The Journal of Applied Science **Vol. 21 No. 1 [2022]: 244482**
วารสารวิทยาศาสตร์ประยุกต์ doi: 10.14416/j.appsci.2022.01.001
ISSN 1513-7805 Printed in Thailand ISSN 2586-9663 (Online)

Research Article

Modelling the Household Debts in Thailand using Fay-Herriot models

Annop Angkunsit and Jiraphan Suntornchost*

Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

***E-mail**: Jiraphan.S@chula.ac.th

Received: 22/05/2021; Revised: 17/10/2021; Accepted: 28/03/2022

Abstract

In this study, we investigate the performance of the Fay-Herriot area level model to estimate the Household Debts of Thailand by using the 2017 household socio-economic survey from the National Statistics Office (NSO) of Thailand. Two versions of the model are considered which are the classical Fay-Herriot model and the Bayesian approach to the model, called as the Hierarchical Bayesian Fay-Herriot model. Results from our study show that the two models can reduce mean square etrrors of the direct estimates, particularly for the areas with large sampling variances. Moreover, we apply results from the two models to discuss behavior of household debts in Thailand by comparing household debts among different regions of Thailand. Moreover, we study the difference in household debts between municipal area and non-municipal area. From our analysis, household debts in municipal areas are higher than those of non-municipal areas. Comparing among different regions, we find that Bangkok has much higher average household debt than other regions. The northern and the northern east regions have higher household debts than the country average while the central, the east, and the southern regions have average household debts lower than the country average.

Keywords: Household debt, Fay-Herriot model, empirical best linear unbiased predictor, Bayesian.

Introduction

Due to current economic crisis, the rise of household debt is currently considered as one of the most important economic issues in many countries including Thailand. The problem of household debts affects the whole economy and living standards of people in the country. Therefore, several studies of debts in Thailand have been conducted in many different aspects. For example, Kiatipong et al. (2007) studied the wealth and debt of Thai household. Muthitacharoen et al. (2015) discussed the effect of the rise of household debts to the economic stability of Thailand. Visitwarakorn (2015) studied the debt of Thai's government due to the change of 2013 salary policy. Moreover, the Thailand's National Statistical Office also conducts regular surveys of the household debts of the Thai population and reports an estimate of annual household debt computed by using the information of household which is collected from all provinces in Thailand. Based on our current knowledge, these studies are performed based on the direct estimates from a survey. Therefore, the estimates are heavily influenced by

 doi: 10.14416/j.appsci.2022.01.001

the sampling design and the sampling errors particularly for areas with small sample sizes. To overcome the issues, alternative model-based methods have been proposed in the context of "small area statistics" (Rao & Molina, 2015) which are the statistical methods used when direct estimates from a survey are not of adequate precision, or in the situation of small sample sizes. The motivation of the method is to use statistical models to link the variable of interest with auxiliary information to define the model-based estimator that "borrow strength" from the related area. One well-known small area model applied in different applications is the Fay-Herriot model. The Fay-Herriot model was first introduced by Fay and Herriot in 1979 to predict the mean per capita income in small places within counties in USA. Since then, the model has been widely applied in different countries. For example, Srivastava et al. (2007) used the Fay-Herriot model to obtain the model-based district level estimates of the amount of loan outstanding per household in India. D'Alò et al. (2008) applied the Fay Herriot model to study socio-economic indicators in Italy. Wawrowski (2016) applied the spatial Fay-Herriot model to study poverty in Poland. For Thai data, Angkunsit & Suntornchost (2021) applied the bivariate Fay-Herriot model to obtain the model-based estimates of the average household income and average household expenditure for Thai data. In this paper, we study the performance of Fay-Herriot area level model to estimate the Household Debts of Thailand. Two versions of the Fay-Herriot model are considered which are the classical Fay-Herriot model and the hierarchical Bayesian Fay-Herriot model. The data used in our study are the Household Socio-Economic Survey 2017 and Population and Housing Census 2010 collected by the Thailand's National Statistical Office.

The organization of this paper is as follows. In Section 2, we introduce the Fay-Herriot model and the hierarchical Bayesian Fay-Herriot model. In this section, we also discuss estimation methods and model diagnostics of the two models. In Section 3, we describe the data source and auxiliary variables used in this study, model comparison, and discuss applications of the two models to the household debt of Thailand. Section 4 gives discussions and conclusions of our study.

Small Area Models

In this section, we first describe the small area models used in this paper which are the Fay-Herriot and the hierarchical Bayesian Fay-Herriot model.

2.1 Fay-Herriot model

The Fay-Herriot model was proposed by Fay & Herriot (1979) to predict the mean per capita income in small places within counties in USA. The structure of the Fay-Herriot model is as follows. Suppose that the population is partitioned into m subpopulations, called as "area". For any area i $(i=1,...,m)$, let θ_i be the true area mean of ith area and y_i be a direct estimator of θ_i. The model assumes that θ_i is linearly related to the auxiliary variables $\mathrm{x}_i'=(1,x_{i1},...,x_{ip})'$, for $i=1,...,m$, through the model

$$\theta_i=\mathrm{x}_i'\beta+v_i,\qquad v_i\sim N(0,A),\qquad i=1,...,m,$$

where $\beta=(\beta_0,\beta_1,...,\beta_p)$ is the vector of regression coefficients and A is the regression variance. The true mean is linked to direct estimator y_i through the following sampling model

 doi: 10.14416/j.appsci.2022.01.001

$$y_i = \theta_i + e_i, \quad e_i \sim N(0, D_i), \quad i = 1, ..., m,$$

where e_i is the sampling error with the known sampling variance D_i.
The Fay-Herriot model can be rewritten as

$$y_i = \mathrm{x}_i' \beta + v_i + e_i, \quad i = 1, ..., m, \tag{1}$$

where the area random effects v_i are independent of the sampling errors e_i.
If the area random effect variance A is known, the true area mean is estimated by minimizing the mean squared error (MSE) of θ_i proposed in Henderson (1975). The obtained estimate is called as the best linear unbiased prediction, or BLUP, defined as follows.

$$\tilde{\theta}_i = \frac{A}{A + D_i} y_i + \frac{D_i}{A + D_i} \mathrm{x}_i' \tilde{\beta}, \tag{2}$$

where $\tilde{\beta} = \left(\sum_{i=1}^m \mathrm{x}_i \mathrm{x}_i' (A + D_i)^{-1} \right)^{-1} \left(\sum_{i=1}^m \mathrm{x}_i \, y_i (A + D_i)^{-1} \right)$. However, A is unknown in general practice. Therefore, it is estimated by an estimate $\hat{A}$. Therefore, by substituting $\hat{A}$ into (2), we obtain the empirical BLUP or EBLUP $\hat{\theta}_i$ of θ_i. That is,

$$\hat{\theta}_i = \frac{\hat{A}}{\hat{A} + D_i} y_i + \frac{D_i}{\hat{A} + D_i} \mathrm{x}_i' \hat{\beta}, \tag{3}$$

where $\hat{\beta} = \tilde{\beta}(\hat{A}) = \left(\sum_{i=1}^m \mathrm{x}_i \mathrm{x}_i' (\hat{A} + D_i)^{-1} \right)^{-1} \left(\sum_{i=1}^m \mathrm{x}_i \, y_i (\hat{A} + D_i)^{-1} \right)$.

There are many estimation methods to obtain the estimate $\hat{A}$ available in literatures, for example, the Prasad-Rao method-of-momemts estimator (Prasad & Rao, 1990), the Fay-Herriot method-of-moment estimator (Fay & Herriot, 1979), the profile maximum likelihood estimator (Hartley & Rao, 1967), and the residual maximum likelihood estimator (Patterson & Thompsonm, 1971). Among these methods, the residual maximum likelihood estimation method, called as REML method, has been shown to outperform other methods because REML estimator is a second-order unbiased estimator of A. Therefore, in this study, we use the REML method to obtain the estimator of variance model variance $\hat{A}$.

The residual maximum likelihood estimator $\hat{A}$ can be obtained by maximizing the residual log-likelihood function

$$\ell(A) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\left| \sum_{i=1}^m \frac{\mathrm{x}_i \mathrm{x}_i'}{A + D_i} \right| - \frac{1}{2}\log\left| \operatorname*{diag}_{1 \le i \le m} A + D_i \right| - \frac{1}{2}\sum_{i=1}^m \frac{(y_i - \mathrm{x}_i' \tilde{\beta})'(y_i - \mathrm{x}_i' \tilde{\beta})}{A + D_i}.$$

 doi: 10.14416/j.appsci.2022.01.001

Using the REML estimator of A, a second-order unbiased (or nearly unbiased) estimator of MSE of the EBLUP is given by (Rao & Molina, 2015)

$$\hat{\mathrm{M}}[\hat{\theta}_i] = g_{1i}(\hat{B}_i) + g_{2i}(\hat{B}_i) + 2g_{3i}(\hat{B}_i), \tag{4}$$

where $\hat{B}_i = \frac{D_i}{\hat{A}+D_i}$, $g_{1i}(\hat{B}_i) = D_i(1-\hat{B}_i)$ is the MSE estimator of θ_i,

$g_{2i}(\hat{B}_i) = \hat{B}_i^2 \mathrm{x}_i' \left(\sum_{j=1}^m \mathrm{x}_j \mathrm{x}_j' \left(\frac{\hat{B}_j}{D_j} \right) \right)^{-1} \mathrm{x}_i$ is the excess in MSE estimator due to estimation of β,

$g_{3i}(\hat{B}_i) = 2\left(\frac{\hat{B}_i^3}{D_i} \right) \left(\sum_{j=1}^m \left(\frac{\hat{B}_j}{D_j} \right)^2 \right)^{-1}$ is the excess in MSE estimator due to estimation of A.

2.2 Hierarchical Bayesian Fay-Herriot model

In this section, we discuss the hierarchical Bayesian approach (Gelman et al., 1995; Albert, 2009) to the Fay-Herriot model, called as the hierarchical Bayesian Fay-Herriot model (Rao & Molina, 2015, Chapter 10). The model is then written as follows.

(i) $y_i \mid \theta_i, D_i \sim N(\theta_i, D_i), \quad i = 1,...,m,$ independent

(ii) $\theta_i \mid \beta, A \sim N(x_i'\beta, A), \quad i = 1,...,m,$ independent (5)

(iii) $A \sim f(A),\ \beta \sim g(\beta),$

where f and g are prior distributions of the model variance A and regression coefficient β, respectively. The prior distributions may be informative priors or noninformative priors. The informative priors can be determined based on substantial prior information such as previous study related to the data. However, informative priors are infrequently available. Therefore, noninformation priors are commonly used in many applications.

The hierarchical Bayes (HB) estimate of θ_i is the posterior mean given by

$$\mathrm{E}[\theta_i \mid \mathrm{y}] = \int \theta_i f(\theta_i \mid \mathrm{y})\, d\theta_i,$$

where $f(\theta_i \mid \mathrm{y})$ is the posterior density.

The corresponding posterior variance of θ_i is given by

$$\mathrm{Var}[\theta_i \mid \mathrm{y}] = \int (\theta_i - \mathrm{E}[\theta_i \mid \mathrm{y}])^2 f(\theta_i \mid \mathrm{y})\, d\theta_i,$$

Numerical estimates of posterior mean and variance can be obtained via the Markov chain Monte Carlo (MCMC) method (Rao & Molina, 2015). Having drawn an MCMC sample $\{(\beta^{(k,l)}, A^{(k,l)}),\ k = d+1,\ldots,d+K,\ l = 1,\ldots,L\}$ from the Markov Chain Monte Carlo method using L chains with the sample size where K of each chain and the burn in size of d, the posterior mean of θ_i can be obtained as

 doi: 10.14416/j.appsci.2022.01.001

$$\hat{\theta}_i^{\mathrm{HB}} = \frac{1}{LK}\sum_{l=1}^{L}\sum_{k=d+1}^{d+K}\hat{\theta}_i^{\mathrm{B}}(\beta^{(k,l)}, A^{(k,l)}) := \theta_i^{(\cdot,\cdot)}. \tag{6}$$

The corresponding posterior variance is

$$\mathrm{Var}[\theta_i \mid y] = \frac{d-1}{d}W_i + \frac{1}{d}B_i, \tag{7}$$

where $B_i = \frac{d}{L-1}\sum_{l=1}^{L}\left(\theta_i^{(\cdot,l)} - \theta_i^{(\cdot,\cdot)}\right)^2$ is the between-run variance and $W_i = \frac{1}{L(d-1)}\sum_{l=1}^{L}\sum_{k=d+1}^{d+K}\left(\theta_i^{(k,l)} - \theta_i^{(\cdot,l)}\right)^2$ is the within-run variance with $\theta_i^{(k,l)}$ is the k th retained value in the l th run of the length $2d$ with the first d burn-in iterations deleted such that $\theta_i^{(\cdot,l)} = \frac{1}{K}\sum_{k=d+1}^{d+K}\theta_i^{(k,l)}$ and $\theta_i^{(\cdot,\cdot)} = \frac{1}{L}\sum_{l=1}^{L}\theta_i^{(\cdot,l)}$.

The convergence of Markov chain Monte Carlo output can be examined by the "potential scale reduction factor (PSRF)" originally proposed by Gelman & Rubin's (1992). The PSRF is an estimated factor by which the scale of the current distribution for the target distribution. Each PSRF reduces to 1 as the number of iterations approaches infinity. The multivariate extension, multivariate potential scale reduction factor (MPSRF), was introduced by Brooks & Gelman (1998). The values of PSRF and MPSRF close to 1 indicate convergence of the chain.

The model validation for Bayesian models can be done via many different approaches such as the posterior predictive assessment introduced by Datta et al. (1999) and the divergence measure introduced by Laud & Ibrahim (1995) described below.

A posterior Predictive Assessment approach. (Datta et al., 1999)

We define $\mathrm{y}_{\mathrm{obs}}$ and $\mathrm{y}_{\mathrm{new}}$ to be the observed and the generated data, respectively. Then define $f(\theta \mid \mathrm{y}_{\mathrm{obs}})$, $f(d(\mathrm{y}_{\mathrm{obs}},\theta) \mid \mathrm{y}_{\mathrm{obs}})$, and $f(d(\mathrm{y}_{\mathrm{new}},\theta) \mid \mathrm{y}_{\mathrm{obs}})$ to be the posterior (predictive) predictions of θ, $d(\mathrm{y}_{\mathrm{obs}},\theta)$, and $d(\mathrm{y}_{\mathrm{new}},\theta)$, respectively. The discrepancy measure is

$$d(\mathrm{y},\theta) = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(y_{ij} - \theta_{ij})^2}{\sigma_{ij}},$$

where m is the length of y, n is the total number of iterations of the θ values, and σ_{ij} is the variance of y_{ij}. The value $P(d(\mathrm{y}_{\mathrm{new}},\theta) \geq d(\mathrm{y}_{\mathrm{obs}},\theta) \mid \mathrm{y}_{\mathrm{obs}})$ of a model close to 0.5 indicates that the model is adequately fit to the data. The extreme value (close to 0 or 1) indicates a lack of fit.

 doi: 10.14416/j.appsci.2022.01.001

Divergence measure. (Laud & Ibrahim, 1995)

The divergence measure between the observed data $\mathrm{y}_{\mathrm{obs}}$ and the generated data $\mathrm{y}_{\mathrm{new}}$ is defined as

$$d(\mathrm{y}_{\mathrm{new}}, \mathrm{y}_{\mathrm{obs}}) = \mathrm{E}[m^{-1} \| \mathrm{y}_{\mathrm{new}} - \mathrm{y}_{\mathrm{obs}} \|^2 | \, \mathrm{y}_{\mathrm{obs}}],$$

where m is the length of $\mathrm{y}_{\mathrm{obs}}$. An adequate model should produce a small value of the estimated divergence measure.

Data Analysis

In this section, we apply the Classical Fay-Herriot model and the Hierarchical Bayesian Fay-Herriot model to the average household debt data in Thailand demonstrated in previous section. The structure of this section is as follows. We first explain data description and variable selection of explanatory variables used in the study. We then discuss model diagnostics of the two models, compare numerical results of the two models with the classical direct estimators. Finally, we discuss the results on the average household debt in Thailand using estimates from the two models.

3.1 Data Description

The data used in this paper is the average household debt data in Thailand from the Household Socio-Economic Survey (SES) 2017 (National Statistical Office, 2017). The SES is conducted yearly by the National Statistical Office Thailand. The total sample of SES 2017 is 43,210 households which are distributed in 77 provinces. Each province (except Bangkok) is divided into two parts according to the type of local administration area, namely, municipal area and non-municipal area. The design sampling of SES is a stratified two-stage sampling. The basic concepts of stratified two-stage sampling is as follows. (Sukhatme, 1984; Särndal et al., 1992; and Cochran, 1997). The population U is partitioned into H separate strata $U_h \ (h = 1, ..., H)$. For each stratum U_h, consists of N_h separate primary sampling units, called as PSU, U_{hi} for $i = 1, ..., N_h$. For each PSU U_{hi}, consists of M_{hi} secondary sampling units, called as SSU. Let y_{hij} be the values of the quantity of interest in the j th SSU of the i th PSU from the h th stratum. The population mean per second-stage unit in the h th stratum is $\bar{Y}_h = \frac{1}{M_{h0}} \sum_{i=1}^{N_h} \sum_{j=1}^{M_{hi}} y_{hij}$, where $M_{h0} = \sum_{i=1}^{N_h} M_{hi} = N_h \bar{M}_h$. The estimate of the population mean per second-stage unit in the h th stratum is

$$\bar{y}_h = \frac{1}{N_h \bar{M}_h} \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij} = \frac{1}{n_h \bar{M}_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij}, \tag{8}$$

where n_h is a sample size of the PSUs from the h th stratum, and m_{hi} is the sample size of SSUs from the i th PSU from the h th stratum. The variance of estimate of the population mean is given by

$$\mathrm{Var}[\bar{y}_h] = \left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_h^2 + \frac{1}{n_h N_h} \sum_{i=1}^{N_h} \frac{M_{hi}^2}{\bar{M}_h^2} \left(\frac{1}{m_{hi}} - \frac{1}{M_{hi}}\right) S_{hi}^2, \tag{9}$$

where $S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(\frac{M_{hi}}{\bar{M}_h} \bar{y}_{hi} - \bar{Y}_h \right)^2$ is the variance among primary unit means and $S_{hi}^2 = \frac{1}{M_{hi} - 1} \sum_{j=1}^{M_{hi}} \left(y_{hij} - \bar{y}_{hi} \right)^2$ is the variance among subunits within primary units. The estimator of the variance is obtained by replacing S_h^2 and S_{hi}^2 with their sample estimators and summing up by sampled PSUs in each h th stratum.

The area-specific explanatory variables used in this paper are chosen by an Akaike's Information Criteria (AIC) backward selection from 12 variables available in the Population and Housing Census 2010 file (National Statistical Office, 2010). The 12 variables in the Housing Census 2010 file, before backward variable selection, are as follows: 1) the proportion of households with detached house; 2) the proportion of households that own living quarters; 3) the proportion of households that rent living quarters; 4) the proportion of households that cement or brick dwellings; 5) the proportion of households that own land; 6) the proportion of households using gas for cooking; 7) the proportion of households using sitting toilet; 8) the proportion of male populations; 9) the average population per private household; 10) the proportion of households that working household head; 11) the proportion of populations that graduated from upper secondary level; 12) the proportion of populations that graduated from higher level.

Having performed the backward variable selection, there are six significantly explanatory variables to be used in our models as shown in Table 1.

Table 1 Explanatory variables

Variable	Notation
Proportion of households that rent living quarters	x_1
Proportion of households that cement or brick dwellings	x_2
Proportion of households that own land	x_3
Proportion of households using gas for cooking	x_4
Proportion of households using sitting toilet	x_5
Proportion of male populations	x_6

3.2 Model diagnostics.

3.2.1 Model diagnostics for the Fay-Herriot model

In this section, we apply the Fay-Herriot model (1) to the average household debt for 153 areas: Bangkok and two areas from each of other 76 provinces which are municipal area and non-municipal area. The direct estimates and sampling variances are obtained from (8) and (9), respectively. The EBLUP estimates for the household debts and a second-order unbiased estimators of MSE of the EBLUP are obtained from (3) and (4), respectively. To diagnose the model, we apply the residual analysis discussed in Erciulescu et al. (2021). We consider the standardized residuals

 doi: 10.14416/j.appsci.2022.01.001

$$r_i = \frac{y_i - x_i'\hat{\beta}}{\sqrt{\mathrm{Var}[y_i - x_i'\hat{\beta}]}},$$

where $\mathrm{Var}[y_i - x_i'\hat{\beta}] = D_i + A$. Since the parameter A is unknown, the variance is estimated by $D_i + \hat{A}$, where $\hat{A}$ is the estimator of A.

The normality analysis used are 1) Q-Q plot; 2) histogram; 3) Kolmogorov-Smirnov test; 4) Shapiro-Wilk normality test, of standardized residuals of direct estimate from synthetic part.

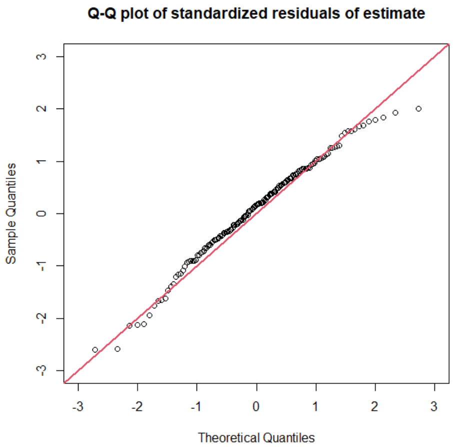


Figure 1 Q-Q plot of standardized residuals of estimate

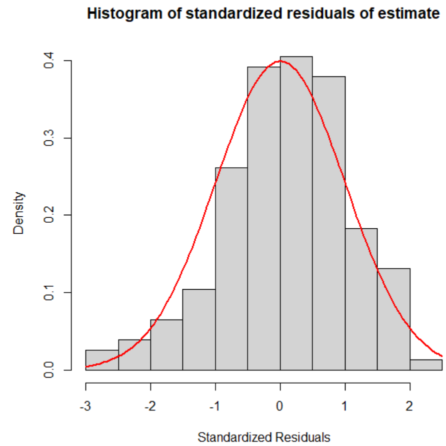


Figure 2 Histogram of standardized residuals of estimate

From Figures 1 - 2, we can see that the standardized residuals from the synthetic part follow normal distribution. Moreover, we also perform the Kolmogorov-Smirnov and Shapiro-Wilk normality tests for the standardized residuals. The corresponding p-values of the tests are 0.3940 and 0.0784, respectively. From the three analyses, we can conclude that the standardized residuals follow normal distribution at the significant level $\alpha = 0.05$.

3.2.2 Model diagnostic for the Hierarchical Bayesian Fay-Herriot model.

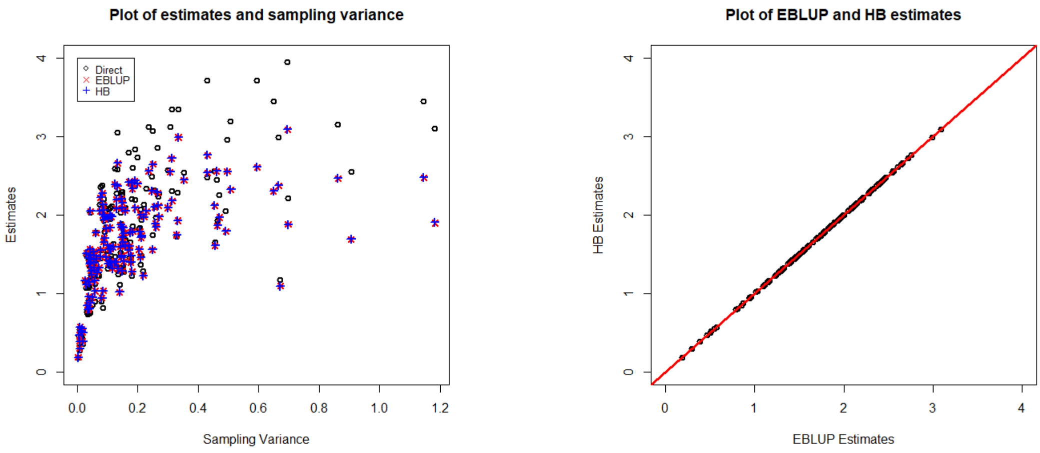
In this section, we apply the hierarchical Bayesian Fay-Herriot model (5) to the average household debts. Since there is no further information available, we use noninformative prior distributions for the unknown parameters. Several choices of noninformation priors can be applied such as uniform distribution, normal distribution, and inverse-gamma distribution. In this study, we use common choices which are the uniform prior for the variance and normal prior for the regression coefficients. To be specific, the prior distributions of the model variance A and regression coefficients β are $A \sim U(0,10^8)$ and $\beta \sim N_7(0,10^6\,\mathrm{I}_7)$, respectively. The MCMC samples are obtained by using **R2OpenBUGS** (Thomas, 2020), **R2WinBUGS** (Ligges, 2015) and **coda** (Plummer, 2020) packages in R program (R Core Team, 2020). The setting of the Monte Carlo simulation is as follows: the number of Markov chains is 3, the sample size per chain is 500000, the length of burn in is 250000, and the thinning rate is 2. Having obtained MCMC sample, we apply the PSRF and MPSRF convergence tests and the model diagnostics mentioned in Section 2. The values of PSRFs are in the range of (0.9999,

 doi: 10.14416/j.appsci.2022.01.001

1.0001) and the MPSRF is 1.0007 which are close to 1. Therefore, the PSRFs and MPSRF results show the convergence of the MCMC chains. The posterior Predictive Assessment value is 0.4792 which is close to 0.5, and the divergence measure value is 0.3922 which is relatively small. The two model validations indicate an adequate fit of the model to the data.

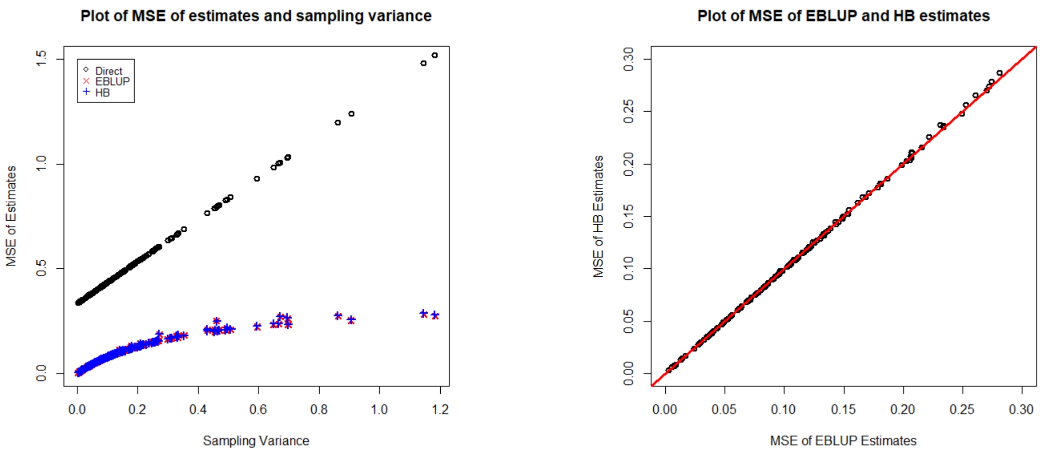
3.3 Model Comparisons.

In this section, we compare performances of the two models with direct estimates. The values of direct, EBLUP and HB estimators of average household debts ordered by the magnitudes of sampling errors are demonstrated in Figure 3. The corresponding MSEs are shown in Figure 4.



(a) (b)

Figure 3 Plot of direct, EBLUP and HB estimates of average household debt



(a) (b)

Figure 4 Plot of MSE of direct, EBLUP and HB estimates of household debt

From Figure 3 (a), we can see that EBLUP and HB estimates are close to direct estimates for the areas with small sampling variances but the two estimates are different from the direct

 doi: 10.14416/j.appsci.2022.01.001

estimates for the areas with large sampling variances. From Figure 3 (b), we can see that the EBLUP and HB estimates are almost identical as the plot between the two estimates goes along with the 45-degree line. From Figure 4 (a), we can see that the MSEs of the direct estimates are larger than MSEs of the EBLUP and HB estimates particularly for areas with large sampling variances. From Figure 4 (b), the MSEs of the direct estimates steady increase when sampling variances increase, while the MSEs of the EBLUP and HB estimates seem to be stable for a certain value of sampling variance, at 0.3 in this dataset. Results from Figure 3 and Figure 4 indicate the lack of fit of the direct estimates for data for areas with large sampling variances while the Fay-Herriot model and the hierarchical Bayesian Fay-Herriot model can improve estimates for such cases by reducing the MSEs. Comparing the the Fay-Herriot model and the hierarchical Bayesian Fay-Herriot model, Figure 3 (b) and Figure 4 (b) show that two estimates are approximately the same.

3.4 Discussions on the Household Debts in Thailand

In this section, we discuss the household debts in Thailand by using estimates from the Fay-Herriot model and the hierarchical Bayesian Fay-Herriot. Table 2 shows average household debts for Bangkok and the five regions of Thailand. The average household debts are presented for municipal area, non-municipal area, and the combined estimates. Figure 5 shows the boxplots of average household debts for each region. From Table 2 and Figure 5, we can see that the average household debts in municipal areas are higher than those of non-municipal areas for all regions. Moreover, the average household debts in Bangkok (BKK) are higher than average debts in other regions of Thailand. Comparing with the country average (1.7023 for EBLUP, and 1.7030 for HB), we can see that Bangkok, the northern, and the northeast regions have higher debts than the country average while the central, the east and the southern regions have lower debts than the country average.

Table 2 The model-based estimates of average household debt (Unit: 100,000 Baht)

Regions	Municipality	Size	Mean		Standard Deviation	
			EBLUP	HB	EBLUP	HB
Bangkok		1	2.0470	2.0467	-	-
Central	**Total**	**36**	**1.6759**	**1.6771**	**0.6792**	**0.6807**
	Municipal area	18	1.7899	1.7910	0.6527	0.6542
	Non- Municipal area	18	1.5619	1.5631	0.7044	0.7061
East	**Total**	**14**	**1.4838**	**1.4839**	**0.3974**	**0.3980**
	Municipal area	7	1.5983	1.5980	0.3673	0.3677
	Non- Municipal area	7	1.3694	1.3697	0.4205	0.4215
North	**Total**	**34**	**1.7448**	**1.7455**	**0.5813**	**0.5819**
	Municipal area	17	1.9571	1.9577	0.5673	0.5680
	Non- Municipal area	17	1.5325	1.5333	0.5286	0.5293
Northeast	**Total**	**40**	**1.8136**	**1.8144**	**0.4798**	**0.4813**
	Municipal area	20	2.0366	2.0381	0.4537	0.4555
	Non- Municipal area	20	1.5907	1.5907	0.4025	0.4033
South	**Total**	**28**	**1.6223**	**1.6231**	**0.6074**	**0.6082**
	Municipal area	14	1.8638	1.8642	0.5789	0.5793
	Non- Municipal area	14	1.3807	1.3819	0.5529	0.5544
Total	**Total**	**153**	**1.7023**	**1.7030**	**0.5726**	**0.5737**
	Municipal area	76	1.8882	1.8890	0.5509	0.5520
	Non- Municipal area	76	1.5118	1.5125	0.5364	0.5375

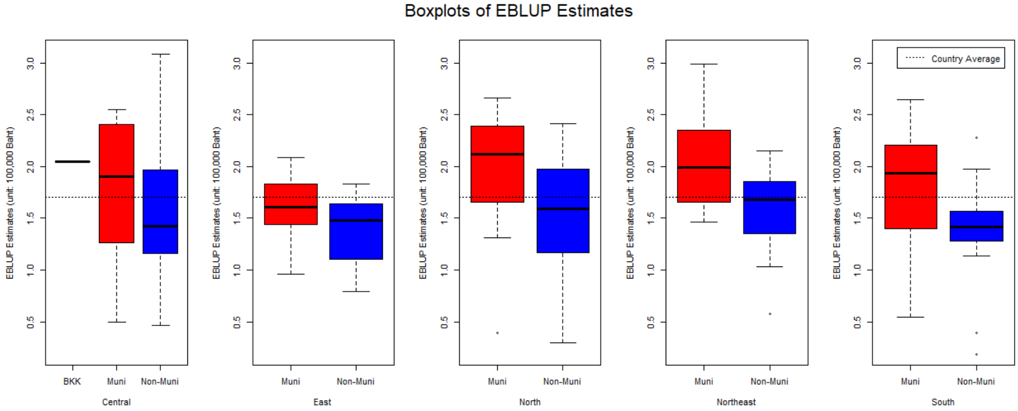


Figure 5 Boxplots of EBLUP estimates for average household debts in different regions

Discussion and Conclusions

In this paper, we applied the two area level models: the Fay-Herriot model and the hierarchical Bayesian Fay-Herriot model to obtain are level estimates of household debts where the area considered in this paper is the municipality by province. The model diagnostics of the two models indicate adequate fit to the data. Model comparisons show that the two small area models outperform the direct estimates especially for areas with large sampling variances. The study indicates that model-based estimates can improve the direct estimates for the household debt data. From our study, we believe that the Fay-Herriot model can be applied to other applications with small sample sizes to overcome the bias of the direct survey estimats caused by sampling errors. Some improvements of our study can be performed such as incorporating time effect into the model to study trends of household debts over years.

Acknowledgement

The authors would like to thank the National Statistical Office Thailand for providing the Household Socio-Economic Survey 2017 data and Population and Housing Census 2010 data used in our study.

References

Albert, J. (2009). *Bayesian Computation with R* (2nd ed.). Springer, New York.

Angkunsit, A. & Suntornchost, J. (2021). Bivariate Fay-Herriot models with application to Thai socio-economic data. *Naresuan University Journal: Science and Technology, 29* (1), 36-49. doi: 10.14456/nujst.2021.3

Brooks, S. P. & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics, 7* (4), 434-455. doi: 10.1080/10618600.1998.10474787

Cochran, W.G. (1997). *Sampling Techniques* (3rd ed.). John Wiley and Sons, Inc., New York.

Datta, G., Lahiri, P., Maiti, T. & Lu, K. (1999). Hierarchical Bayes estimation of unemployment rates for the states of the U.S. *Journal of the American Statistical Association, 94* (448), 1074-1082. doi: 10.2307/2669921

D'Alò, M., Consiglio, L.D., Falorsi, S. & Solari, F. (2008). Small area estimation methods for socio-economic indicators in household surveys. *Rivista Internazionale Di Scienze Sociali, 116* (4), 419-441.

 doi: 10.14416/j.appsci.2022.01.001

Erciulescu, A. L., Franco, C. & Lahiri, P. (2021). Use of Administrative Records in Small Area Estimation. In A. Y. Chun & M. Larsen (Eds.), *Administrative Records for Survey Methodology*, pp. 231-267. John Wiley and Sons, Inc., New York.

Fay, R. E. & Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedure to census data. *Journal of the American Statistical Association, 74* (366), 269-277. doi: 10.2307/2286322

Gelman, A., Carlin, J., Stern, H. & Rubin, D. (1995). *Bayesian data analysis.* Chapman & Hall, London.

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7* (4), 457-472. doi: 10.1214/ss/1177011136

Hartley, H.O. & Rao, J.N.K. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, *54* (1/2), 93–108.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under selection model. *Biometrics, 31* (2), 423-447. doi: 10.2307/2529430

Kiatipong A., Wilatluk S. & Nalin C. (2007). *The Wealth and Debt of Thai Household: Risk Management and Financial Access.* Bangkok, Thailand: Bank of Thailand Symposium 2007.

Laud, P.W. & Ibrahim, J.G. (1995). Predictive model selection. *Journal of the Royal Statistical Society. Series B (Methodological), 57* (1), 247-262. doi: 10.1111/j.2517-6161.1995.tb02028.x

Liggers, U. (2015) Running 'WinBUGS' and 'OpenBUGS' from 'R' / 'S-PLUS'. https://cran.r-project.org/web/packages/R2WinBUGS/index.html

Muthitacharoen, A., Nuntramas P., & Chotewattanakul P. (2015). Rising household debt: Implications for economic stability. *Thailand and The World Economy, 33* (3), 43-65.

National Statistical Office (2010). Population and Housing Census 2010. Thailand.

National Statistical Office (2017). The Household Socio-Economic Survey 2017. Thailand.

Patterson, H.D., Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika, 58* (3), 545–554.

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

Plummer M. (2020). *Output Analysis and Diagnostics for MCMC.* https://cran.r-project.org/web/packages/coda/index.html

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rao, J. N. K. & Molina, I. (2015). *Small Area Estimation* (2nd ed.). John Wiley and Sons Inc., New York.

Särndal, C.E., Swensson B. & Wretman, J. (1992). *Model Assisted Survey Sampling.* Spring-Verlag, New York.

Srivastava, A.K., Sud, U.C. & Chandra, H. (2007). Small area estimation – an application to national sample survey data. *Journal of the Indian Society of Agricultural Statistics, 61* (2), 249-254.

Sukhatme, P.V. (1984). *Sampling Theory of Surveys with Applications.* Ioxa State University Press, Ames, IA.

Thomas N. (2020). *Running OpenBUGS from R.* https://cran.r-project.org/web/packages/R2 OpenBUGS/index.html

 doi: 10.14416/j.appsci.2022.01.001

Visitwarakorn, W. (2015). Debt of government officers after the 2013 salary policy: A case study of the office of the permanent secretary, ministry of natural resources and environment, *Varidian E Journal, 1*, 515-529.

Wawrowski, L. (2016). The spatial Fay-Herriot model in poverty estimation. *Folia Oeconomica Stetinensia, 16* (2), 191-202. doi: 10.1515/foli-2016-0034