Multivariate Fay-Herriot models for small area estimation with application to household consumption per capita expenditure in Indonesia

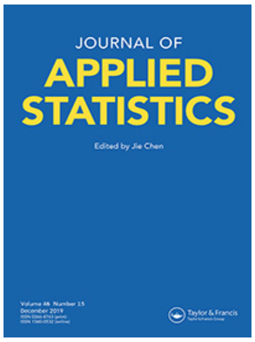
Article *in* Journal of Applied Statistics · May 2019

DOI: 10.1080/02664763.2019.1615420

Azka Ubaidillah
Politeknik Statistika STIS

Khairil Anwar Notodiputro
Bogor Agricultural University

Anang Kurnia
IPB University (Bogor Agricultural University)



Taylor & Francis
Taylor & Francis Group

Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Multivariate Fay-Herriot models for small area estimation with application to household consumption per capita expenditure in Indonesia

Azka Ubaidillah, Khairil Anwar Notodiputro, Anang Kurnia & I. Wayan Mangku

To cite this article: Azka Ubaidillah, Khairil Anwar Notodiputro, Anang Kurnia & I. Wayan Mangku (2019) Multivariate Fay-Herriot models for small area estimation with application to household consumption per capita expenditure in Indonesia, Journal of Applied Statistics, 46:15, 2845-2861, DOI: 10.1080/02664763.2019.1615420

To link to this article: https://doi.org/10.1080/02664763.2019.1615420

Published online: 13 May 2019.

JOURNAL OF APPLIED STATISTICS
2019, VOL. 46, NO. 15, 2845–2861
https://doi.org/10.1080/02664763.2019.1615420

Taylor & Francis
Taylor & Francis Group

Check for updates

APPLICATION NOTE

Multivariate Fay-Herriot models for small area estimation with application to household consumption per capita expenditure in Indonesia

Azka Ubaidillah [a], Khairil Anwar Notodiputro [b], Anang Kurnia [b] and I. Wayan Mangku [c]

[a]STIS Polytechnic of Statistics, Jakarta, Indonesia; [b]Departement of Statistics, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Bogor, Indonesia; [c]Departement of Mathematics, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Bogor, Indonesia

ABSTRACT

Multivariate Fay-Herriot (MFH) models become popular methods to produce reliable parameter estimates of some related multiple characteristics of interest that are commonly produced from many surveys. This article studies the application of MFH models for estimating household consumption per capita expenditure (HCPE) on food and HCPE of non-food. Both of those associated direct estimates, which are obtained from the National Socioeconomic Surveys conducted regularly by Statistics Indonesia, have a strong correlation. The effects of correlation in MFH models are evaluated by employing a simulation study. The simulation showed that the strength of correlation between variables of interest, instead of the number of domains, plays a prominent role in MFH models. The application showed that MFH models have more efficient than univariate models in terms of standard errors of regression parameter estimates. The roots of mean squared errors (RMSEs) of the estimates obtained from the empirical best linear unbiased prediction (EBLUP) estimators of MFH models are smaller than RMSEs obtained from the direct estimators. Based on MFH model, the HCPE estimates of food by districts in Central Java, Indonesia, are higher than the HCPE estimates of non-food. The average of HCPE estimates of food and non-food in Central Java, Indonesia in 2015 are IDR 383,100.6 and IDR 280,653.6, respectively.

ARTICLE HISTORY
Received 19 March 2018
Accepted 1 May 2019

KEYWORDS
Correlation; EBLUP; household consumption per capita expenditure; multivariate Fay-Herriot models; small area estimation

1. Introduction

Surveys, for the past few decades, have not completely replaced census or complete enumeration as a more cost-effective means for obtaining information on wide-ranging topics of interest. Surveys can be used to derive reliable estimates for large areas (see [14,32] for theoretical references and [4–6,9–11] for more extensive practical references of inferential methods). However, direct estimates for small area, based on sample data from surveys, are likely inefficient due to its large standard errors. Small area estimation (SAE) can improve

CONTACT Khairil Anwar Notodiputro khairil@apps.ipb.ac.id

the effectiveness of sample size of surveys by borrowing the strength of neighboring areas and the relation between the set of auxiliary variables and the variables of interest [28].

Originally, SAE application using Empirical Best Linear Unbiased Prediction (EBLUP) estimator, which is named as Fay-Herriot (FH) model, was initiated by [20] to estimate log per-capita income (PCI) in small places of the US. In many years later, the FH models for SAE have been developed by many researchers. The measurement of the variability of area level means of prediction was studied in [15,16,18,21,22]. An extension of FH model to handle time series and cross-sectional data was studied by [24,29]. Then, the FH model based on the spatiotemporal model was developed by [23]. For multivariate cases, [19] proposed the multivariate Fay-Herriot (MFH) models for SAE.

By using an advantage of the correlation between variables of interest [2,3,7,12,16,17,26] demonstrated that multivariate models can lead more efficient estimators than univariate models. Therefore, MFH models become popular methods to produce reliable parameter estimates of some related multiple characteristics of interest that are commonly produced form many surveys. [17] used a MFH model to estimate the median income of four-person families for the US states by employing hierarchical Bayes estimation. [8] extended a MFH model with allowing different covariance structures of area random effects to estimate poverty proportions at province level in Spanish.

It is well known that household consumption per capita expenditure is one of the economic indicators that is often used for measuring the prosperity and well-being [1,33]. Therefore, producing reliable small area statistics of this data is becoming important for both public and private sectors. Based on the National Socioeconomic Surveys which conducted regularly by Statistics Indonesia, household consumption per capita expenditure (HCPE) data is formed from two components, i.e. HCPE of food and HCPE of non-food. The HCPE of food and non-food most likely have a strong correlation.

Until now, without being exhaustive, small area estimation of per capita expenditure or per capita income has been studied by [16,17,20,21,31] with many models. However, there is no research that applicate MFH models to estimate HCPE by using an advantage of the correlation between its two components, i.e. HCPE of food and non-food. This paper is purposed to estimate HCPE of food and non-food using MFH models.

This paper is organized as follows. Section 2 describes the data used in this research and the suitable method for analysis data, particularly in the case of small area estimation. Section 3 gives main formulas of MFH models including the EBLUP estimator and mean squared errors (MSE) of EBLUP estimation. These main formulas are motivated by [8,28] with many modifications of mathematics notation. Section 4 presents a simulation to analyze the behavior of parameter estimates of MFH models with the different correlation between components of sampling errors. Section 5 describes household consumption expenditure data, related surveys, sample size and gives some results of model application to real data. Section 6 provides the conclusions.

2. Data description

Household consumption per capita expenditure (HCPE) data in Indonesia have been collected by the National Socioeconomic Surveys (SUSENAS) which conducted regularly by Statistics Indonesia. Since 2015, SUSENAS is held periodically twice a year, in March and in September. The SUSENAS held in March was designed to produce estimates up to the district level, while the SUSENAS held in September was designed to produce estimates only for the provincial level.

This paper studies HCPE at the district levels in Central Java Province which is located in the heart of Java Island, the greatest population of the island in Indonesia. The total sample is 6695 households which distributed in 35 districts including 29 regencies and 6 municipalities for a period of the survey on September 2015. Because the survey (SUSENAS) held in September was designed to produce estimates only to the provincial level, it is necessary to apply the small area estimation method to produce reliable estimates up to the district level.

Based on SUSENAS, HCPE data is formed from two components, i.e. HCPE of food and HCPE of non-food. Both of them are most likely have a strong correlation. For this reason, the multivariate small area estimation model is considered appropriate to obtain small area mean of HCPE of food and HCPE of non-food estimates at the district level in Central Java Province on September 2015.

Table 1 reports that the mean of sample of HCPE of food is higher than the mean of HCPE of non-food, but the sample value of HCPE of food is less diverse than HCPE of non-food. The mean of sample of HCPE in regency area is lower of and less diverse than HCPE in municipality area. Figure 1 shows that, in regency area, the mean of sample of HCPE of

Table 1. Sample size, mean and standard deviation of HCPE in the period of the survey on september 2015 in Central Java, Indonesia.

HCPE	District	Size	Mean	Std.Deviation
Food	Regency	5788	439,739.6	326,405.2
	Municipality	907	608,604.0	443,173.9
	All	6695	462,616.4	349,323.8
Non-food	Regency	5788	392,205.5	894,904.8
	Municipality	907	705,619.2	992,182.5
	All	6695	434,665.0	914,924.5
Food + Non-food	Regency	5788	831,945.1	1072,504.9
	Municipality	907	1314,223.2	1272,554.1
	All	6695	897,281.4	1113,927.5

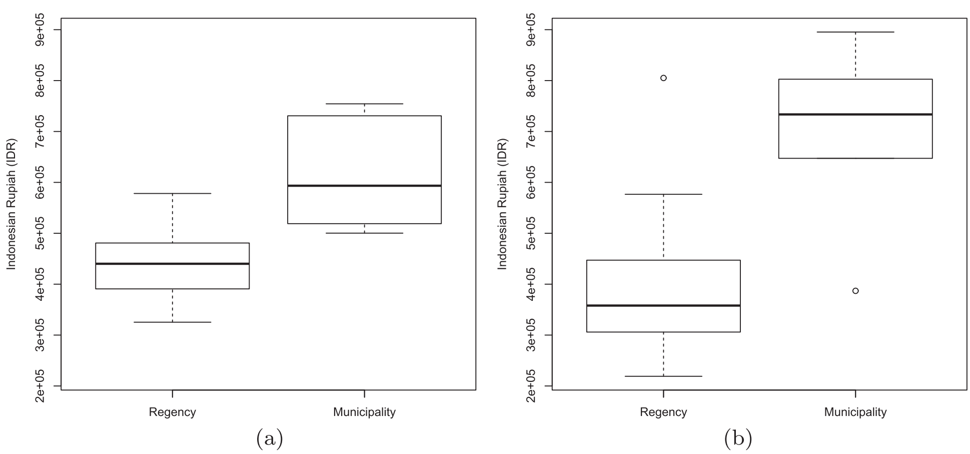


Figure 1. Boxplots of the sample mean of HCPE of food and non-food in the period of the survey on September 2015 in Central Java, Indonesia. (a) HCPE of food and (b) HCPE of non-food.

Table 2. Auxiliary variables.

Variable	Notation	Unit of measure
Number of pre-prosperous family	X_1	unit of family
Districts minimum wage	X_2	IDR
Gross regional domestic product at 2010 constant market price	X_3	billion (IDR)
Wide harvested area of wetland paddy	X_4	hectare
Ratio of total number of senior high school per 10,000 residents	X_5	school per 10,000 residents

food is higher but less diverse than HCPE of non-food. Meanwhile, in municipality area, the mean of sample of HCPE of food is lower but more diverse than HCPE of non-food.

For the correspond area-specific auxiliary variables, we use five variables: number of pre-prosperous family, districts minimum wage, gross regional domestic product at 2010 constant market price, wide harvested area of wetland paddy, and ratio of total number of senior high school per 10,000 residents. Those five auxiliary variables, shown in Table 2, are administrative data sourced from the Central Java province 2016 in figures publication.

3. Multivariate Fay-Herriot models

Suppose the population is partitioned into m area. Let $\boldsymbol{\mu}_d = (\mu_{1d}, \dots, \mu_{md})^{\mathrm{T}}$ be a vector of the dth characteristics of interest, with $d = 1, \dots, D$. Let $\mathbf{y}_d = (y_{1d}, \dots, y_{md})^{\mathrm{T}}$ be a vector of the dth direct estimators of $\boldsymbol{\mu}_d$. We assume that $\boldsymbol{\mu}_d$ is related to p_d area-specific auxiliary variables $\mathbf{X}_d = (\mathbf{X}_1, \dots, \mathbf{X}_D)^{\mathrm{T}}$ through a linear model

$$\boldsymbol{\mu}_d = \mathbf{X}_d{}^{\mathrm{T}}\boldsymbol{\beta}_d + \mathbf{u}_d, \quad \mathbf{u}_d \overset{iid}{\sim} N(\mathbf{0}, \mathbf{V}_{u_d}), \qquad d = 1, \dots, D, \tag{1}$$

where $\mathbf{u}_d = (u_{1d}, \dots, u_{md})^{\mathrm{T}}$ be a vector of area random effects with $D \times D$ covariance matrix $\mathbf{V}_{u_d} = \sigma^2_{u_d}\mathbf{I}_m$, which $\mathbf{I}_m$ is $m \times m$ identity matrix. The matrix $\mathbf{X}_d$ be the dth matrix of area-specific auxiliary variables of size $m \times p_d$ with $p = \sum_{d=1}^{D} p_d$ and $\boldsymbol{\beta}_d = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_D)^{\mathrm{T}}_{p_d \times 1}$ be a vector of regression coefficients corresponding with $\mathbf{X}_d$. The sampling model is

$$\mathbf{y}_d = \boldsymbol{\mu}_d + \mathbf{e}_d, \quad \mathbf{e}_d \overset{iid}{\sim} N(0, \mathbf{V}_{e_d}), \quad d = 1, \dots, D, \tag{2}$$

where $\mathbf{e}_d$ be a vector of sampling error with known $m \times m$ covariance matrix $\mathbf{V}_{e_d}$. Combining (1) and (2), we obtain

$$\mathbf{y}_d = \mathbf{X}_d\boldsymbol{\beta}_d + \mathbf{u}_d + \mathbf{e}_d, \quad d = 1, \dots, D, \quad i = 1, \dots, m, \tag{3}$$

where $\mathbf{u}_d$ and $\mathbf{e}_d$ are independent.

In matrix form, model (3) can be written as [see 28]

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad \mathbf{u} \sim N(0, \mathbf{G}), \quad \mathbf{e} \sim N(0, \mathbf{R}), \tag{4}$$

where $\mathbf{u} = \mathrm{col}_{1 \le d \le D}(\mathbf{u}_d)$ and $\mathbf{e} = \mathrm{col}_{1 \le d \le D}(\mathbf{e}_d)$ are independent. The matrix $\mathbf{Z}$ is a constant matrix that assumed to be known. The matrix $\mathbf{X} = \mathrm{diag}_{1 \le d \le D}(\mathbf{X}_d)$ is the $Dm \times 1$ matrix of area-specific auxiliary variables and $\mathbf{y} = \mathrm{col}_{1 \le d \le D}(\mathbf{y}_d)$ is the $Dm \times 1$ vector of variable of interest. The col operator means stacking matrix by column. $\mathbf{G} = \mathbf{V}_u \otimes \mathbf{I}_m$ is the covariance matrix of area effect where $\mathbf{V}_u = \mathrm{diag}_{1 \le d \le D}(\sigma^2_{u_d})$. The $\mathbf{R}$ is a covariance matrix of sampling errors of size $Dm \times Dm$ which is assumed to be known from particular surveys.

Unlike [8] that used the subscript d for the dth domain, we use subscript d for the dth characteristics of interest. The use of different subscript leads distinct forms of the covariance matrices $\mathbf{G}$ and $\mathbf{R}$.

3.1. Empirical best linear unbiased prediction (EBLUP)

Under model (4), it holds that $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathrm{var}(\mathbf{y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}} + \mathbf{R} = \boldsymbol{\Omega}$. The best linear unbiased prediction (BLUP) of $\boldsymbol{\mu} = \mathrm{col}_{1\le i\le m}(\boldsymbol{\mu}_i)$ is [see 8]

$$\tilde{\boldsymbol{\mu}} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}), \tag{5}$$

where $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{y}$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ with covariance matrix $\mathrm{cov}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}$.

The multivariate BLUP estimator (5) depends on unknown variance parameter component $\boldsymbol{\delta} = (\delta_{u1}^2, \dots, \delta_{uq}^2)^{\mathrm{T}}$. We use restricted maximum likelihood (REML) to estimate $\boldsymbol{\delta}$. The restricted log-likelihood function of model 4 is [see 8]

$$\ell_R(\boldsymbol{\delta}) = -\frac{Dm-p}{2}\log(2\pi) + \frac{1}{2}\log|\mathbf{X}^{\mathrm{T}}\mathbf{X}| - \frac{1}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}\log|\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{P}\mathbf{y}, \tag{6}$$

where $\mathbf{P} = \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}$. By taking partial derivatives of (6) with respect to $\boldsymbol{\delta}$ with kth element, $k = 1, \dots, q$, we obtain the score vector $\mathbf{s}(\boldsymbol{\delta}) = (s_1(\delta), \dots, s_q(\delta))^{\mathrm{T}}$ where

$$s_k(\delta) = \frac{\partial \ell_R(\boldsymbol{\delta})}{\partial \delta_k} = -\frac{1}{2}\mathbf{tr}\left(\mathbf{P}\boldsymbol{\Omega}_{(k)}\right) + \frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{P}\boldsymbol{\Omega}_{(k)}\mathbf{P}\mathbf{y}, \quad k = 1, \dots, q, \tag{7}$$

where $\boldsymbol{\Omega}_{(k)} = \partial\boldsymbol{\Omega}/\partial\delta_k$ is the partial derivative of $\boldsymbol{\Omega}$ with respect to kth element of $\boldsymbol{\delta}$ [see 8]. By taking second-order partial derivatives of (6) with respect to $\boldsymbol{\delta}$ with the klth element, changing sign and taking expectations, we obtain the Fisher information matrix [see 8]

$$\Im_{kl}(\boldsymbol{\delta}) = \tfrac{1}{2}\mathrm{tr}(\mathbf{P}\boldsymbol{\Omega}_{(l)}\mathbf{P}\boldsymbol{\Omega}_{(k)}), \quad k, l = 1, \dots, q. \tag{8}$$

The ath iterative of Fisher-scoring algorithm for REML estimation of $\boldsymbol{\delta}$ is

$$\hat{\boldsymbol{\delta}}^{(a+1)} = \hat{\boldsymbol{\delta}}^{(a)} + \Im_{kl}^{-1}(\hat{\boldsymbol{\delta}}^{(a)})\mathbf{s}(\hat{\boldsymbol{\delta}}^{(a)}), \quad k, l = 1, \dots, q. \tag{9}$$

By plugging $\hat{\boldsymbol{\delta}}$ in $\mathbf{G}$, we get $\mathbf{G}(\hat{\boldsymbol{\delta}})$ and $\boldsymbol{\Omega}(\hat{\boldsymbol{\delta}})$. Then substituting them in (5), we obtain the multivariate EBLUP estimator

$$\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\mathbf{G}(\hat{\boldsymbol{\delta}})Z^{\mathrm{T}}\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\delta}})(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \tag{10}$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\delta}})\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\delta}})\mathbf{y}$ with covariance matrix $\mathrm{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}(\hat{\boldsymbol{\delta}})\mathbf{X})^{-1}$. For $\hat{\beta}_{dr} = \beta_0$, the p-value for testing the hypothesis $H_0 : \beta_{dr} = 0, \quad d = 1, \dots, D, \quad r = 1, \dots, p_d$ is [see 8]

$$p\text{−value} = 2P_{H_0}(\hat{\beta}_{dr} > |\beta_0|) = 2P\left(N(0,1) > \frac{|\beta_0|}{\sqrt{\mathrm{cov}(\hat{\beta}_{dr})}}\right).$$

3.2. Estimation of mean squared errors (MSE)

The mean squared errors (MSE) of multivariate EBLUP estimator is obtained by taking the diagonal of covariance matrix of $\hat{\boldsymbol{\mu}}$ as follows:

$$\mathrm{MSE}(\hat{\boldsymbol{\mu}}) = \mathrm{col}_{1\le j\le Dm}\left[\mathrm{cov}(\hat{\boldsymbol{\mu}})\right]_j, \quad j = 1, \dots, Dm, \tag{11}$$

where $[\mathrm{cov}(\hat{\boldsymbol{\mu}})]_j$ is the jth diagonal element of $Dm \times Dm$ of covariance matrix $\mathrm{cov}(\hat{\boldsymbol{\mu}})$ that given by

$$\mathrm{cov}(\hat{\boldsymbol{\mu}}) \approx \boldsymbol{\Phi}_1(\boldsymbol{\delta}) + \boldsymbol{\Phi}_2(\boldsymbol{\delta}) + \boldsymbol{\Phi}_3(\boldsymbol{\delta}), \tag{12}$$

with

$$\begin{aligned} \boldsymbol{\Phi}_1(\boldsymbol{\delta}) &= \boldsymbol{\Gamma}\mathbf{R}, \\ \boldsymbol{\Phi}_2(\boldsymbol{\delta}) &= (\mathbf{I} - \boldsymbol{\Gamma})\mathbf{X}(\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{I} - \boldsymbol{\Gamma})^T, \\ \boldsymbol{\Phi}_3(\boldsymbol{\delta}) &\approx \sum_{k=1}^{q}\sum_{l=1}^{q} \mathrm{cov}(\hat{\delta}_k, \hat{\delta}_l)\boldsymbol{\Gamma}_{(k)}\boldsymbol{\Omega}\boldsymbol{\Gamma}_{(l)}^T, \qquad k, l = 1, \dots, q, \end{aligned} \tag{13}$$

where $\boldsymbol{\Gamma}_{(k)} = \partial\boldsymbol{\Gamma}/\partial\delta_k$ and $\mathrm{cov}(\hat{\delta}_k, \hat{\delta}_l)$ is the klth element of $\Im_{kl}^{-1}(\hat{\boldsymbol{\delta}})$, the inverse of REML Fisher information matrix (8). Furthermore, (11) can be written as

$$\mathrm{MSE}(\hat{\boldsymbol{\mu}}) \approx \mathbf{g}_1(\boldsymbol{\delta}) + \mathbf{g}_2(\boldsymbol{\delta}) + \mathbf{g}_3(\boldsymbol{\delta}), \tag{14}$$

where $\mathbf{g}_t(\boldsymbol{\delta}) = \mathrm{col}_{1\le j\le Dm}[\boldsymbol{\Phi}_t(\boldsymbol{\delta})]_j$, $j = 1, \dots, Dm$, $t = 1,2,3$, and $[\boldsymbol{\Phi}_t(\boldsymbol{\delta})]_j$ is the jth diagonal element of matrix $\boldsymbol{\Phi}_t(\boldsymbol{\delta})$ which is stated in (13).

Finally, similarly as [8,27,28], by substituting the REML estimator $\hat{\boldsymbol{\delta}}$ in (14), we estimate $\mathrm{MSE}(\hat{\boldsymbol{\mu}})$ with

$$\mathrm{mse}(\hat{\boldsymbol{\mu}}) \approx \mathbf{g}_1(\hat{\boldsymbol{\delta}}) + \mathbf{g}_2(\hat{\boldsymbol{\delta}}) + 2\mathbf{g}_3(\hat{\boldsymbol{\delta}}). \tag{15}$$

The asymptotic distribution of the EBLUP estimator $\hat{\boldsymbol{\mu}}$ can be formed as $\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \mathrm{cov}(\hat{\boldsymbol{\mu}}))$, then we can obtain $(1-\alpha)$-level confidence intervals for μ_j as

$$\hat{\mu}_j \pm z_{\alpha/2}\left[\mathrm{mse}(\hat{\boldsymbol{\mu}})\right]_j^{1/2}, \quad j = 1, \dots, Dm, \tag{16}$$

where $[\mathrm{mse}(\hat{\boldsymbol{\mu}})]_j$ is the jth element of vector $\mathrm{mse}(\hat{\boldsymbol{\mu}})$ and $z_{\alpha/2}$ is the α-quantile of the $N(0,1)$ distribution.

4. Simulation

The simulation in this paper is designed to analyze the behavior of the EBLUP estimates based on MFH models with the different correlations between components of sampling errors. Let us write model (4) in the form

$$\begin{aligned} Y_1 &= \beta_{01} + \beta_{11}X_1 + \beta_{21}X_2 + \beta_{31}X_3 + u_1 + e_1, \\ Y_2 &= \beta_{02} + \beta_{12}X_4 + \beta_{22}X_5 + u_2 + e_2, \end{aligned} \tag{17}$$

where $\binom{u_1}{u_2} \sim N\left[\binom{0}{0}, \begin{pmatrix}\sigma_{u11} & 0\\ 0 & \sigma_{u22}\end{pmatrix}\right]$ and $\binom{e_1}{e_2} \sim N\left[\binom{0}{0}, \begin{pmatrix}\sigma_{e11} & \sigma_{e12}\\ \sigma_{e12} & \sigma_{e22}\end{pmatrix}\right]$. The correlation between components of sampling errors is given by $\rho_e = \sigma_{e12}/\sqrt{\sigma_{e11}\sigma_{e11}}$.

In this simulation study, we employ three particularizations of model (17). Model 1 (or M_1) is assumed that the correlation is zero ($\rho_e = 0$), so the covariance matrix of sampling errors becomes the diagonal matrix with $\sigma_{e_{12}} = 0$. M_1 can also called the univariate Fay-Herriot (UFH) model. Model 2 (or M_2) is assumed that the correlation exists with the homogeneity of the covariance matrix of area random effects ($\sigma_{u_{11}} = \sigma_{u_{22}}$). Model 3 (or M_3) is the heteroscedastic version of M_2.

The simulation is run into three values of correlation i.e. 0.1, 0.5 and 0.9. We take $\beta_{01} = 5$, $\beta_{02} = 10$, $\beta_{11} = -0.05$, $\beta_{21} = 0.5$, $\beta_{31} = 0.15$, $\beta_{12} = 0.2$, $\beta_{22} = -0.1$, $\sigma_{e_{11}} = 0.02$, $\sigma_{e_{22}} = 0.03$. For M_2 we set $\sigma_{u_{11}} = \sigma_{u_{22}} = 0.02$, meanwhile for M_3, we set $\sigma_{u_{22}} = 0.02$ and $\sigma_{u_{22}} = 0.04$. For explanatory variables, we set $X_1 \sim N(10, 1), X_2 \sim \text{Burr}(0.079, 1995.5, 14.07), X_3 \sim N(10, 0.5), X_4 \sim U(10, 12)$, and $X_5 \sim \text{Burr}(0.2, 14.74, 0.64)$, where N, U, and Burr denote the Normal distribution, Uniform distribution, and Burr distribution, respectively. The Burr distribution is a continuous probability distribution for a non-negative random variable that is most commonly used to model household income (see [13,30] for references).

The steps of the simulation are

(1) For each correlation $\rho_e = 0.1, 0.5$, and 0.9, repeat $B = 1000$ times
   (a) Generate $\{e_{id}^{(b)}, u_{id}^{(b)}, Y_{id}^{(b)}, X_{id}\}$, $i = 1, \ldots, m$, $d = 1, 2$ from model M_2 and model M_3. We take $m = 50, 100, 200$, and 400.
   (b) Calculate the target of parameter value, $\mu_{id}^{(b)} = \mathbf{x}_{id}^{\mathrm{T}} \boldsymbol{\beta} + u_{id}^{(b)}$.
   (c) Calculate the EBLUP estimates, $\hat{\mu}_{id}^{(b)}$.
   (d) Calculate the bias, $\hat{\mu}_{id}^{(b)} - \mu_{id}^{(b)}$.
   (e) Calculate the mean squared error, $1/m \sum_{i=1}^{m} (\hat{\mu}_{id}^{(b)} - \mu_{id}^{(b)})^2$

(2) Calculate the average of absolute relative error ($\overline{\text{ARE}}$) and the average of relative efficiency ($\overline{\text{EFF}}$) as

$$\overline{\text{ARE}}_d = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{B} \sum_{b=1}^{B} \left| \frac{\hat{\mu}_{id}^{(b)}}{\mu_{id}^{(b)}} - 1 \right|, \quad \overline{\text{EFF}}_d = \left[\frac{\text{mse}(Y_{id}^{(b)})}{\text{mse}(\hat{\mu}_{id}^{(b)})} \right]^{1/2}, \tag{18}$$

where

$$\text{mse}(\hat{\mu}_{id}^{(b)}) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{B} \sum_{b=1}^{B} \left(\hat{\mu}_{id}^{(b)} - \mu_{id}^{(b)} \right)^2 \quad \text{and}$$

$$\text{mse}(Y_{id}^{(b)}) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{B} \sum_{b=1}^{B} \left(Y_{id}^{(b)} - \mu_{id}^{(b)} \right)^2 .$$

Table 3 presents the simulation results of $\overline{\text{ARE}}$ and $\overline{\text{EFF}}$ values for the components $d = 1,2$, for the classes of correlation ρ_e, and for the numbers of domains $m = 50,100,200$ and 400. The four first columns of Table 3 state the target of the variables (Resp), the model generating the data (Genr), the correlation (ρ_e) and the model for estimating $\hat{\mu}$ (Model).

Table 3 gives an information that MFH models have more efficient estimates than UFH models. The MFH models have lower $\overline{\text{ARE}}$ and higher $\overline{\text{EFF}}$ than the UFH models. Table 3 shows that all of $\overline{\text{ARE}}$ decrease as the number of domains m increase, while all of $\overline{\text{EFF}}$

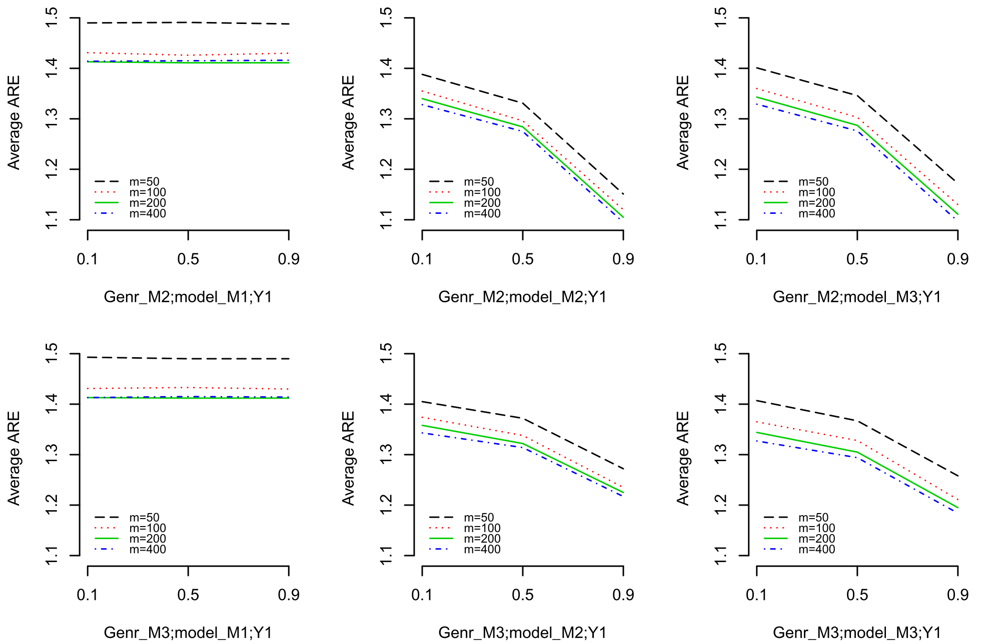
Table 3. Average absolute error ($\overline{ARE}$) and average relative efficiency ($\overline{EFF}$) of model estimates.

Resp	Genr	ρ_e	Model	$\overline{ARE}$(%) 50	100	200	400	$\overline{EFF}$(%) 50	100	200	400
Y_1	M2	0.1	M1	1.490	1.431	1.413	1.414	125.514	130.796	132.796	132.251
			M2	1.388	1.355	1.340	1.328	134.785	137.982	140.014	140.815
			M3	1.401	1.360	1.343	1.329	133.710	137.480	139.726	140.684
		0.5	M1	1.491	1.426	1.411	1.415	126.107	131.468	132.968	132.366
			M2	1.331	1.296	1.284	1.275	141.507	144.769	146.105	146.798
			M3	1.346	1.303	1.287	1.276	140.114	144.043	145.756	146.605
		0.9	M1	1.488	1.430	1.411	1.416	126.695	131.388	133.004	132.151
			M2	1.151	1.120	1.105	1.095	164.226	167.875	170.011	170.917
			M3	1.172	1.130	1.111	1.097	161.291	166.404	169.181	170.511
	M3	0.1	M1	1.493	1.431	1.413	1.413	126.170	131.110	132.727	132.221
			M2	1.405	1.374	1.358	1.343	133.823	136.442	138.064	138.998
			M3	1.407	1.365	1.344	1.327	134.024	137.481	139.653	140.702
		0.5	M1	1.490	1.433	1.412	1.415	126.884	131.096	132.354	132.363
			M2	1.372	1.338	1.322	1.314	137.819	140.190	141.361	142.497
			M3	1.367	1.328	1.305	1.294	138.515	141.446	143.262	144.798
		0.9	M1	1.490	1.430	1.412	1.414	127.361	130.998	132.832	132.508
			M2	1.272	1.235	1.225	1.217	149.888	151.378	153.191	154.000
			M3	1.258	1.211	1.195	1.184	151.658	154.563	157.181	158.353
Y_2	M2	0.1	M1	0.864	0.949	0.870	0.842	133.971	120.226	132.294	137.002
			M2	0.772	0.751	0.741	0.735	150.774	154.321	155.868	157.148
			M3	0.784	0.757	0.745	0.736	148.712	153.066	155.152	156.806
		0.5	M1	0.869	0.949	0.869	0.839	133.791	120.452	132.686	137.201
			M2	0.750	0.730	0.718	0.714	155.437	159.099	161.157	161.770
			M3	0.762	0.737	0.721	0.715	153.142	157.567	160.346	161.445
		0.9	M1	0.868	0.953	0.870	0.840	133.821	120.337	132.593	137.200
			M2	0.682	0.666	0.655	0.651	170.635	174.599	176.663	177.385
			M3	0.698	0.673	0.659	0.652	166.912	172.760	175.651	176.933
	M3	0.1	M1	0.951	0.995	0.951	0.933	121.862	115.478	121.583	123.651
			M2	0.910	0.899	0.893	0.888	128.029	128.944	129.809	130.166
			M3	0.903	0.888	0.881	0.875	128.824	130.255	131.347	131.982
		0.5	M1	0.953	0.994	0.947	0.932	121.647	115.593	121.620	123.621
			M2	0.879	0.865	0.857	0.857	132.761	133.700	134.196	134.673
			M3	0.872	0.854	0.844	0.843	133.690	135.303	136.291	136.942
		0.9	M1	0.953	0.997	0.951	0.934	121.853	115.220	121.705	123.607
			M2	0.792	0.780	0.770	0.765	147.404	148.429	150.458	150.895
			M3	0.784	0.765	0.751	0.745	149.075	151.480	154.298	155.017

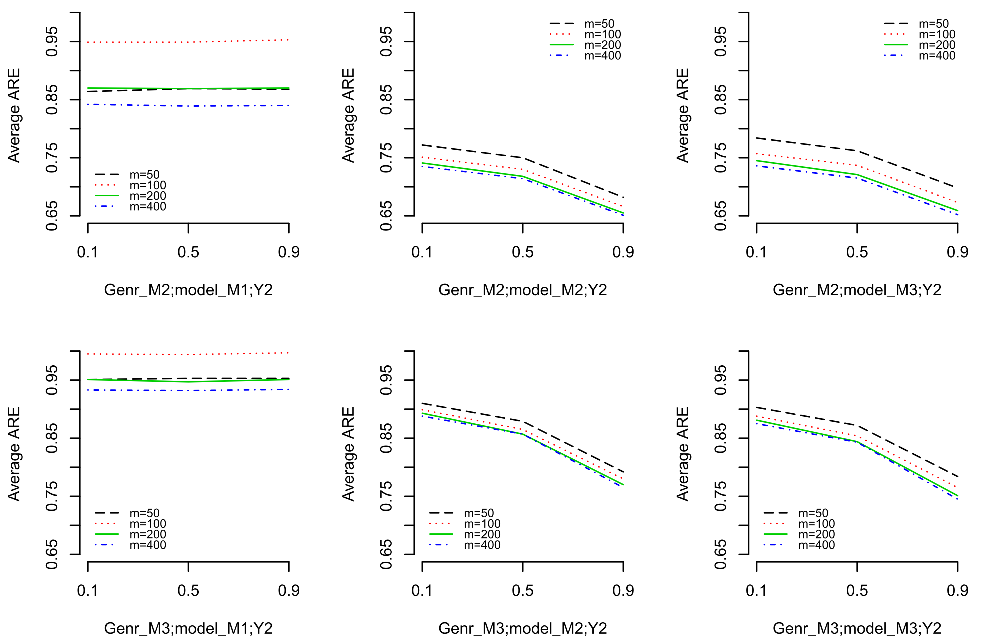
increase as the number of domains m increase. Table 3 also shows that the performances of MFH models increase as the number of domains m and the correlations ρ_e increase.

Figure 2 shows the average of absolute relative error ($\overline{ARE}$) values from the simulation results. The correlation values ($\rho_e = 0.1, 0.5, 0.9$) are labeled as the X axis, whereas $\overline{ARE}$ values are labeled as the Y axis. The number of domains ($m = 50, 100, 200, 400$) is presented as the series with different colors and line types. Figure 2(a) shows the $\overline{ARE}$ values of Y_1 estimates using model M1, M2 and M3 from the data generated by model M2 (placed in the upper side) and the data generated by model M3 (placed in the lower side). The $\overline{ARE}$ values of Y_2 estimates are presented in Figure 2(b). The label 'Genr_M2;model_M1;Y1' in Figure 2(a) in topleft side denotes the $\overline{ARE}$ values of Y_1 estimates using M1 model from the data generated by M2, etc.

Figure 2 demonstrates that the correlations, instead of the sample sizes, play prominent roles in the performance of MFH models. Figure 2 shows that the declining of $\overline{ARE}$ values, caused by the increasing of the correlations, are more rapidly than the declining of $\overline{ARE}$ values caused by the increasing of the number of domains m. Figure 2 also shows that the



(a)

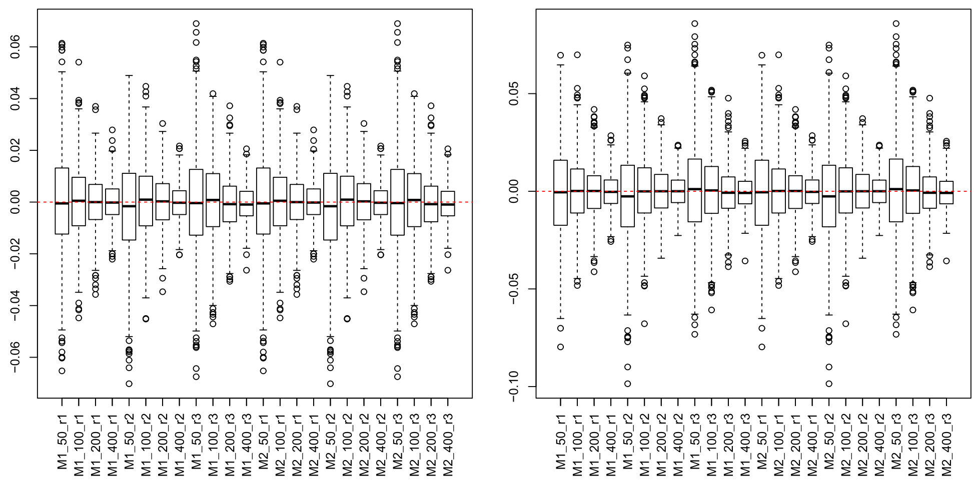


(b)

Figure 2. The $\overline{\text{ARE}}$s of Y_1 and Y_2 estimates with $m = 50,100,200,400$ and $\rho_e = 0.1, 0.5, 0.9$. (a) The $\overline{\text{ARE}}$s of Y_1 estimates.

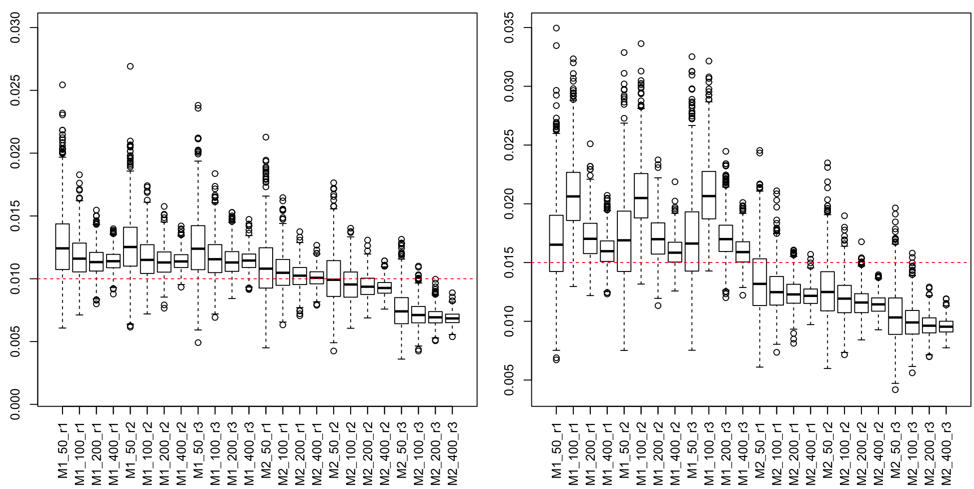
model from which data are generated performs the best model with the smallest values of $\overline{\text{ARE}}$.

Figures 3 and 4 show the biases and the mean squared errors (MSEs) of EBLUP estimates based on UFH models and MFH models with different amount of domain (m) and correlation (ρ_e). The r1, r2, and r3 denote the correlation $\rho_e = 0.1, 0.5$, and 0.9, respectively. The M1_50_r1 denotes the M_1 model with $m = 50$ and $\rho_e = 0.1$. The M2_100_r2



(a) (b)

Figure 3. Boxplots of biases of Y_1 and Y_2 based on UFH model estimates (M1) and MFH models (M2), with $m = 50, 100, 200, 400$ and $\rho_e = 0.1, 0.5, 0.9$. (a) Boxplots of biases of Y_1 estimates and (b) Boxplots of biases of Y_2 estimates.



(a) (b)

Figure 4. Boxplots of mean squared errors of Y_1 and Y_2 based on UFH model estimates (M1) and MFH models (M2), with $m = 50, 100, 200, 400$ and $\rho_e = 0.1, 0.5, 0.9$. (a) Boxplots of mean squared errors of Y_1 estimates and (b) Boxplots of mean squared errors of Y_2 estimates.

denotes the M_2 model with $m = 100$ and $\rho_e = 0.5$, etc. For the sake of brevity, we do not present M3 estimates, where the similar results are also obtained.

Figure 3 shows that the behavior of biases is almost similar between UFH and MFH model estimates. The median of all model estimates is around at zero value. The interquartile range of biases of all model estimates decreases as m increases. The similarity behavior of those biases are obtained because both of UFH and MFH models are in the same class of the linear unbiased prediction.

Figure 4 shows that the MSEs of EBLUP estimates obtained from MFH models are smaller than the MSEs of EBLUP estimates obtained from UFH models. The decreasing of MSEs of EBLUP estimates obtained from MFH model if the domain m increases is more consistent than the decreasing of MSEs of EBLUP estimates obtained from UFH models. In addition, the stronger of correlation (ρ_e) causes the smaller value of MSEs of MFH model estimates.

5. Small area estimation of household consumption per capita expenditure (HCPE)

he target variables are log HCPE of food (Y_1) and log HCPE of non_food (Y_2). The direct estimator of a total $Y_{id} = \sum_{n=1}^{N_i} y_{idn}$ is $\hat{Y}_{id}^{dir} = \sum_{n\in\varpi_i} w_{in} y_{idn}$, where ϖ_i is the sample and w_{in}s are the official sampling weights of the nth household within the ith area, $d = 1,2$, $n = 1,\ldots,N_i$, $i = 1,\ldots,m$. The direct estimator of the ith area population is $\hat{N}_i^{dir} = \sum_{n\in\varpi_i} w_{in}$. The direct estimator of the area mean is $\bar{y}_{id} = \hat{Y}_{id}^{dir}/\hat{N}_i^{dir}$ and the direct estimator of covariance matrix is approximated by

$$\mathrm{cov}(\bar{y}_{i1}, \bar{y}_{i2}) = \frac{1}{(\hat{N}_i^{dir})^2} \sum_{n\in\varpi_i} w_{in}(w_{in} - 1)(y_{i1n} - \bar{y}_{i1})(y_{i2n} - \bar{y}_{i2}).$$

Those direct estimator formulas are obtained from [8].

We use two direct estimators, Y_1 and Y_2, as two target variables of interest. For the correspond area-specific auxiliary variables, we use five variables as described in Section 2. We use logarithm transformation for variables X_1, X_2, X_3, and X_4 to obtain an appropriate model. The significantly regression parameter estimates (with alpha 5%) based on UFH and MFH models are shown in Table 4.

Table 4 presents the regression parameter estimates $\hat{\beta}_{dr}$, $d = 1,2$, $r = 1,\ldots,p_d$, the standard errors of parameter estimates s.e($\hat{\beta}_{dr}$), and the p-values for testing $H_0 : \beta_{dr} = 0$ from

Table 4. Regression parameter estimates of HCPE modeling using UFH and MFH models.

		UFH			MFH		
Resp	Variables	Estimate	Std.Error	*p*-value	Estimate	Std.Error	*p*-value
Y_1	Constant	5.44241	3.33046	0.10223	5.44606	3.12171	0.08106
	log X_1	−0.05864	0.02736	0.03209	−0.05539	0.02568	0.03103
	log X_2	0.55699	0.23255	0.01661	0.55389	0.21795	0.01104
	X_5	0.14896	0.06089	0.01443	0.15389	0.05739	0.00733
Y_2	Constant	11.53249	0.56738	0.00000	11.52768	0.53204	0.00000
	log X_3	0.18940	0.05028	0.00017	0.18855	0.04710	0.00006
	log X_4	−0.10870	0.02671	0.00005	−0.10786	0.02505	0.00002
	X_5	0.24575	0.09617	0.01061	0.24970	0.09040	0.00574

both the UFH and MFH models. The two first columns of Table 4 states the response of variables of interest (Resp) and the area-specific auxiliary variables (Variables). Observing the standard errors of parameter estimates, we concluded that MFH model is more efficient than UFH model. This is because the direct estimators, Y_1 and Y_2, have a strong correlation with the coefficient correlation 0.86 and p-value $=$ .000 at the level significance of 0.05.

Table 5 shows that the random effect variance estimates $\hat{\boldsymbol{\delta}} = (\hat{\sigma}^2_{u_1}, \hat{\sigma}^2_{u_2})^{\mathrm{T}}$ and their standard errors obtained from MFH model are smaller than the estimates of variance components obtained from UFH model. Therefore, the MFH model is regarded to improve the effectiveness of the sample size obtained from the survey. The efficiency of model estimates obtained from the EBLUP estimators of MFH model compared to direct estimators is presented in Figure 5.

Figure 5 plots the RMSEs of the direct estimates and the EBLUP estimates based on MFH models for Y_1 and Y_2. Figure 5, which small areas (districts) have been sorted by increasing sample size, shows that the RMSEs of the EBLUP estimators based on MFH models are smaller than the RMSEs of the direct estimators for all areas. Figure 5 also shows a decreasing pattern of RMSEs as the sample sizes increase for both direct and EBLUP estimates based on MFH models. This is coherent with [8] where increasing sample size has an impact on the decrease in RMSE pattern.

Table 6 shows the direct estimates and EBLUP estimates of MFH models for HCPE of food, HCPE of non-food, and HCPE of the total (HCPE of food + HCPE of non-food).

Table 5. The estimates of variance components.

Model	Components	Estimate	Std. error
UFH	$\sigma^2_{u_1}$	0.01387419	0.003994492
	$\sigma^2_{u_2}$	0.02870398	0.008095995
MFH	$\sigma^2_{u_1}$	0.01203362	0.003501326
	$\sigma^2_{u_2}$	0.02495372	0.007091886

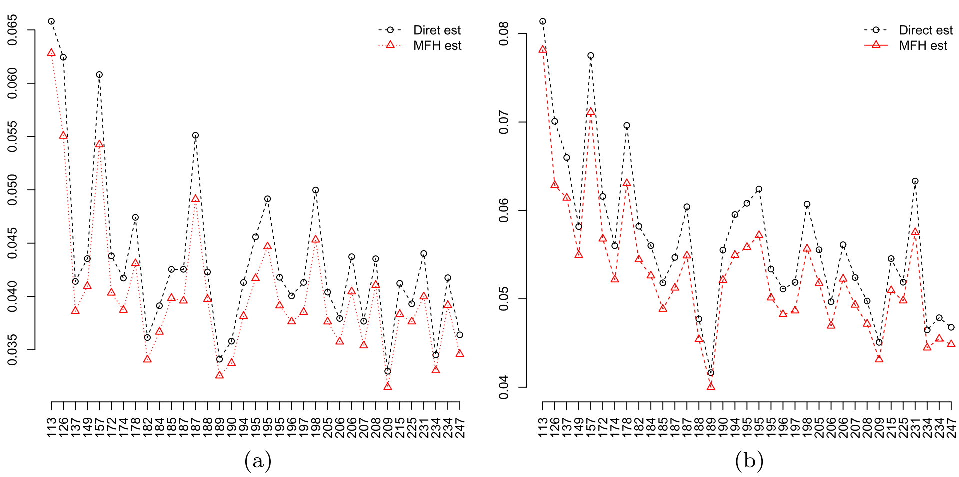


Figure 5. RMSE of the direct estimates and the EBLUP estimates obtained from MFH model which districts sorted by increasing sample size. (a) RMSE of the direct and MFH estimates of Y_1 and (b) RMSE of the direct and MFH estimates of Y_2

Table 6. Direct estimates and MFH model estimates of HCPE of food, HCPE of non-food, and HPCE of the total at a period of the survey on September 2015 in Central Java, Indonesia.

		HCPE of food (IDR)		HCPE of non-food (IDR)		HCPE of total (IDR)	
District	Sample size	Direct	MFH	Direct	MFH	Direct	MFH
Regency							
Cilacap	247	387,802	387,057	272,848	273,319	660,650	660,376
Banyumas	231	358,087	355,407	270,798	267,792	628,885	623,200
Purbalingga	187	329,539	332,877	218,181	220,354	547,720	553,231
Banjarnegara	188	277,629	287,765	175,471	182,285	453,100	470,050
Kebumen	209	355,347	356,008	235,840	236,217	591,187	592,225
Purworejo	172	359,640	360,384	210,896	212,058	570,536	572,442
Wonosobo	198	364,328	359,384	242,406	238,376	606,734	597,760
Magelang	195	287,736	298,285	210,845	219,505	498,582	517,789
Boyolali	194	358,266	356,609	271,091	268,369	629,357	624,977
Klaten	207	388,392	384,040	271,704	267,398	660,096	651,438
Sukoharjo	178	458,461	430,443	400,641	360,977	859,101	791,420
Wonogiri	195	385,012	373,634	294,263	282,013	679,275	655,647
Karanganyar	182	322,333	324,912	270,422	270,248	592,755	595,160
Sragen	197	393,757	384,803	277,870	270,468	671,628	655,271
Grobogan	234	303,477	308,167	172,362	176,662	475,838	484,829
Blora	185	287,650	296,652	186,212	192,738	473,862	489,390
Rembang	174	389,345	385,462	225,217	223,700	614,562	609,162
Pati	206	392,087	386,654	262,228	258,433	654,316	645,087
Kudus	184	318,663	333,254	245,286	261,132	563,949	594,386
Jepara	205	360,105	362,780	233,844	237,100	593,949	599,879
Demak	208	444,449	441,053	277,745	274,716	722,194	715,770
Semarang	196	425,440	416,310	336,529	326,455	761,969	742,765
Temanggung	187	318,227	323,808	243,482	246,166	561,709	569,974
Kendal	195	391,605	393,811	234,280	237,130	625,885	630,940
Batang	190	328,526	335,451	181,981	188,164	510,507	523,615
Pekalongan	189	346,979	349,181	226,974	228,113	573,953	577,294
Pemalang	206	366,195	360,740	221,365	217,927	587,560	578,667
Tegal	215	414,306	404,819	253,729	247,704	668,035	652,524
Brebes	234	359,918	358,198	200,747	201,901	560,664	560,099
Municipality							
Magelang	113	537,073	549,926	550,592	561,481	1,087,666	1,111,407
Surakarta	157	473,255	479,541	521,169	537,615	994,424	1,017,156
Salatiga	126	631,896	581,420	473,800	435,794	1,105,696	1,017,214
Semarang	225	509,426	500,798	525,697	513,872	1,035,123	1,014,670
Pekalongan	149	410,762	417,637	249,103	255,405	659,865	673,042
Tegal	137	433,575	431,248	437,819	431,290	871,394	862,538

Those estimation values are in Indonesian Rupiah (IDR) and given by

$$\text{HCPE of food} = \exp(Y_1),$$
$$\text{HCPE of non} - \text{food} = \exp(Y_2),$$
$$\text{HCPE of total} = \text{HCPE of food } + \text{HCPE of non} - \text{food}.$$

By observing in Table 6, we showed that the EBLUP estimates obtained from MFH models track the direct estimates but seem slightly to be more stable. Table 3 also shows that the estimates of HCPE of food are still more dominant than the estimates of HCPE of non-food. Most people in Central Java use their income to suffice their food needs. This shows the pattern of developing areas that have not been a good level of welfare [see 25].

Table 7 reports the average and percentage of the estimates of HCPE of food and non-food by regency and municipality obtained from the EBLUP of MFH model. The estimates

Table 7. Average and percentage of the estimates of HCPE of food, HCPE of non-food and HCPE of the total obtained from the EBLUP of MFH model in Central Java, Indonesia.

	HCPE		
District	Food	Non-food	Total
Regency	360,274.1	244,393.7	604,667.8
	59.58%	*40.42%*	*100.00%*
Municipality	493,428.3	455,909.5	949,337.8
	51.98%	*48.02%*	*100.00%*
All	383,100.6	280,653.6	663,754.1
	57.72%	*42.28%*	*100.00%*

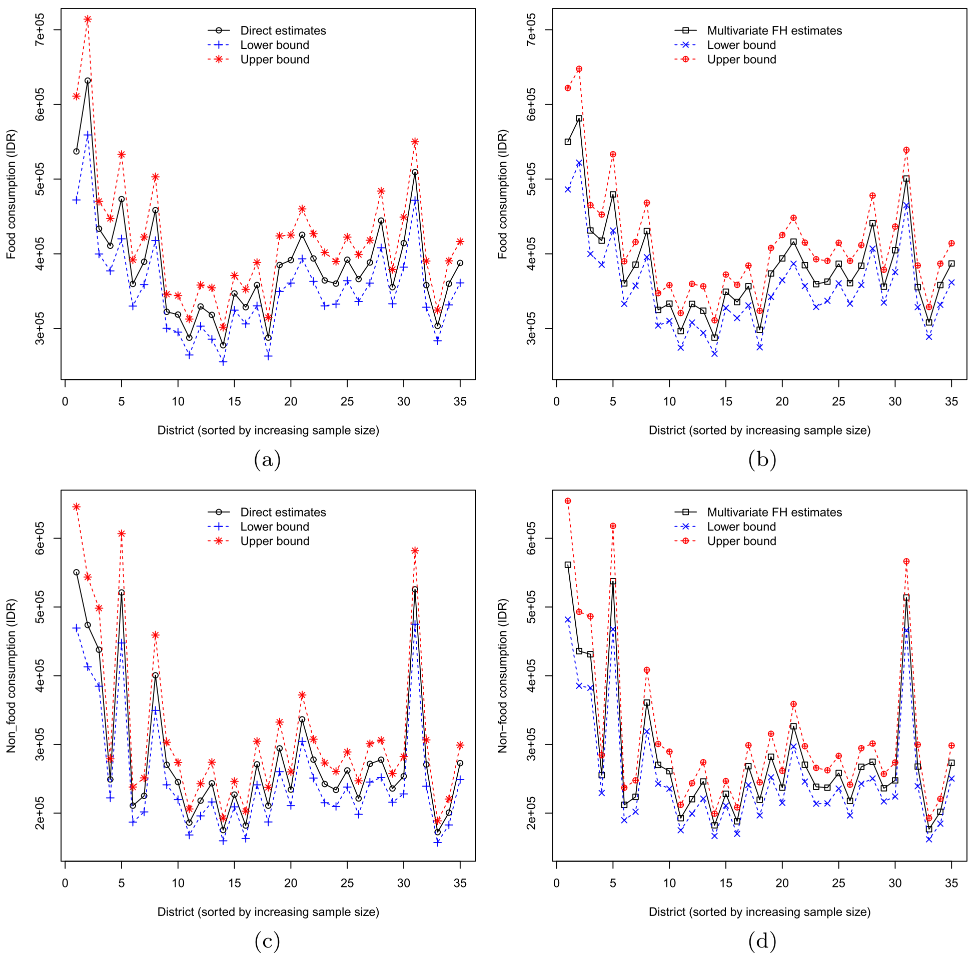


Figure 6. Confidence intervals of the direct estimates and the EBLUP of MFH model estimates of HCPE on food and non-food during the period of the survey on September 2015 in Central Java, Indonesia. (a) Direct estimates of HCPE of food. (b) The EBLUP of MFH estimates of HCPE of food. (c) Direct estimates of HCPE of non-food and (d) The EBLUP of MFH estimates of HCPE of non-food.

of HCPE of food are more dominant than the estimates of HCPE of non-food. For all district, the average of the estimates of HCPE of food is IDR 383,100.6 or 57.72% of HCPE of the total estimates, while the average of the estimates of HCPE of non-food is only 42.28% of HCPE of the total estimates during the period of the survey on September 2015 in Central Java.

For regency area, the average of the estimates of HCPE of the total is IDR 604,667.8 of which HCPE of food spends about 59.58% and HCPE of non-food spends about 40.42%. But slightly different, in municipality area, 51.96% of the average of the total of HCPE estimates is spent on food consumption, and 42.28% of the average of the total of HCPE estimates is spent on non-food consumption.

The confidence intervals of the direct estimates and the EBLUP estimates of MFH models for HCPE of food and non-food are plotted in Figure 6 which small areas (districts) have been sorted by increasing sample size. Figure 6 shows that the confidence intervals of the EBLUP estimates of MFH models are slightly shorter than the direct estimates for all districts. Thus, in this application, the EBLUP estimators of MFH models are more reliable than the direct estimators.

6. Conclusions

We have shown that the MFH models produced more efficient of parameter estimates than the UFH models. By employing a simulation, the MFH models showed better performance compared to UFH models. The application to real data also gives the same information that the MFH model produced more efficient than the direct estimates. These results are coherent with [8,16] who studied MFH models for small area estimation.

The HCPE estimates based on MFH model in Central Java during the period of the survey on September 2015 are still dominated by HCPE of food. The highest value of HCPE estimates is Magelang City, while the least value of HCPE estimates is Banjarnegara Regency. The average of HCPE estimates of food and non-food in Central Java during the period of the survey in September 2015 is IDR 383,100.6 and IDR 280,653.6, respectively.

The interesting thing from this paper is the investigation of the effect of correlation between the variables of interest which often can be approximated from the correlation of components of sampling errors. The simulation showed that the increase of correlation between components of sampling errors plays a prominent role in the MFH model's performance.

For the next study, we will apply the simultaneous equation FH models for SAE. Because, in the real application, variables that are produced from many surveys may not only have the correlation, but also the interdependent relationship. Therefore, by taking an advantage of interdependent relationship among variables of interest, the substantial improvements in parameter estimates of SAE are possibly achieved by using the simultaneous equation FH models.

Acknowledgments

The authors would like to express a sincere gratitude to Statistics Indonesia for providing funds for this research as a part of the PhD program and for its human resource development projects in cooperation with Bogor Agricultural University, Indonesia.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Azka Ubaidillah http://orcid.org/0000-0002-3597-0459

References

[1] T. Akita and A. Pirmansyah, *Urban inequality in Indonesia*, IUJ Research Institute, Minami-uonuma, 2011.

[2] G.R. Arboretti and S. Bonnini, *Moment-based multivariate permutation tests for ordinal categorical data*, J. Nonpar. Stat. 18 (2008), pp. 383–393.

[3] G.R. Arboretti and S. Bonnini, *Some new results on univariate and multivariate permutation tests for ordinal categorical variables under restricted alternatives*, Stat. Method. Appl. J. Ital. Stat. Soc. 18 (2009), pp. 221–236.

[4] G.R. Arboretti, S. Bonnini, and L. Salmaso, *Employment status and education/employment relationship of PhD graduates from the University of Ferrara*, J. Appl. Stat. 36 (2009), pp. 1329–1344.

[5] G.R. Arboretti, S. Bonnini, and F. Pesarin, *A permutation approach for testing heterogeneity in two-sample problems*, Stat. Comput. 19 (2009), pp. 209–216.

[6] G.R. Arboretti, S. Bonnini, L. Corain, and D. Vidotto, *Environmental odor perception: testing regional differences on heterogeneity with application to odor perceptions in the area of Este (Italy)*, Environmetrics, 2015. doi:10.1002/env.2341.

[7] G.R. Arboretti, S. Bonnini, L. Corain, and L. Salmaso, *Dependency and truncated forms of combinations in multivariate combination-based permutation tests and ordered categorical variables*, J. Stat. Comp. Sim. (2016), pp. 1–12. doi:10.1080/00949655.2016.1177826.

[8] R. Benavent and D. Morales, *Multivariate Fay-Herriot models for small area estimation*, Comput. Stat. Data Anal. 94 (2016), pp. 372–390.

[9] S. Bonnini, D. Piccolo, L. Salmaso, and F. Solmi, *Permutation inference for a class of mixture models*, Comm. Stat. Theory Methods 41 (2012), pp. 2879–2895.

[10] S. Bonnini, *Testing for heterogeneity with categorical data:permutation solution vs bootstrap method*, Comm. Stat. Theory Methods 43 (2014), pp. 906–917.

[11] S. Bonnini, N. Prodi, L. Salmaso, and C. Visentin, *Permutation approaches for stochastic ordering*, Comm. Stat. Theory Methods 43 (2014), pp. 2227–2235.

[12] S. Bonnini, *Multivariate approach for comparative evaluations of customer satisfaction with application to transport services*, Comm. Stat.: Simul. Comput. (2015). doi:10.1080/03610918.2014.941685.

[13] I.W. Burr, *Cumulative frequency functions*, Annals Math. Stat. 13 (1942), pp. 215 232. doi:10.1214/aoms/1177731607.

[14] W.G. Cochran, *Sampling techniques*, 3rd ed., John Wiley & Son, New York, 1977.

[15] K. Das, J. Jiang, and J.N.K. Rao, *Mean squared error of empirical predictor*, Ann. Stat. 32 (2004), pp. 818–840.

[16] G.S. Datta, R.E. Fay, and M. Ghosh, *Hierarchical and empirical bayes multivariate analysis in small area estimation*, Proceedings of Bureau of The Census 1991 Annual Research Conference. U.S. Bureau of The Census, Washington, DC, 1991, pp. 63–67.

[17] G.S. Datta, M. Ghosh, N. Nangia, and K. Natarajan, *Estimation of median income of four-person families: a bayesian approach*, In W.A. Berry, K.M. Chaloner and J.K. Geweke, eds., Bayesian Analysis in Statistics and Econometrics, John Wiley & Son, Inc, New York, 1996, pp. 129–140.

[18] G.S. Datta and P. Lahiri, *A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems*, Stat. Sinica. 10 (2000), pp. 613–627.

[19] R.E. Fay, *Application of multivariate regression of small domain estimation*, In: R. Platek, J.N.K. Rao, C.E. Särndal, M.P. Singh, eds., Small Area Statistics, John Wiley & Son. Inc, New York, 1987, 91–102.
[20] R.E. Fay and R.A. Herriot, *Estimates of income for small places: an application of James-Stein procedures to census data*, J. Amer. Stat. Ass. 74 (1979), pp. 269–277.
[21] A. Kurnia and K.A. Notodiputro, *EB-EBLUP MSE estimator on small area estimation with application to BPS'data*, Proceeding at The First International Conference on Mathematics and Statistics (ICoMS-1), Bandung, 2006, June pp. 19–21.
[22] P. Lahiri and J.N.K. Rao, *Robust estimation of mean squared error of small area estimators*, J. Amer. Stat. Ass. 82 (1995), pp. 758–766.
[23] Y. Marhuenda, I. Molina, and D. Morales, *Small area estimation with spatio-temporal Fay-Herriot models*, Comput. Stat. Data Anal. 58 (2013), pp. 308–325.
[24] S. Muchlisoh, A. Kurnia, K.A. Notodiputro, and I.W. Mangku, *Estimation of unemployment rates using small area estimation model by combining time series and cross-sectional data*, AIP. Conf. Proc. 1707 (2016), pp. 080013.
[25] L. Pasinetti, *Structural Change and Economic Growth: A Theoretical Essay on the Dynamics of the Wealth of Nations*, Cambridge University Press, Cambridge, 1981.
[26] F. Pesarin and L. Salmaso, *Permutation Tests for Complex Data*, Applications and Software-Pesarin, Salmaso, Wiley, Chichester, 2010.
[27] N.G.N. Prasad and J.N.K. Rao, *The estimation of the mean squared error of small-area estimators*, J. Amer. Stat. Ass. 85 (1990), pp. 163–171.
[28] J.N.K. Rao and I. Molina, *Small Area Estimation*, 2nd Ed., John Wiley and Sons Inc., New York, 2015.
[29] J.N.K. Rao and M. Yu, *Small area estimation by combining time series and cross-sectional data*, Can. J. Stat. 22 (1994), pp. 511–528.
[30] R.N. Rodriguez, *A guide to Burr Type XII distributions*, Biometrika 64 (1977), pp. 129134. doi:10.1093/biomet/64.1.129.
[31] A. Salma, K. Sadik, and K.A. Notodiputro, *Small area estimation of per capita expenditures using robust empirical best linear unbiased prediction (REBLUP)*, AIP Conf. Proc. 1827 (2017). doi:10.1063/1.4979443.
[32] M.E. Thompson, *Theory of Sample Surveys*, Chapman and Hall, London, 1997.
[33] World Bank, *World Development Report 2000/2001: Attacking Poverty*, World Bank, Washington, DC, 2000.