d^{bp}u'_d\big]$ is

$$E\big[\hat{u}_d^{bp}u'_d\big] = \Phi_d V_{ed}^{-1}E\left[\begin{pmatrix} u_{d1} + e_{d2} \\ u_{d2} + e_{d2} \end{pmatrix}(u_{d1}, u_{d2})\right] = \Phi_d V_{ed}^{-1}V_{ud}.$$

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

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

3.2 *Empirical Best Predictors in Domains of Groups* $\mathbb{D}_1$ *and* $\mathbb{D}_2$

As the mathematical derivations for groups $\mathbb{D}_1$ and $\mathbb{D}_2$ are somehow symmetric, we only present those corresponding to group $\mathbb{D}_1$. Let us consider the group $\mathbb{D}_1$. The MSE of the EBP is

$$\begin{aligned} \mathrm{MSE}(\hat{\mu}_d^{ebp}) &= E\big[(\hat{\mu}_d^{ebp}-\mu_d)(\hat{\mu}_d^{ebp}-\mu_d)'\big] = E\big[(\hat{\mu}_d^{bp}-\mu_d)(\hat{\mu}_d^{bp}-\mu_d)'\big] \\ &+ E\big[(\hat{\mu}_d^{ebp}-\hat{\mu}_d^{bp})(\hat{\mu}_d^{ebp}-\hat{\mu}_d^{bp})'\big] + 2E\big[(\hat{\mu}_d^{ebp}-\hat{\mu}_d^{bp})(\hat{\mu}_d^{bp}-\mu_d)'\big]. \end{aligned}$$

The first summand is $S_{d1} = E\big[(\hat{\mu}_d^{bp}-\mu_d)(\hat{\mu}_d^{eb}-\mu_d)'\big] = \mathrm{MSE}(\hat{\mu}_d^{bp})$.
Let us now approximate the second summand $S_{d2} = E\big[(\hat{\mu}_d^{ebp}-\hat{\mu}_d^{bp})(\hat{\mu}_d^{ebp}-\hat{\mu}_d^{bp})'\big]$. 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

$$h_d(\hat{\beta},\hat{\theta}) \triangleq \hat{\mu}_d^{ebp} = X_d\hat{\beta} + \hat{\Phi}_{d1}A_{d1}(y_{\bar{d}1} - X_d\hat{\beta}) = \begin{pmatrix} x_{d1}'\hat{\beta}_1 \\ x_{d2}\hat{\beta}_2 \end{pmatrix} + \frac{y_{d1}-x_{d1}'\hat{\beta}_1}{\sigma_{ed1}^2}\begin{pmatrix} \hat{\phi}_{d1,11} \\ \hat{\phi}_{d1,12} \end{pmatrix}.$$

where

$$\hat{\Phi}_{d1} = \Phi_{d1}(\hat{\theta}) = \Big(A_{d1} + \hat{V}_{ud}^{-1}\Big)^{-1}, \quad \hat{V}_{ud} = V_{ud}(\hat{\theta}) = \begin{pmatrix} \hat{\sigma}_{u1}^2 & \hat{\rho}\hat{\sigma}_{u1}\hat{\sigma}_{u2} \\ \hat{\rho}\hat{\sigma}_{u1}\hat{\sigma}_{u2} & \hat{\sigma}_{u2}^2 \end{pmatrix}.$$

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

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

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

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

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 $H_{d\beta_{k_1}\beta_{k_2},ab} = h_{d\beta_{k_1},a}h_{d\beta_{k_2},b}'$, $k_1,k_2,a,b=1,2$.
The 3×1 vectors containing the derivatives with respect to θ are

$$h_{d\theta,1} = \underset{1\le\ell\le 3}{\mathrm{col}}\big(h_{d\theta_\ell,1}\big),\; g_{d\theta,1} = \underset{1\le\ell\le 3}{\mathrm{col}}\big(g_{d\theta_\ell,1}\big), \quad h_{d\theta,2} = \underset{1\le\ell\le 3}{\mathrm{col}}\big(h_{d\theta_\ell,2}\big),\; g_{d\theta,2} = \underset{1\le\ell\le 3}{\mathrm{col}}\big(g_{d\theta_\ell,2}\big)$$

and the corresponding 3×3 matrices are $H_{d\theta\theta,ab} = h_{d\theta,a}h_{d\theta,b}'$, $G_{d\theta\theta,ab} = g_{d\theta,a}g_{d\theta,b}'$, $a,b=1,2$.
The crossed $p_k\times 3$ matrices are $H_{d\beta_k\theta,ab} = h_{d\beta_k,a}h_{d\theta,b}'$, $G_{d\beta_k\theta,ab} = h_{d\beta_k,a}g_{d\theta,b}'$, $k,a,b=1,2$.

We further define the $p_k \times 2$ and 3×2 matrices

$$\begin{aligned} h_{d\beta_k} &= \frac{\partial h_d}{\partial \beta_k} = \operatorname*{col}_{1\le j\le p_k}\Big(\frac{\partial h'_d(\beta,\theta)}{\partial \beta_{kj}}\Big) = \mathrm{row}(h_{d\beta_k,1}, h_{d\beta_k,2}), \quad k=1,2, \\ h_{d\theta} &= \frac{\partial h_d}{\partial \theta} = \operatorname*{col}_{1\le \ell\le 3}\Big(\frac{\partial h'_d(\beta,\theta)}{\partial \theta_\ell}\Big) = \mathrm{row}(h_{d\theta,1}, h_{d\theta,2}). \end{aligned}$$

A Taylor series expansion of the 2×1 vector $h_d(\hat\beta,\hat\theta)$ around (β,θ) yields

$$h_d(\hat\beta,\hat\theta) = h_d(\beta,\theta)+\sum_{k=1}^{2}\sum_{j=1}^{p_k}\frac{\partial h'_d(\beta,\theta)}{\partial \beta_{kj}}(\hat\beta_{kj}-\beta_{kj})+\sum_{\ell=1}^{3}\frac{\partial h'_d(\beta,\theta)}{\partial \theta_\ell}(\hat\theta_\ell-\theta_\ell)+o_p(\|\hat\beta-\beta\|_2)+o_p(\|\hat\theta-\theta\|_2).$$

Therefore,

$$h_d(\hat\beta,\hat\theta) = h_d(\beta,\theta)+h'_{d\beta_1}(\hat\beta_1-\beta_1)+h'_{d\beta_2}(\hat\beta_2-\beta_2)+h'_{d\theta}(\hat\theta-\theta)+o_p(\|\hat\beta-\beta\|_2)+o_p(\|\hat\theta-\theta\|_2)$$

and

$$\begin{aligned} S_{d2} &= E\big[(\hat\mu_d^{ebp}-\hat\mu_d^{bp})(\hat\mu_d^{ebp}-\hat\mu_d^{bp})'\big] = E\big[(h_d(\hat\beta,\hat\theta)-h_d(\beta,\theta))(h_d(\hat\beta,\hat\theta)-h_d(\beta,\theta))'\big] \\ &= E\big[h'_{d\beta_1}(\hat\beta_1-\beta_1)(\hat\beta_1-\beta_1)'h_{d\beta_1}\big] + E\big[h'_{d\beta_2}(\hat\beta_2-\beta_2)(\hat\beta_2-\beta_2)'h_{d\beta_2}\big] \\ &+ E\big[h'_{d\theta}(\hat\theta-\theta)(\hat\theta-\theta)'h_{d\theta}\big] \\ &+ \big\{E\big[h'_{d\beta_1}(\hat\beta_1-\beta_1)(\hat\beta_2-\beta_2)'h_{d\beta_2}\big] + E\big[h'_{d\beta_2}(\hat\beta_2-\beta_2)(\hat\beta_1-\beta_1)'h_{d\beta_1}\big]\big\} \\ &+ \big\{E\big[h'_{d\beta_1}(\hat\beta_1-\beta_1)(\hat\theta-\theta)'h_{d\theta}\big] + E\big[h'_{d\theta}(\hat\theta-\theta)(\hat\beta_1-\beta_1)'h_{d\beta_1}\big]\big\} \\ &+ \big\{E\big[h'_{d\beta_2}(\hat\beta_2-\beta_2)(\hat\theta-\theta)'h_{d\theta}\big] + E\big[h'_{d\theta}(\hat\theta-\theta)(\hat\beta_2-\beta_2)'h_{d\beta_2}\big]\big\} \\ &+ o_p(\|\hat\beta-\beta\|_2^2)+o_p(\|\hat\theta-\theta\|_2^2). \end{aligned}$$

Concerning the first summand of S_{d2}, $E_{d11} = E\big[h'_{d\beta_1}(\hat\beta_1-\beta_1)(\hat\beta_1-\beta_1)'h_{d\beta_1}\big]$, we observe that

$$h'_{d\beta_1}(\hat\beta_1-\beta_1)(\hat\beta_1-\beta_1)'h_{d\beta_1} = \begin{pmatrix} h'_{d\beta_1,1}(\hat\beta_1-\beta_1)(\hat\beta_1-\beta_1)'h_{d\beta_1,1} & h'_{d\beta_1,1}(\hat\beta_1-\beta_1)(\hat\beta_1-\beta_1)'h_{d\beta_1,2} \\ h'_{d\beta_1,2}(\hat\beta_1-\beta_1)(\hat\beta_1-\beta_1)'h_{d\beta_1,1} & h'_{d\beta_1,2}(\hat\beta_1-\beta_1)(\hat\beta_1-\beta_1)'h_{d\beta_1,2} \end{pmatrix}.$$

Note that $h'_{d\beta_1,1}(\hat\beta_1-\beta_1)$ and $h'_{d\beta_1,2}(\hat\beta_1-\beta_1)$ are scalars. Then, for $a,b=1,2$, we have

$$h'_{d\beta_1,a}(\hat\beta_1-\beta_1)(\hat\beta_1-\beta_1)'h_{d\beta_1,b} = (\hat\beta_1-\beta_1)'h_{d\beta_1,b}h'_{d\beta_1,a}(\hat\beta_1-\beta_1) = (\hat\beta_1-\beta_1)'H_{d\beta_1\beta_1,ba}(\hat\beta_1-\beta_1).$$

Therefore,

$$E_{d11} = E\left[\begin{pmatrix} (\hat\beta_1-\beta_1)'H_{d\beta_1\beta_1,11}(\hat\beta_1-\beta_1) & (\hat\beta_1-\beta_1)'H_{d\beta_1\beta_1,21}(\hat\beta_1-\beta_1) \\ (\hat\beta_1-\beta_1)'H_{d\beta_1\beta_1,12}(\hat\beta_1-\beta_1) & (\hat\beta_1-\beta_1)'H_{d\beta_1\beta_1,22}(\hat\beta_1-\beta_1) \end{pmatrix}\right].$$

We apply to $z=\hat\beta_1-\beta_1$ the formula

$$E[z'Qz] = \mathrm{tr}\{Q\,\mathrm{var}(z)\}+E[z]'QE[z].$$

As the matrices $H_{d\beta_1\beta_1,ab}$ are not random and $E[\hat\beta_1-\beta_1]=O(\|\hat\beta_1-\beta_1\|_2)$, we obtain

$$E_{d11}=\begin{pmatrix} \mathrm{tr}\{H_{d\beta_1\beta_1,11}\,\mathrm{var}(\hat\beta_1)\} & \mathrm{tr}\{H_{d\beta_1\beta_1,21}\,\mathrm{var}(\hat\beta_1)\} \\ \mathrm{tr}\{H_{d\beta_1\beta_1,12}\,\mathrm{var}(\hat\beta_1)\} & \mathrm{tr}\{H_{d\beta_1\beta_1,22}\,\mathrm{var}(\hat\beta_1)\} \end{pmatrix}+O_{2\times 2}(\|\hat\beta_1-\beta_1\|_2^2).$$

For the second summand of S_{d2}, $E_{d22}=E\big[h'_{d\beta_2}(\hat\beta_2-\beta_2)(\hat\beta_2-\beta_2)'h_{d\beta_2}\big]$, we have

$$E_{d22}=\begin{pmatrix} \mathrm{tr}\{H_{d\beta_2\beta_2,11}\,\mathrm{var}(\hat\beta_2)\} & \mathrm{tr}\{H_{d\beta_2\beta_2,21}\,\mathrm{var}(\hat\beta_2)\} \\ \mathrm{tr}\{H_{d\beta_2\beta_2,12}\,\mathrm{var}(\hat\beta_2)\} & \mathrm{tr}\{H_{d\beta_2\beta_2,22}\,\mathrm{var}(\hat\beta_2)\} \end{pmatrix}+O_{2\times 2}(\|\hat\beta_2-\beta_2\|_2^2).$$

The fourth summand of S_{d2} is $E_{d12}+E_{d21}$, where $E'_{d21}=E_{d12}=E\big[h'_{d\beta_1}(\hat\beta_1-\beta_1)(\hat\beta_2-\beta_2)'h_{d\beta_2}\big]$,

$$E_{d12}=\begin{pmatrix} \mathrm{tr}\{H_{d\beta_2\beta_1,11}\,\mathrm{cov}(\hat\beta_1,\hat\beta_2)\} & \mathrm{tr}\{H_{d\beta_2\beta_1,21}\,\mathrm{cov}(\hat\beta_1,\hat\beta_2)\} \\ \mathrm{tr}\{H_{d\beta_2\beta_1,12}\,\mathrm{cov}(\hat\beta_1,\hat\beta_2)\} & \mathrm{tr}\{H_{d\beta_2\beta_1,22}\,\mathrm{cov}(\hat\beta_1,\hat\beta_2)\} \end{pmatrix}+O_{2\times 2}(\|\hat\beta_1-\beta_1\|_2\|\hat\beta_1-\beta_1\|_2).$$

The third summand of S_{d2}, $E_{d00}=E\big[h'_{d\theta}(\hat\theta-\theta)(\hat\theta-\theta)'h_{d\theta}\big]$ is

$$E_{d00}=E\left[\begin{pmatrix} (\hat\theta-\theta)'H_{d\theta\theta,11}(\hat\theta-\theta) & (\hat\theta-\theta)'H_{d\theta\theta,21}(\hat\theta-\theta) \\ (\hat\theta-\theta)'H_{d\theta\theta,12}(\hat\theta-\theta) & (\hat\theta-\theta)'H_{d\theta\theta,22}(\hat\theta-\theta) \end{pmatrix}\right].$$

We recall that

$$H_{d\theta\theta,ab}=\frac{(y_{d1}-x'_{d1}\beta_1)^2}{\sigma^4_{ed1}}\begin{pmatrix} \phi^2_{d1\ell,11} & \phi_{d1\ell,11}\phi_{d1\ell,12} \\ \phi_{d1\ell,11}\phi_{d1\ell,12} & \phi^2_{d1\ell,12} \end{pmatrix}\triangleq\frac{(y_{d1}-x'_{d1}\beta_1)^2}{\sigma^4_{ed1}}G_{d\theta\theta,ab}$$

is random, but $G_{d\theta\theta,ab}$ is not. This is why we calculate E_{d00} for the estimators $\hat\theta_{d-}$ and $\hat\beta_{kd-}$, $k=1,2$, based on $y_{d-}=\big(y_{d'}\big)_{d'\neq d}$. As the target data y can be split into the independent parts y_{d-} and y_{d_1}, we have

$$\begin{aligned}
E_{d-00,ab} &= E\big[(\hat\theta_{d-}-\theta)'H_{d\theta\theta,ab}(\hat\theta_{d-}-\theta)\big]=E_{y_{d-}}\left[E_{y_{d_1}}\big[(\hat\theta_{d-}-\theta)'H_{d\theta\theta,ab}(\hat\theta_{d-}-\theta)\big]\right] \\
&= E_{y_{d-}}\left[(\hat\theta_{d-}-\theta)'G_{d\theta\theta,ab}(\hat\theta_{d-}-\theta)\sigma^{-4}_{ed1}E_{y_{d_1}}\big[(y_{d1}-x'_{d1}\beta_1)^2\big]\right] \\
&= \frac{\sigma^2_{ud1}+\sigma^2_{ed1}}{\sigma^4_{ed1}}E_{y_{d-}}\left[(\hat\theta_{d-}-\theta)'G_{d\theta\theta,ab}(\hat\theta_{d-}-\theta)\right] \\
&= \frac{\sigma^2_{ud1}+\sigma^2_{ed1}}{\sigma^4_{ed1}}\,\mathrm{tr}\{G_{d\theta\theta,b}\,\mathrm{var}(\hat\theta_{d-})\}+O_{2\times 2}(\|\hat\theta_{d-}-\theta\|_2^2).
\end{aligned}$$

Taken out (y_{d1},x_{d1}) from the data file should not have great influence on $\mathrm{var}(\hat\theta)$. Therefore, we take

$$E_{d00}=\frac{\sigma^2_{ud1}+\sigma^2_{ed1}}{\sigma^4_{ed1}}\begin{pmatrix} \mathrm{tr}\{G_{d\theta\theta,11}\,\mathrm{var}(\hat\theta)\} & \mathrm{tr}\{G_{d\theta\theta,21}\,\mathrm{var}(\hat\theta)\} \\ \mathrm{tr}\{G_{d\theta\theta,12}\,\mathrm{var}(\hat\theta)\} & \mathrm{tr}\{G_{d\theta\theta,22}\,\mathrm{var}(\hat\theta)\} \end{pmatrix}+O_{2\times 2}(\|\hat\theta-\theta\|_2^2).$$

The fifth and sixth summands of S_{d2} are $E_{d10}+E_{d01}$ and $E_{d20}+E_{d02}$ respectively, where $E'_{d0k}=E_{dk0}=E\big[h'_{d\beta_k}(\hat\beta_k-\beta_k)(\hat\theta-\theta)'h_{d\theta}\big]$, $k=1,2$, and

$$E_{dk0}=E\left[\begin{pmatrix} (\hat\beta_k-\beta_k)'H_{d\beta_k\theta,11}(\hat\theta-\theta) & (\hat\beta_k-\beta_k)'H_{d\beta_k\theta,12}(\hat\theta-\theta) \\ (\hat\beta_k-\beta_k)'H_{d\beta_k\theta,21}(\hat\theta-\theta) & (\hat\beta_k-\beta_k)'H_{d\beta_k\theta,22}(\hat\theta-\theta) \end{pmatrix}\right].$$

We recall that

$$H_{d\beta_k\theta,ab} = \frac{y_{d1} - x'_{d1}\beta_1}{\sigma^2_{ed1}} G_{d\beta_k\theta,ab}$$

is random, but $G_{d\beta_k\theta,ab}$ is not. It holds that

$$\begin{aligned} E_{d\_k0,ab} &= E\big[(\hat\beta_{kd\_} - \beta)' H_{d\beta_k\theta,ab}(\hat\theta_{d\_} - \theta)\big] = E_{y_{d\_}}\big[E_{y_{d_1}}\big[(\hat\beta_{kd\_} - \beta)' H_{d\beta_k\theta,ab}(\hat\theta_{d\_} - \theta)\big]\big] \\ &= E_{y_{d\_}}\big[(\hat\beta_{kd\_} - \beta)' G_{d\beta_k\theta,ab}(\hat\theta_{d\_} - \theta)\sigma^{-2}_{ed1} E_{y_{d_1}}\big[y_{d1} - x'_{d1}\beta_1\big]\big] = 0. \end{aligned}$$

Therefore, we take $E_{dk0} = 0_{2\times 2}$, $k = 1, 2$.

Let us now approximate the third summand $S_{d3} = E\big[(\hat\mu^{ebp}_d - \hat\mu^{bp}_d)(\hat\mu^{bp}_d - \mu_d)'\big]$. We recall that

$$\hat\mu^{ebp}_d - \hat\mu^{bp}_d = h'_{d\beta_1}(\hat\beta_1 - \beta_1) + h'_{d\beta_2}(\hat\beta_2 - \beta_2) + h'_{d\theta}(\hat\theta - \theta) + o_p(\|\hat\beta - \beta\|_2) + o_p(\|\hat\theta - \theta\|_2),$$

$$\hat\mu^{bp}_d - \mu_d = \hat u^{bp}_d - u_d = \frac{u_{d1} + e_{d1}}{\sigma^2_{ed1}} \begin{pmatrix} \phi_{d1,11} \\ \phi_{d1,12} \end{pmatrix} - \begin{pmatrix} u_{d1} \\ u_{d2} \end{pmatrix}.$$

For $k = 1, 2$, it holds that

$$\begin{aligned} E_{d3k} &= E\Big[h'_{d\beta_k}(\hat\beta_{kd\_} - \beta_k)\big(\hat u^{bp}_d - u_d\big)'\Big] \\ &= E_{(u_{d\_}, e_{d\_})}\left[h'_{d\beta_{kd\_}}(\hat\beta_k - \beta_k) E_{(u_d, e_{d1})}\left[\frac{u_{d1} + e_{d1}}{\sigma^2_{ed1}} \begin{pmatrix} \phi_{d1,11} \\ \phi_{d1,12} \end{pmatrix}' - \begin{pmatrix} u_{d1} \\ u_{d2} \end{pmatrix}'\right]\right] = 0_{2\times 2}, \\ E_{d30} &= E\Big[h'_{d\theta}(\hat\theta_{kd\_} - \theta)\big(\hat u^{bp}_d - u_d\big)'\Big] = E\left[h'_{d\theta}(\hat\theta_{kd\_} - \theta)\left\{\frac{u_{d1} + e_{d1}}{\sigma^2_{ed1}} \begin{pmatrix} \phi_{d1,11} \\ \phi_{d1,12} \end{pmatrix} - \begin{pmatrix} u_{d1} \\ u_{d2} \end{pmatrix}\right\}'\right] \\ &= E_{(u_{d\_}, e_{d\_})}\left[h'_{d\theta}(\hat\theta_{kd\_} - \theta) E_{(u_d, e_{d1})}\left[\frac{u_{d1} + e_{d1}}{\sigma^2_{ed1}} \begin{pmatrix} \phi_{d1,11} \\ \phi_{d1,12} \end{pmatrix}' - \begin{pmatrix} u_{d1} \\ u_{d2} \end{pmatrix}'\right]\right] = 0_{2\times 2}. \end{aligned}$$

Therefore, we take $S_{d3} = 0_{2\times 2}$.

Let us assume that $\|\hat\beta - \beta\|_2 = O_p(D^{-1/2})$ and $\|\hat\theta - \theta\|_2 = O_p(D^{-1/2})$. Then

$$O_{2\times 2,p}(\|\hat\theta - \theta\|_2^2) = \frac{O_{2\times 2,p}(\|\hat\theta - \theta\|_2^2)}{\|\hat\theta - \theta\|_2^2} \frac{\|\hat\theta - \theta\|_2^2}{D} D = O_{2\times 2,p}(1) O_{2\times 2,p}(1) O_{2\times 2,p}(D^{-1}) = O_{2\times 2,p}(D^{-1}).$$

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

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

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

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

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

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

Note that when using ML instead of REML estimation an additional bias correction term has to be considered, see Datta and Lahiri (2000) for further details.

3.2.1 *Matrix Calculations*

The vectors $h_{d\beta_k,a}$, $k,a=1,2$, are

$$\begin{aligned} h_{d\beta_1,1} &= \underset{1\le j\le p_1}{\mathrm{col}}\big(h_{d\beta_{1j},1}\big)=\frac{\sigma_{ed1}^2-\phi_{d1,11}}{\sigma_{ed1}^2}\underset{1\le j\le p_1}{\mathrm{col}}(x_{d1j})=\frac{\sigma_{ed1}^2-\phi_{d1,11}}{\sigma_{ed1}^2}\,x_{d1},\\ h_{d\beta_1,2} &= -\frac{\phi_{d1,12}}{\sigma_{ed1}^2}\,x_{d1},\quad h_{d\beta_2,1}=0_{p_2\times 1},\quad h_{d\beta_2,2}=x_{d2}. \end{aligned}$$

The matrices $H_{d\beta_{k_1}\beta_{k_2},ab}=h_{d\beta_{k_1},a}h'_{d\beta_{k_2},b}$, $k_1,k_2,a,b=1,2$ are

$$H_{d\beta_1\beta_1,11}=\frac{(\sigma_{ed1}^2-\phi_{d1,11})^2}{\sigma_{ed1}^4}\,x_{d1}x'_{d1},\;\; H_{d\beta_1\beta_1,12}=-\frac{\phi_{d1,12}(\sigma_{ed1}^2-\phi_{d1,11})}{\sigma_{ed1}^4}\,x_{d1}x'_{d1},$$

$$H_{d\beta_1\beta_1,22}=\frac{\phi_{d1,12}^2}{\sigma_{ed1}^4}\,x_{d1}x'_{d1},\;\; H_{d\beta_2\beta_2,11}=0_{p_2\times p_2},\;\; H_{d\beta_2\beta_2,12}=0_{p_2\times p_2},$$

$$H_{d\beta_2\beta_2,22}=x_{d2}x'_{d2},\;\; H_{d\beta_1\beta_2,11}=0_{p_1\times p_2},\;\; H_{d\beta_1\beta_2,12}=\frac{\sigma_{ed1}^2-\phi_{d1,11}}{\sigma_{ed1}^2}\,x_{d1}x'_{d2},$$

$$H_{d\beta_1\beta_2,21}=0_{p_1\times p_2},\;\; H_{d\beta_1\beta_2,22}=-\frac{\phi_{d1,12}}{\sigma_{ed1}^2}\,x_{d1}x'_{d2},\;\; H_{d\beta_2\beta_1,11}=0_{p_2\times p_1},\;\; H_{d\beta_2\beta_1,12}=0_{p_2\times p_1},$$

$$H_{d\beta_2\beta_1,21}=\frac{\sigma_{ed1}^2-\phi_{d1,11}}{\sigma_{ed1}^2}\,x_{d2}x'_{d1},\; H_{d\beta_2\beta_1,22}=-\frac{\phi_{d1,12}}{\sigma_{ed1}^2}\,x_{d2}x'_{d1},$$

and $H_{d\beta_k\beta_k,12}=H_{d\beta_k\beta_k,21}$, $k=1,2$.
The vectors $g_{d\theta,a}$, $a=1,2$, are $g_{d\theta,1}=\underset{1\le\ell\le 3}{\mathrm{col}}\big(\phi_{d1\ell,11}\big)$, $g_{d\theta,2}=\underset{1\le\ell\le 3}{\mathrm{col}}\big(\phi_{d1\ell,12}\big)$ and the corresponding matrices are $G_{d\theta\theta,ab}=g_{d\theta,a}g'_{d\theta,b}$, $a,b=1,2$.

3.3 *Empirical Best Predictors in Domains of Group* $\mathbb{D}_3$

Let us consider the group $\mathbb{D}_3$. The MSE of the EBP is

$$\begin{aligned} \mathrm{MSE}(\hat\mu_d^{ebp}) &= E\big[(\hat\mu_d^{ebp}-\mu_d)(\hat\mu_d^{ebp}-\mu_d)'\big]=E\big[(\hat\mu_d^{bp}-\mu_d)(\hat\mu_d^{bp}-\mu_d)'\big]\\ &+ E\big[(\hat\mu_d^{ebp}-\hat\mu_d^{bp})(\hat\mu_d^{ebp}-\hat\mu_d^{bp})'\big]+2E\big[(\hat\mu_d^{ebp}-\hat\mu_d^{bp})(\hat\mu_d^{bp}-\mu_d)'\big]. \end{aligned}$$

The first summand is $S_{d1}=E\big[(\hat\mu_d^{bp}-\mu_d)(\hat\mu_d^{bp}-\mu_d)'\big]=\mathrm{MSE}(\hat\mu_d^{bp})$.
Let us now approximate the second summand $S_{d2}=E\big[(\hat\mu_d^{ebp}-\hat\mu_d^{bp})(\hat\mu_d^{ebp}-\hat\mu_d^{bp})'\big]$. The EBP is a function of the estimators $(\hat\beta,\hat\theta)$ and of the target variable y_d. For the sake of brevity, we write

$$h_d(\hat\beta,\hat\theta)\triangleq\hat\mu_d^{ebp}=X_d\hat\beta+\hat\Phi_dV_{ed}^{-1}(y_d-X_d\hat\beta),$$

where

$$\hat\Phi_d=\Phi_d(\hat\theta)=\Big(V_{ed}^{-1}+\hat V_{ud}^{-1}\Big)^{-1},\quad \hat V_{ud}=V_{ud}(\hat\theta)=\begin{pmatrix}\hat\sigma_{u1}^2 & \hat\rho\hat\sigma_{u1}\hat\sigma_{u2}\\ \hat\rho\hat\sigma_{u1}\hat\sigma_{u2} & \hat\sigma_{u2}^2\end{pmatrix}.$$

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

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

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

$$\begin{aligned}
\frac{\partial h_d}{\partial \beta_{1j}} &= \begin{pmatrix} x_{d1j} \\ 0 \end{pmatrix} - \Phi_d V_{ed}^{-1} \begin{pmatrix} x_{d1j} \\ 0 \end{pmatrix} = \left(I - \Phi_d V_{ed}^{-1}\right) \begin{pmatrix} x_{d1j} \\ 0 \end{pmatrix} \triangleq \begin{pmatrix} h_{d\beta_{1j},1} \\ h_{d\beta_{1j},2} \end{pmatrix}, \\
\frac{\partial h_d}{\partial \beta_{2j}} &= \begin{pmatrix} 0 \\ x_{d2j} \end{pmatrix} - \Phi_d V_{ed}^{-1} \begin{pmatrix} 0 \\ x_{d2j} \end{pmatrix} = \left(I - \Phi_d V_{ed}^{-1}\right) \begin{pmatrix} 0 \\ x_{d2j} \end{pmatrix} \triangleq \begin{pmatrix} h_{d\beta_{2j},1} \\ h_{d\beta_{2j},2} \end{pmatrix}, \\
\frac{\partial h_d}{\partial \theta_\ell} &= \frac{\partial \Phi_d}{\partial \theta_\ell} V_{ed}^{-1} (y_d - X_d\beta) \triangleq \begin{pmatrix} h_{d\theta_\ell,1} \\ h_{d\theta_\ell,2} \end{pmatrix}.
\end{aligned}$$

For $k, k_1, k_2 = 1,2$, the $p_k \times 1$ vectors containing the derivatives with respect to β_k and the corresponding $p_{k_1} \times p_{k_2}$ matrices are

$$h_{d\beta_k,1} = \underset{1\le j\le p_k}{\mathrm{col}}\left(h_{d\beta_{kj},1}\right),\;\; h_{d\beta_k,2} = \underset{1\le j\le p_k}{\mathrm{col}}\left(h_{d\beta_{kj},2}\right),\;\; H_{d\beta_{k_1}\beta_{k_2},ab} = h_{d\beta_{k_1},a} h'_{d\beta_{k_2},b},\;\; a,b = 1,2.$$

The 3×1 vectors containing the derivatives with respect to θ are

$$\begin{aligned}
h_{d\theta,1} &= \underset{1\le \ell\le 3}{\mathrm{col}}\left(h_{d\theta_\ell,1}\right) = \frac{y_{d1} - x'_{d1}\beta_1}{\sigma^2_{ed1}} g_{d\theta,11} + \frac{y_{d2} - x'_{d2}\beta_2}{\sigma^2_{ed2}} g_{d\theta,12}, \\
h_{d\theta,2} &= \underset{1\le \ell\le 3}{\mathrm{col}}\left(h_{d\theta_\ell,2}\right) = \frac{y_{d1} - x'_{d1}\beta_1}{\sigma^2_{ed1}} g_{d\theta,12} + \frac{y_{d2} - x'_{d2}\beta_2}{\sigma^2_{ed2}} g_{d\theta,22}, \\
g_{d\theta,11} &= \underset{1\le \ell\le 3}{\mathrm{col}}\left(\phi_{d\ell,11}\right),\;\; g_{d\theta,12} = \underset{1\le \ell\le 3}{\mathrm{col}}\left(\phi_{d\ell,12}\right),\;\; g_{d\theta,22} = \underset{1\le \ell\le 3}{\mathrm{col}}\left(\phi_{d\ell,22}\right)
\end{aligned}$$

and the corresponding 3×3 matrices are

$$\begin{aligned}
H_{d\theta\theta,11} &= h_{d\theta,1} h'_{d\theta,1} = \frac{(y_{d1} - x'_{d1}\beta_1)^2}{\sigma^4_{ed1}}\, g_{d\theta,11} g'_{d\theta,11} + \frac{(y_{d2} - x'_{d2}\beta_2)^2}{\sigma^4_{ed2}}\, g_{d\theta,12} g'_{d\theta,12} \\
&+ \frac{(y_{d1} - x'_{d1}\beta_1)(y_{d2} - x'_{d2}\beta_2)}{\sigma^2_{ed1}\sigma^2_{ed2}} \left(g_{d\theta,11} g'_{d\theta,12} + g_{d\theta,12} g'_{d\theta,11}\right),
\end{aligned}$$

$$\begin{aligned}
H_{d\theta\theta,12} &= H'_{d\theta\theta,21} = h_{d\theta,1} h'_{d\theta,2} = \frac{(y_{d1} - x'_{d1}\beta_1)^2}{\sigma^4_{ed1}} g_{d\theta,11} g'_{d\theta,12} + \frac{(y_{d2} - x'_{d2}\beta_2)^2}{\sigma^4_{ed2}} g_{d\theta,12} g'_{d\theta,22} \\
&+ \frac{(y_{d1} - x'_{d1}\beta_1)(y_{d2} - x'_{d2}\beta_2)}{\sigma^2_{ed1}\sigma^2_{ed2}} \left(g_{d\theta,11} g'_{d\theta,22} + g_{d\theta,12} g'_{d\theta,12}\right), \\
H_{d\theta\theta,22} &= h_{d\theta,2} h'_{d\theta,2} = \frac{(y_{d1} - x'_{d1}\beta_1)^2}{\sigma^4_{ed1}}\, g_{d\theta,12} g'_{d\theta,12} + \frac{(y_{d2} - x'_{d2}\beta_2)^2}{\sigma^4_{ed2}}\, g_{d\theta,22} g'_{d\theta,22} \\
&+ \frac{(y_{d1} - x'_{d1}\beta_1)(y_{d2} - x'_{d2}\beta_2)}{\sigma^2_{ed1}\sigma^2_{ed2}} \left(g_{d\theta,12} g'_{d\theta,22} + g_{d\theta,22} g'_{d\theta,12}\right).
\end{aligned}$$

The crossed $p_k \times 3$ matrices, $H_{d\beta_k\theta,ab} = h_{d\beta_k,a}h'_{d\theta,b}$, $k,a,b=1,2$, are

$$
\begin{aligned}
H_{d\beta_k\theta,a1} &= \frac{y_{d1}-x'_{d1}\beta_1}{\sigma^2_{ed1}} h_{d\beta_k,a}g'_{d\theta,11} + \frac{y_{d2}-x'_{d2}\beta_2}{\sigma^2_{ed2}} h_{d\beta_k,a}g'_{d\theta,12}, \; a=1,2, \\
H_{d\beta_k\theta,a2} &= \frac{y_{d1}-x'_{d1}\beta_1}{\sigma^2_{ed1}} h_{d\beta_k,a}g'_{d\theta,12} + \frac{y_{d2}-x'_{d2}\beta_2}{\sigma^2_{ed2}} h_{d\beta_k,a}g'_{d\theta,22}, \; a=1,2.
\end{aligned}
$$

We further define the $p_k \times 2$ and 3×2 matrices

$$
h_{d\theta} = \text{row}(h_{d\theta,1}, h_{d\theta,2}), \quad h_{d\beta_k} = \text{row}(h_{d\beta_k,1}, h_{d\beta_k,2}), \quad k=1,2.
$$

A Taylor series expansion of the 2×1 vector $h_d(\hat{\beta},\hat{\theta})$ around (β,θ) yields

$$
\begin{aligned}
h_d(\hat{\beta},\hat{\theta}) &= h_d(\beta,\theta) + \sum_{k=1}^{2}\sum_{j=1}^{p_k} \frac{\partial h'_d(\beta,\theta)}{\partial \beta_{kj}}(\hat{\beta}_{kj}-\beta_{kj}) + \sum_{\ell=1}^{3} \frac{\partial h'_d(\beta,\theta)}{\partial \theta_\ell}(\hat{\theta}_\ell - \theta_\ell) \\
&+ o_p(\|\hat{\beta}-\beta\|_2) + o_p(\|\hat{\theta}-\theta\|_2).
\end{aligned}
$$

Therefore,

$$
h_d(\hat{\beta},\hat{\theta}) = h_d(\beta,\theta) + h'_{d\beta_1}(\hat{\beta}_1-\beta_1) + h'_{d\beta_2}(\hat{\beta}_2-\beta_2) + h'_{d\theta}(\hat{\theta}-\theta) + o_p(\|\hat{\beta}-\beta\|_2) + o_p(\|\hat{\theta}-\theta\|_2)
$$

and

$$
\begin{aligned}
S_{d2} &= E\big[(\hat{\mu}_d^{ebp}-\hat{\mu}_d^{bp})(\hat{\mu}_d^{ebp}-\hat{\mu}_d^{bp})'\big] = E\big[(h_d(\hat{\beta},\hat{\theta})-h_d(\beta,\theta))(h_d(\hat{\beta},\hat{\theta})-h_d(\beta,\theta))'\big] \\
&= E\big[h'_{d\beta_1}(\hat{\beta}_1-\beta_1)(\hat{\beta}_1-\beta_1)'h_{d\beta_1}\big] + E\big[h'_{d\beta_2}(\hat{\beta}_2-\beta_2)(\hat{\beta}_2-\beta_2)'h_{d\beta_2}\big] \\
&+ E\big[h'_{d\theta}(\hat{\theta}-\theta)(\hat{\theta}-\theta)'h_{d\theta}\big] \\
&+ \big\{E\big[h'_{d\beta_1}(\hat{\beta}_1-\beta_1)(\hat{\beta}_2-\beta_2)'h_{d\beta_2}\big] + E\big[h'_{d\beta_2}(\hat{\beta}_2-\beta_2)(\hat{\beta}_1-\beta_1)'h_{d\beta_1}\big]\big\} \\
&+ \big\{E\big[h'_{d\beta_1}(\hat{\beta}_1-\beta_1)(\hat{\theta}-\theta)'h_{d\theta}\big] + E\big[h'_{d\theta}(\hat{\theta}-\theta)(\hat{\beta}_1-\beta_1)'h_{d\beta_1}\big]\big\} \\
&+ \big\{E\big[h'_{d\beta_2}(\hat{\beta}_2-\beta_2)(\hat{\theta}-\theta)'h_{d\theta}\big] + E\big[h'_{d\theta}(\hat{\theta}-\theta)(\hat{\beta}_2-\beta_2)'h_{d\beta_2}\big]\big\} \\
&+ o_p(\|\hat{\beta}-\beta\|_2^2) + o_p(\|\hat{\theta}-\theta\|_2^2).
\end{aligned}
$$

Concerning the first summand of S_{d2}, $E_{d11} = E\big[h'_{d\beta_1}(\hat{\beta}_1-\beta_1)(\hat{\beta}_1-\beta_1)'h_{d\beta_1}\big]$, we observe that

$$
h'_{d\beta_1}(\hat{\beta}_1-\beta_1)(\hat{\beta}_1-\beta_1)'h_{d\beta_1} = \begin{pmatrix} h'_{d\beta_1,1}(\hat{\beta}_1-\beta_1)(\hat{\beta}_1-\beta_1)'h_{d\beta_1,1} & h'_{d\beta_1,1}(\hat{\beta}_1-\beta_1)(\hat{\beta}_1-\beta_1)'h_{d\beta_1,2} \\ h'_{d\beta_1,2}(\hat{\beta}_1-\beta_1)(\hat{\beta}_1-\beta_1)'h_{d\beta_1,1} & h'_{d\beta_1,2}(\hat{\beta}_1-\beta_1)(\hat{\beta}_1-\beta_1)'h_{d\beta_1,2} \end{pmatrix}.
$$

Note that $h'_{d\beta_1,1}(\hat{\beta}_1-\beta_1)$ and $h'_{d\beta_1,2}(\hat{\beta}_1-\beta_1)$ are scalars. Then, for $a,b=1,2$, we have

$$
h'_{d\beta_1,a}(\hat{\beta}_1-\beta_1)(\hat{\beta}_1-\beta_1)'h_{d\beta_1,b} = (\hat{\beta}_1-\beta_1)'h_{d\beta_1,b}h'_{d\beta_1,a}(\hat{\beta}_1-\beta_1) = (\hat{\beta}_1-\beta_1)'H_{d\beta_1\beta_1,ba}(\hat{\beta}_1-\beta_1).
$$

Therefore,

$$
E_{d11} = E\left[\begin{pmatrix} (\hat{\beta}_1-\beta_1)'H_{d\beta_1\beta_1,11}(\hat{\beta}_1-\beta_1) & (\hat{\beta}_1-\beta_1)'H_{d\beta_1\beta_1,21}(\hat{\beta}_1-\beta_1) \\ (\hat{\beta}_1-\beta_1)'H_{d\beta_1\beta_1,12}(\hat{\beta}_1-\beta_1) & (\hat{\beta}_1-\beta_1)'H_{d\beta_1\beta_1,22}(\hat{\beta}_1-\beta_1) \end{pmatrix}\right].
$$

We apply to $z = \hat\beta_1 - \beta_1$ the formula

$$E[z'Qz] = \operatorname{tr}\{Q \operatorname{var}(z)\} + E[z]'QE[z].$$

As the matrices $H_{d\beta_1\beta_1,ab}$ are not random and $E[\hat\beta_1 - \beta_1] = O(\|\hat\beta_1 - \beta_1\|_2)$, we obtain

$$E_{d11} = \begin{pmatrix} \operatorname{tr}\{H_{d\beta_1\beta_1,11} \operatorname{var}(\hat\beta_1)\} & \operatorname{tr}\{H_{d\beta_1\beta_1,21} \operatorname{var}(\hat\beta_1)\} \\ \operatorname{tr}\{H_{d\beta_1\beta_1,12} \operatorname{var}(\hat\beta_1)\} & \operatorname{tr}\{H_{d\beta_1\beta_1,22} \operatorname{var}(\hat\beta_1)\} \end{pmatrix} + O_{2\times 2}(\|\hat\beta_1 - \beta_1\|_2^2).$$

For the second summand of S_{d2}, $E_{d22} = E\big[h'_{d\beta_2}(\hat\beta_2 - \beta_2)(\hat\beta_2 - \beta_2)'h_{d\beta_2}\big]$, we have

$$E_{d22} = \begin{pmatrix} \operatorname{tr}\{H_{d\beta_2\beta_2,11} \operatorname{var}(\hat\beta_2)\} & \operatorname{tr}\{H_{d\beta_2\beta_2,21} \operatorname{var}(\hat\beta_2)\} \\ \operatorname{tr}\{H_{d\beta_2\beta_2,12} \operatorname{var}(\hat\beta_2)\} & \operatorname{tr}\{H_{d\beta_2\beta_2,22} \operatorname{var}(\hat\beta_2)\} \end{pmatrix} + O_{2\times 2}(\|\hat\beta_2 - \beta_2\|_2^2).$$

The fourth summand of S_{d2} is $E_{d12}+E_{d21}$, where $E'_{d21} = E_{d12} = E\big[h'_{d\beta_1}(\hat\beta_1-\beta_1)(\hat\beta_2-\beta_2)'h_{d\beta_2}\big]$,

$$E_{d12} = \begin{pmatrix} \operatorname{tr}\{H_{d\beta_2\beta_1,11} \operatorname{cov}(\hat\beta_1, \hat\beta_2)\} & \operatorname{tr}\{H_{d\beta_2\beta_1,21} \operatorname{cov}(\hat\beta_1, \hat\beta_2)\} \\ \operatorname{tr}\{H_{d\beta_2\beta_1,12} \operatorname{cov}(\hat\beta_1, \hat\beta_2)\} & \operatorname{tr}\{H_{d\beta_2\beta_1,22} \operatorname{cov}(\hat\beta_1, \hat\beta_2)\} \end{pmatrix} + O_{2\times 2}(\|\hat\beta_1-\beta_1\|_2\|\hat\beta_2-\beta_2\|_2).$$

The third summand of S_{d2}, $E_{d00} = E\big[h'_{d\theta}(\hat\theta - \theta)(\hat\theta - \theta)'h_{d\theta}\big]$, is

$$E_{d00} = E\left[\begin{pmatrix} (\hat\theta - \theta)'H_{d\theta\theta,11}(\hat\theta - \theta) & (\hat\theta - \theta)'H_{d\theta\theta,21}(\hat\theta - \theta) \\ (\hat\theta - \theta)'H_{d\theta\theta,12}(\hat\theta - \theta) & (\hat\theta - \theta)'H_{d\theta\theta,22}(\hat\theta - \theta) \end{pmatrix}\right].$$

We recall that $H_{d\theta\theta,ab}$ is random. This is why we calculate E_{d00} for the estimators $\hat\theta_{d_-}$ and $\hat\beta_{kd_-}$, $k = 1, 2$, based on $y_{d_-} = \big(y_{d'}\big)_{d'\neq d}$. As the target data y can be split into the independent parts y_{d_-} and y_d, for $a, b = 1$ we have

$$\begin{aligned}
E_{d_-00,11} &= E\big[(\hat\theta_{d_-} - \theta)'H_{d\theta\theta,11}(\hat\theta_{d_-} - \theta)\big] = E_{y_{d_-}}\big[E_{y_d}\big[(\hat\theta_{d_-} - \theta)'H_{d\theta\theta,11}(\hat\theta_{d_-} - \theta)\big]\big] \\
&= E_{y_{d_-}}\big[(\hat\theta_{d_-} - \theta)'g_{d\theta,11}g'_{d\theta,11}(\hat\theta_{d_-} - \theta)\sigma_{ed1}^{-4}E_{y_d}\big[(y_{d1} - x'_{d1}\beta_1)^2\big]\big] \\
&+ E_{y_{d_-}}\big[(\hat\theta_{d_-} - \theta)'g_{d\theta,12}g'_{d\theta,12}(\hat\theta_{d_-} - \theta)\sigma_{ed2}^{-4}E_{y_d}\big[(y_{d2} - x'_{d2}\beta_2)^2\big]\big] \\
&+ E_{y_{d_-}}\big[(\hat\theta_{d_-} - \theta)'\big(g_{d\theta,11}g'_{d\theta,12} + g_{d\theta,12}g'_{d\theta,11}\big)(\hat\theta_{d_-} - \theta)\sigma_{ed1}^{-2}\sigma_{ed2}^{-2}E_{y_d}\big[(y_{d1} - x'_{d1}\beta_1)(y_{d2} - x'_{d2}\beta_2)\big]\big] \\
&= \frac{\sigma_{ud1}^2 + \sigma_{ed1}^2}{\sigma_{ed1}^4}\operatorname{tr}\{g_{d\theta,11}g'_{d\theta,11}\operatorname{var}(\hat\theta_{d_-})\} + \frac{\sigma_{ud2}^2 + \sigma_{ed2}^2}{\sigma_{ed2}^4}\operatorname{tr}\{g_{d\theta,12}g'_{d\theta,12}\operatorname{var}(\hat\theta_{d_-})\} \\
&+ \frac{\rho\sigma_{ud1}\sigma_{ud2} + \sigma_{ed12}}{\sigma_{ed1}^2\sigma_{ed2}^2}\operatorname{tr}\{\big(g_{d\theta,11}g'_{d\theta,12} + g_{d\theta,12}g'_{d\theta,11}\big)\operatorname{var}(\hat\theta_{d_-})\} + O_{2\times 2}(\|\hat\theta_{d_-} - \theta\|_2^2).
\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}
E_{d_-00,12} &= E_{d_-00,21} = \frac{\sigma_{ud1}^2 + \sigma_{ed1}^2}{\sigma_{ed1}^4}\operatorname{tr}\{g_{d\theta,11}g'_{d\theta,12}\operatorname{var}(\hat\theta_{d_-})\} + \frac{\sigma_{ud2}^2 + \sigma_{ed2}^2}{\sigma_{ed2}^4}\operatorname{tr}\{g_{d\theta,12}g'_{d\theta,22}\operatorname{var}(\hat\theta_{d_-})\} \\
&+ \frac{\rho\sigma_{ud1}\sigma_{ud2} + \sigma_{ed12}}{\sigma_{ed1}^2\sigma_{ed2}^2}\operatorname{tr}\{\big(g_{d\theta,11}g'_{d\theta,22} + g_{d\theta,12}g'_{d\theta,12}\big)\operatorname{var}(\hat\theta_{d_-})\} + O_{2\times 2}(\|\hat\theta_{d_-} - \theta\|_2^2), \\
E_{d_-00,22} &= \frac{\sigma_{ud1}^2 + \sigma_{ed1}^2}{\sigma_{ed1}^4}\operatorname{tr}\{g_{d\theta,12}g'_{d\theta,12}\operatorname{var}(\hat\theta_{d_-})\} + \frac{\sigma_{ud2}^2 + \sigma_{ed2}^2}{\sigma_{ed2}^4}\operatorname{tr}\{g_{d\theta,22}g'_{d\theta,22}\operatorname{var}(\hat\theta_{d_-})\} \\
&+ \frac{\rho\sigma_{ud1}\sigma_{ud2} + \sigma_{ed12}}{\sigma_{ed1}^2\sigma_{ed2}^2}\operatorname{tr}\{\big(g_{d\theta,12}g'_{d\theta,22} + g_{d\theta,22}g'_{d\theta,12}\big)\operatorname{var}(\hat\theta_{d_-})\} + O_{2\times 2}(\|\hat\theta_{d_-} - \theta\|_2^2).
\end{aligned}$$

Taken out (y_d, X_d) from the data file should not have great influence on $\text{var}(\hat{\theta})$. Therefore, we approximate $E_{d00,ab}$ by substituting $\hat{\theta}_{d-}$ by $\hat{\theta}_d$ in $E_{d-00,ab}$, i.e.

$$E_{d00} = \begin{pmatrix} E_{d00,11} & E_{d00,12} \\ E_{d00,21} & E_{d00,22} \end{pmatrix} + O_{2\times 2}(\|\hat{\theta}-\theta\|_2^2).$$

The fifth and sixth summands of S_{d2} are $E_{d10} + E_{d01}$ and $E_{d20} + E_{d02}$ respectively, where $E'_{d0k} = E_{dk0} = E\big[h'_{d\beta_k}(\hat{\beta}_k - \beta_k)(\hat{\theta}-\theta)' h_{d\theta}\big]$, $k = 1, 2$, and

$$E_{dk0} = E\left[\begin{pmatrix} (\hat{\beta}_k - \beta_k)' H_{d\beta_k\theta,11}(\hat{\theta}-\theta) & (\hat{\beta}_k - \beta_k)' H_{d\beta_k\theta,21}(\hat{\theta}-\theta) \\ (\hat{\beta}_k - \beta_k)' H_{d\beta_k\theta,12}(\hat{\theta}-\theta) & (\hat{\beta}_k - \beta_k)' H_{d\beta_k\theta,22}(\hat{\theta}-\theta) \end{pmatrix}\right].$$

We recall that $H_{d\beta_k\theta,ab}$ is random. For $b = 1$, it holds that

$$\begin{aligned} E_{d-k0,a1} &= E\big[(\hat{\beta}_{kd-} - \beta)' H_{d\beta_k\theta,a1}(\hat{\theta}_{d-} - \theta)\big] = E_{y_{d-}}\big[E_{y_d}\big[(\hat{\beta}_{kd-} - \beta)' H_{d\beta_k\theta,a1}(\hat{\theta}_{d-} - \theta)\big]\big] \\ &= E_{y_{d-}}\left[(\hat{\beta}_{kd-} - \beta)' h_{d\beta_k,a} g'_{d\theta,11}(\hat{\theta}_{d-} - \theta)\sigma_{ed1}^{-2} E_{y_{d_1}}\left[y_{d1} - x'_{d1}\beta_1\right]\right] \\ &+ E_{y_{d-}}\left[(\hat{\beta}_{kd-} - \beta)' h_{d\beta_k,a} g'_{d\theta,12}(\hat{\theta}_{d-} - \theta)\sigma_{ed2}^{-2} E_{y_{d_2}}\left[y_{d2} - x'_{d2}\beta_2\right]\right] = 0. \end{aligned}$$

Similarly, for $b = 2$ we get $E_{d-k0,a2} = 0$, $a = 1, 2$. Therefore, we take $E_{dk0} = 0_{2\times 2}$, $k = 1, 2$.

Let us now approximate the third summand $S_{d3} = E\big[(\hat{\mu}_d^{ebp} - \hat{\mu}_d^{bp})(\hat{\mu}_d^{bp} - \mu_d)'\big]$. We recall that

$$\hat{\mu}_d^{ebp} - \hat{\mu}_d^{bp} = h'_{d\beta_1}(\hat{\beta}_1 - \beta_1) + h'_{d\beta_2}(\hat{\beta}_2 - \beta_2) + h'_{d\theta}(\hat{\theta}-\theta) + o_p(\|\hat{\beta}-\beta\|_2) + o_p(\|\hat{\theta}-\theta\|_2),$$

$$\hat{\mu}_d^{bp} - \mu_d = \hat{u}_d^{bp} - u_d = \Phi_d V_{ed}^{-1}(y_d - X_d\beta) - u_d = \big(\Phi_d V_{ed}^{-1} - I_{2\times 2}\big)u_d + \Phi_d V_{ed}^{-1} e_d.$$

It holds that

$$\begin{aligned} E_{d-3k} &= E\Big[h'_{d\beta_k}(\hat{\beta}_{kd-} - \beta_k)\big(\hat{u}_d^{bp} - u_d\big)'\Big] \\ &= E\Big[h'_{d\beta_k}(\hat{\beta}_{kd-} - \beta_k)\big\{\big(\Phi_d V_{ed}^{-1} - I_{2\times 2}\big)u_d + \Phi_d V_{ed}^{-1} e_d\big\}'\Big] \\ &= E_{(u_{d-}, e_{d-})}\Big[h'_{d\beta_{kd-}}(\hat{\beta}_k - \beta_k)\big\{\big(\Phi_d V_{ed}^{-1} - I_{2\times 2}\big)E_{u_d}[u_d] + \Phi_d V_{ed}^{-1} E_{e_d}[e_d]\big\}'\Big] = 0_{2\times 2}, \\ &\quad k = 1, 2, \\ E_{d-30} &= E\Big[h'_{d\theta}(\hat{\theta}_{kd-} - \theta)\big(\hat{u}_d^{bp} - u_d\big)'\Big] = E\Big[h'_{d\theta}(\hat{\theta}_{kd-} - \theta)\big\{\big(\Phi_d V_{ed}^{-1} - I_{2\times 2}\big)u_d + \Phi_d V_{ed}^{-1} e_d\big\}'\Big] \\ &= E_{(u_{d-}, e_{d-})}\Big[h'_{d\theta}(\hat{\theta}_{kd-} - \theta)\big\{\big(\Phi_d V_{ed}^{-1} - I_{2\times 2}\big)E_{u_d}[u_d] + \Phi_d V_{ed}^{-1} E_{e_d}[e_d]\big\}'\Big] = 0_{2\times 2}. \end{aligned}$$

Therefore, we take $S_{d3} = 0_{2\times 2}$.

We further assume that $\|\hat{\beta}-\beta\|_2 = O_p(D^{-1/2})$ and $\|\hat{\theta}-\theta\|_2 = O_p(D^{-1/2})$.
For $D_1 + D_2 + 1 \le d \le D$, 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

$$
\begin{aligned}
G_{d1}(\theta) &= \Phi_d(\theta)V_{ed}^{-1}(\theta)(V_{ud}(\theta) + V_{ed}(\theta))V_{ed}^{-1}(\theta)\Phi_d(\theta) + V_{ud}(\theta) - 2\Phi_d(\theta)V_{ed}^{-1}(\theta)V_{ud}(\theta), \\
G_{d2,ab}(\theta) &= \begin{pmatrix} \mathrm{tr}\{H_{d\beta_b\beta_a,11}(\theta)\,\mathrm{cov}(\hat{\beta}_a, \hat{\beta}_b)\} & \mathrm{tr}\{H_{d\beta_b\beta_a,21}(\theta)\,\mathrm{cov}(\hat{\beta}_a, \hat{\beta}_b)\} \\ \mathrm{tr}\{H_{d\beta_b\beta_a,12}(\theta)\,\mathrm{cov}(\hat{\beta}_a, \hat{\beta}_b)\} & \mathrm{tr}\{H_{d\beta_b\beta_a,22}(\theta)\,\mathrm{cov}(\hat{\beta}_a, \hat{\beta}_b)\} \end{pmatrix}, \;\; a, b = 1, 2, \\
G_{d3}(\theta) &= \begin{pmatrix} \mathrm{tr}\{H_{d\theta\theta,11}(\theta)\,\mathrm{var}(\hat{\theta})\} & \mathrm{tr}\{H_{d\theta\theta,21}(\theta)\,\mathrm{var}(\hat{\theta})\} \\ \mathrm{tr}\{H_{d\theta\theta,12}(\theta)\,\mathrm{var}(\hat{\theta})\} & \mathrm{tr}\{H_{d\theta\theta,22}(\theta)\,\mathrm{var}(\hat{\theta})\} \end{pmatrix}.
\end{aligned}
$$

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

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

Note that when using ML instead of REML estimation an additional bias correction term has to be considered, see Datta and Lahiri (2000) for further details.

4 Bootstrap Approximations of the Mean Squared Errors of the MBFH Predictor

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

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

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

$$\text{mse}^1(\hat{\mu}_d^{ebp}) = B^{-1}\sum_{b=1}^{B}(\hat{\mu}_d^{*ebp(b)} - \mu_d^{*(b)})(\hat{\mu}_d^{*ebp(b)} - \mu_d^{*(b)})', \quad \forall d \in \{1, \dots, D\}. \tag{4.1}$$

Observe that $\text{mse}^1(\hat{\mu}_d^{*ebp})$ is consistent for $MSE_*(\hat{\mu}_d^{*ebp})$ 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

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

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

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

A Monte Carlo approximation $\text{mse}^a(\hat{\mu}_d^{*ebp})$ of the bootstrap matrix $\text{MSE}_*(\hat{\mu}_d^{*ebp})$, for $a = 2, 3$, can be obtained similarly as (4.1).

5 Approximation of the Mean Squared Error of the Synthetic Fay-Herriot Predictor

Let us consider a Fay-Herriot model

$$y_d = x_d\beta + u_d + e_d, \quad u_d \sim N(0,\sigma_u^2), \quad e_d \sim N(0,\sigma_{ed}^2), \quad d = 1,\dots,D,$$

where the u_d's and e_d's are mutually independent. Let us assume that some of the y_d are missing, while all the x_d are recorded and known. More concretely, the set of domains is partitioned in two subsets $\mathcal{D}_0 = \{1,\dots,D_0\}$ and $\mathcal{D}_1 = \{D_0+1,\dots,D\}$, where $D_0 < D$. We assume that $n_d > 0$ and y_d is observed if $d \in \mathcal{D}_0$, and that y_d is not observed or $n_d = 0$ if $d \in \mathcal{D}_1$. Further, $\sigma_{ed}^2 > 0$ is known for $d \in \mathcal{D}_0$, but it is missing for $d \in \mathcal{D}_1$. For the subset $\mathcal{D}_0$, we consider the vector and matrices

$$X_0 = \underset{1\le d\le D_0}{\operatorname{col}}(x_d), \quad V_0 = \underset{1\le d\le D_0}{\operatorname{diag}}(\sigma_u^2+\sigma_{ed}^2), \quad u_0 = \underset{1\le d\le D_0}{\operatorname{col}}(u_d), \quad e_0 = \underset{1\le d\le D_0}{\operatorname{col}}(e_d).$$

The synthetic predictor is

$$\hat{\mu}_d^{syn} = x_d\hat{\beta}, \quad d = 1,\dots,D,$$

where $\hat{\beta}$ and $\hat{\sigma}_u^2$ are calculated by using the data from $\mathcal{D}_0$. If $\hat{\beta}$ and $\hat{\sigma}_u^2$ are asymptotically consistent and independent estimators of β and σ_u^2, as the ML and REML estimators are, then the mean and variance of $\hat{\mu}_d^{syn}$ are

$$E[\hat{\mu}_d^{syn}] \approx x_d\beta, \quad \operatorname{var}(\hat{\mu}_d^{syn}) = x_d\operatorname{var}(\hat{\beta})x_d' \approx x_d(X_0'V_0^{-1}X)_0^{-1}x_d'.$$

As $\mu_d = x_d\beta + u_d$, the MSE of $\hat{\mu}_d^{syn} = x_d\hat{\beta}$ is

$$\begin{aligned} MSE(\hat{\mu}_d^{syn}) &= E\big[(\hat{\mu}_d^{syn} - \mu_d)^2\big] = E\big[(x_d\hat{\beta} - x_d\beta - u_d)^2\big] = E\big[(x_d(\hat{\beta}-\beta) - u_d)^2\big] \\ &= x_dE\big[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\big]x_d' + E[u_d^2] - 2E\big[x_d(\hat{\beta}-\beta)u_d\big] \\ &\approx x_d(X_0'V_0^{-1}X_0)^{-1}x_d' + \sigma_u^2 - 2E\big[x_d\hat{\beta}u_d\big]. \end{aligned}$$

If $d \in \mathcal{D}_1$, then u_d and $\hat{\beta}$ are independent, so that $E[x_d\hat{\beta}u_d] = x_dE[\hat{\beta}]E[u_d] = 0$ and

$$MSE(\hat{\mu}_d^{syn}) \approx x_d(X_0'V_0^{-1}X_0)^{-1}x_d' + \sigma_u^2. \tag{5.1}$$

6 Additional results of the model-based simulation

The main manuscript presents the simulation results for 20% missing values in each variable. For comparison, this section gives the same plots for 5% and 10% missing values of each variable. Figures 6.1-6.8 show boxplots of the RRMSEs of the predictors MBFH-EBP, FH-EBP, and FH-SYN for different sampling error correlation $\rho_{ed12} \in \{0.6, -0.3, 0\}$.
The non-zero sampling error correlations in Figures 6.1-6.2, 6.4-6.6, and 6.8 correspond to situations where both variables stem from the same survey. A zero sampling error correlation as in Figures 6.3 and 6.7 applies if both variables stem from independent surveys. It is visible for all six Figures 6.1-6.2, 6.4-6.6, and 6.8 that in terms of RRMSE the introduced MBFH-EBP is at least as efficient as the respective FH-EBLUP or FH-SYN. This observation holds for positively, negatively, and uncorrelated sampling errors. Thus, the application of the MBFH instead of the FH estimator is profitable for various combinations of sampling error and random effect correlations.
Further, we note that the benefits of using the MBFH predictor is more visible for higher percentages of missing direct estimates. Another interesting conclusion is that for the target variable with a missing direct estimate the MBFH predictor gives significant performance gains over the corresponding FH synthetic predictor when the variable random effects are highly correlated.

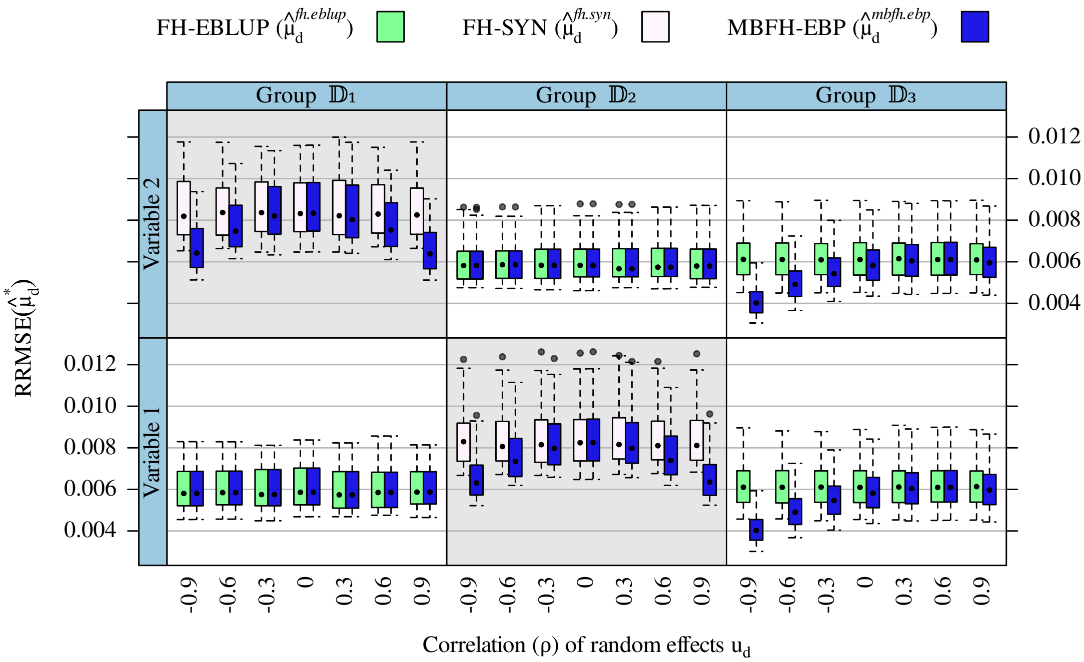

Figure 6.1: RRMSE of point estimates, 5% missing domains, $\rho_{ed12} = 0.6$

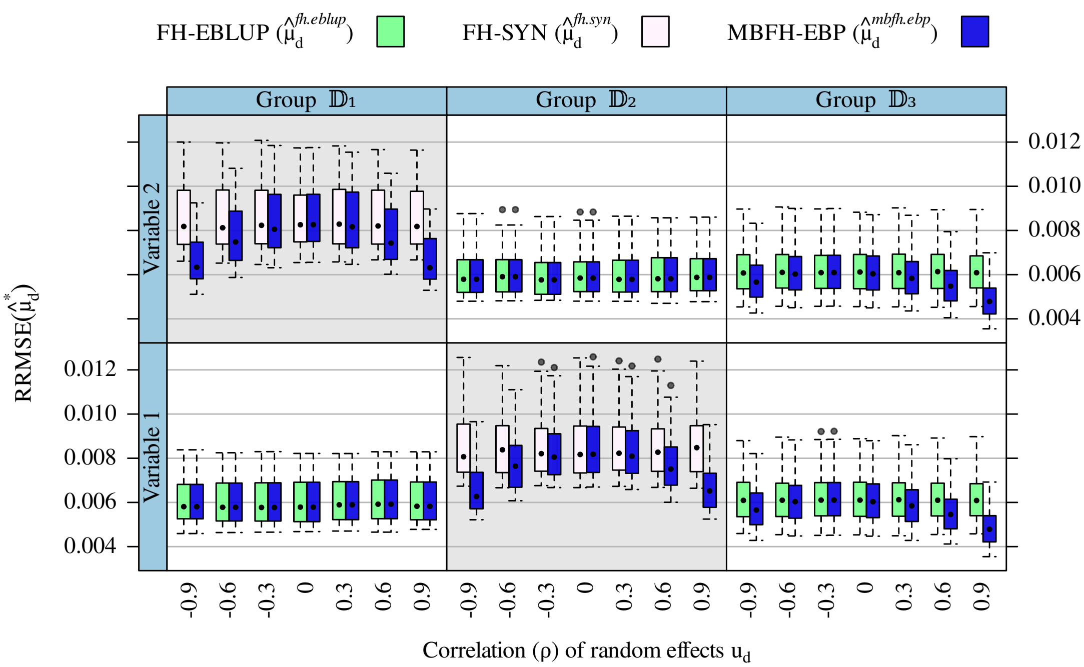

Figure 6.2: RRMSE of point estimates, 5% missing domains, $\rho_{ed12} = -0.3$

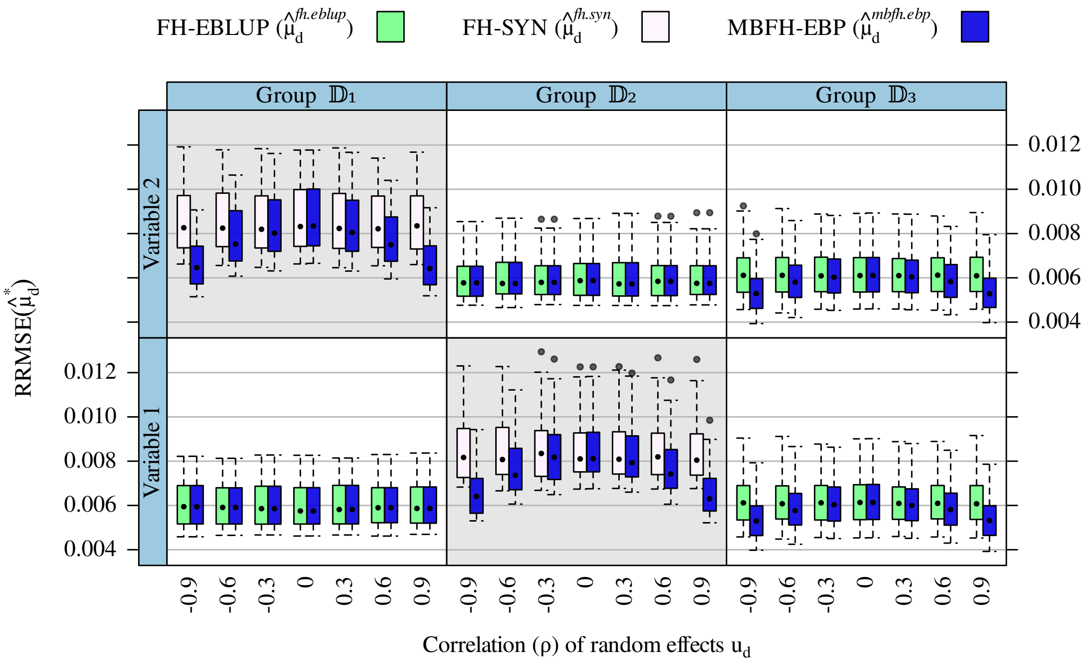

Figure 6.3: RRMSE of point estimates, 5% missing domains, $\rho_{ed12} = 0$

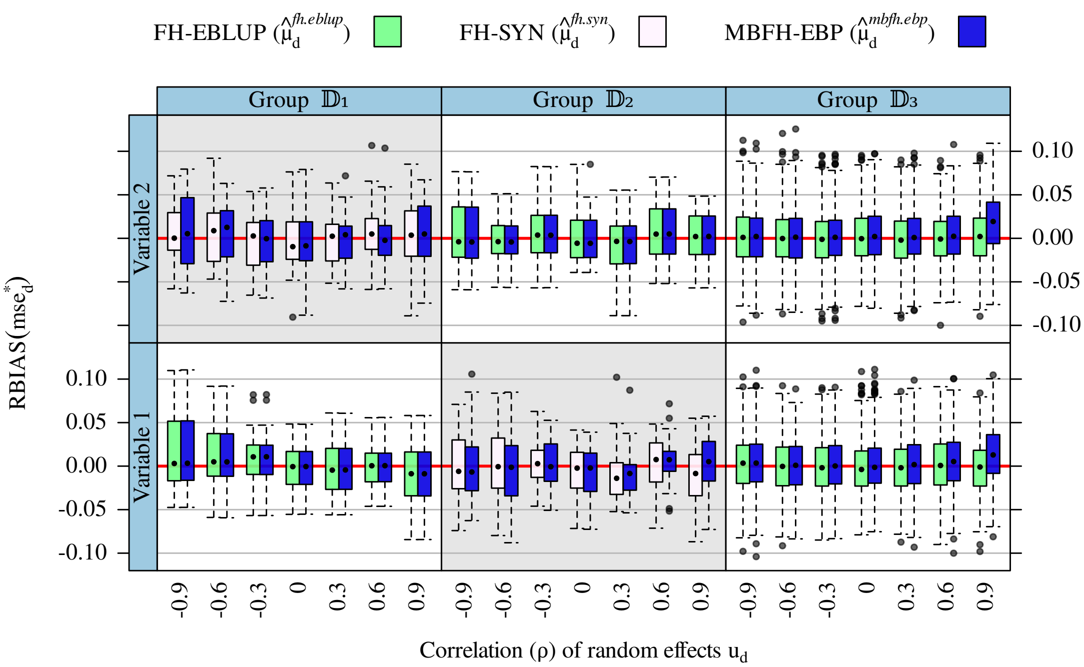

Figure 6.4: RBias of MSE estimates, 5% missing domains, $\rho_{ed12} = -0.3$

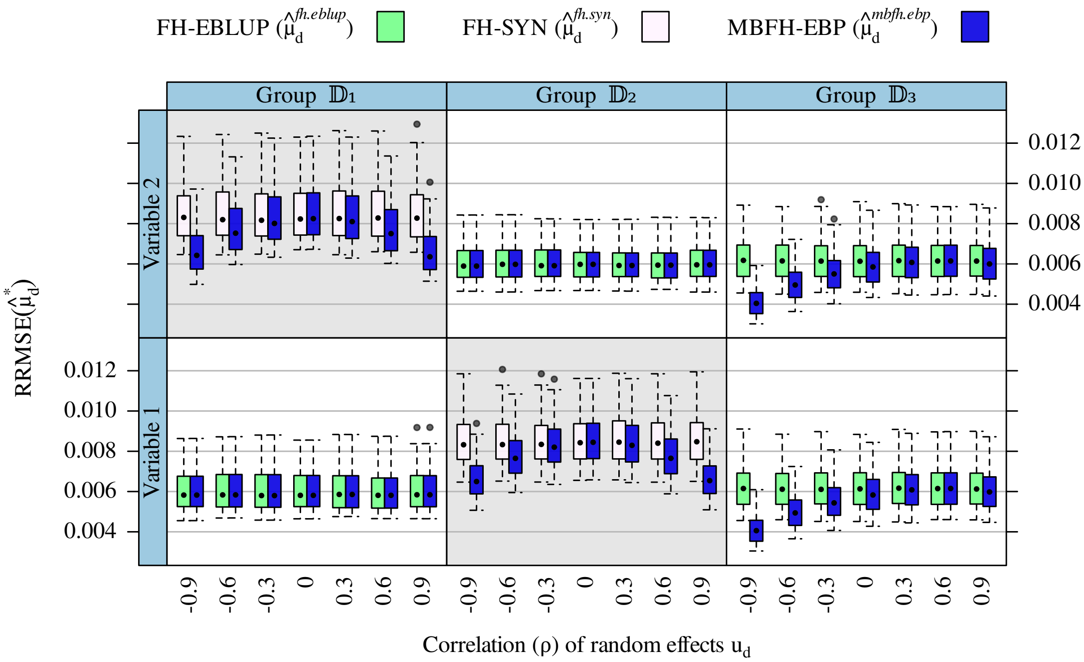

Figure 6.5: RRMSE of point estimates, 10% missing domains, $\rho_{ed12} = 0.6$

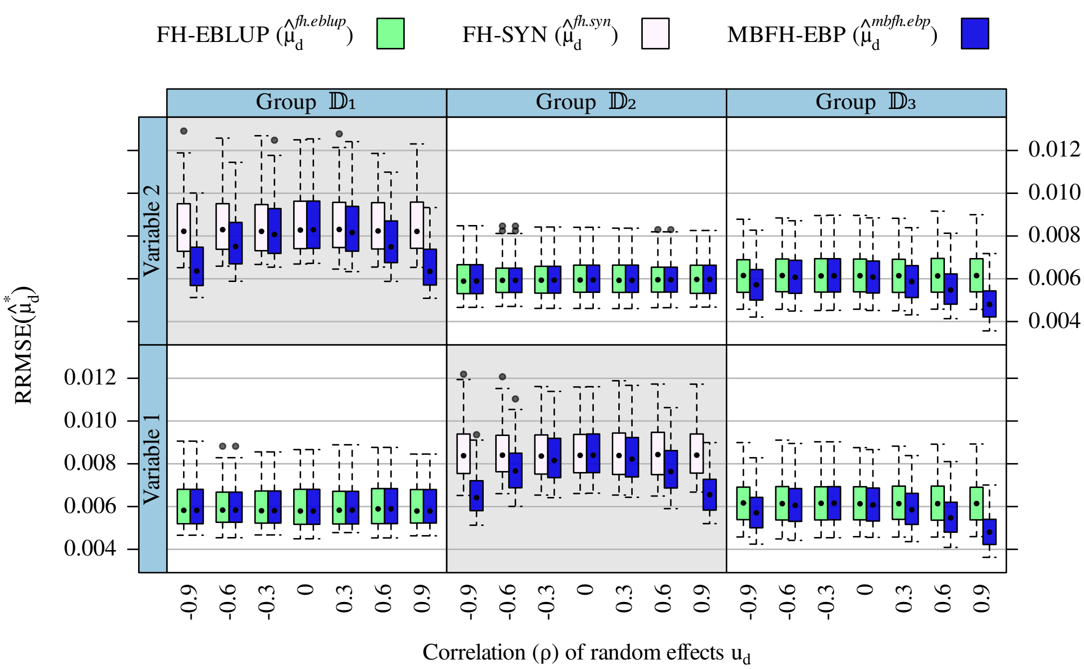

Figure 6.6: RRMSE of point estimates, 10% missing domains, $\rho_{ed12} = -0.3$

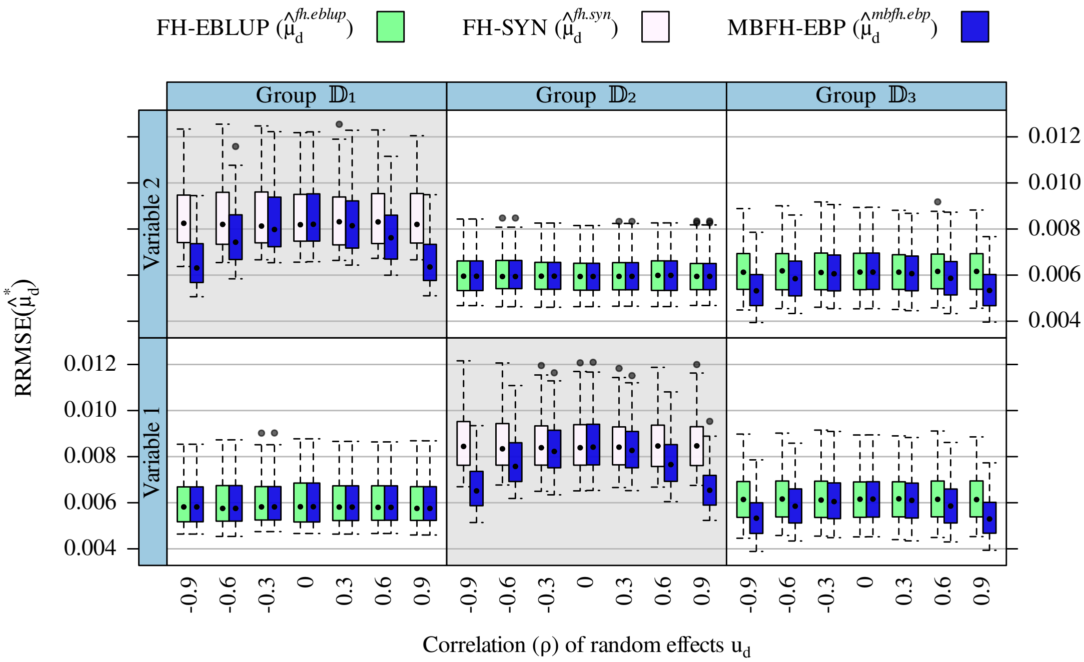

Figure 6.7: RRMSE of point estimates, 10% missing domains, $\rho_{ed12} = 0$

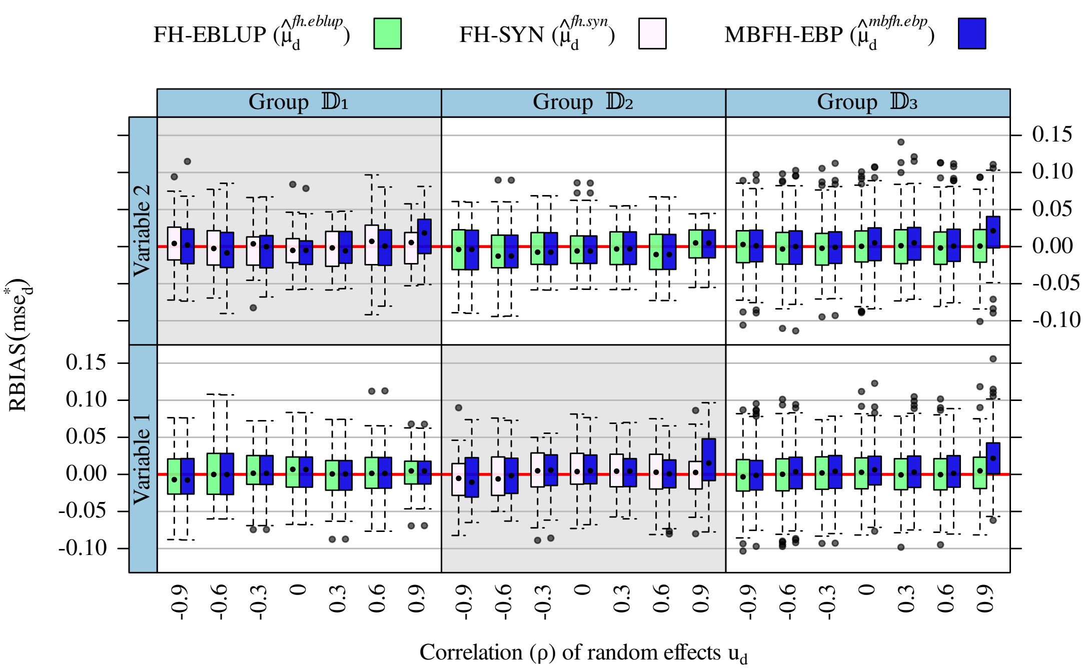

Figure 6.8: RBias of MSE estimates, 10% missing domains, $\rho_{ed12} = -0.3$