Statistics and Applications {ISSN 2454-7395 (online)}
Special Issue in Memory of Prof. C R Rao
Volume 22, No. 3, 2024 (New Series), pp 449–469
http://www.ssca.org.in/journal

# Hierarchical Bayes Small Area Estimation from Aggregated Data using Various Spatial Models

**Jiacheng Li[1], Hee Cheol Chung[2], David Okech[3] and Gauri S. Datta[4,5]**
[1] *Wells Fargo Bank, Charlotte, NC*
[2] *Department of Mathematics and Statistics, University of North Carolina, Charlotte, NC*
[3] *School of Social Work, University of Georgia, Athens, GA*
[4] *Department of Statistics, University of Georgia, Athens, GA*
[5] *Center for Statistical Research and Methodology, U.S. Census Bureau, Suitland, MD*

Received: 04 August 2024; Revised: 17 September 2024; Accepted: 20 September 2024

## Abstract

Small area estimation methods are important tools for applied statisticians to help policymakers in need of reliable statistics for lower level disaggregated populations. While aggregated statistics at the higher level may be available from surveys, they are not useful to estimate characteristics for lower level subpopulations. Often useful covariates for these subpopulations are available, which can be integrated through innovative small area estimation methodology to leverage aggregated data to produce better estimates and measures of uncertainty for the disaggregated subpopulation means.

To serve our need we generalize the celebrated Fay-Herriot model, which has been extensively used for several decades by many National Statistical Offices around the world, to produce reliable small area statistics. We consider the traditional independence for the Fay-Herriot linking model errors as well as various important spatially dependent models for these errors. We conduct a hierarchical Bayesian analysis for all these models based on a popular class of noninformative improper prior densities for the linking model parameters. We illustrate the usefulness of our proposal by producing estimates of statewide four-person family median incomes for the U.S. states for the year 1990. We create for our illustration the aggregated statistics from the 1990 Current Population Survey. We evaluate the accuracy of our state predictions against the corresponding incomes, deemed to be reliable, produced by the 1990 Census. For all models and for all improper prior densities for the model parameters considered here we prove the propriety of the resulting posterior distributions. The result in Corollary 1 of Chung and Datta (2022, *Survey Methodology*, vol. 48, No. 2, pp. 463-489) follows as a special case. Our empirical assessments amply demonstrate the usefulness of our novel approach.

*Key words:* Aggregated statistics; Conditional autoregression; Current Population Survey; Fay-Herriot model; LCAR; Simple CAR; Simultaneous AR.

**AMS Subject Classifications:** 62K05, 05B05

Corresponding Author: Gauri S. Datta
Email: gauri@uga.edu

## 1. Introduction

Preparation and implementation of effective social welfare and human development policy proposals require reliable statistics measuring important population characteristics, for example, income, employment, education, healthcare, agricultural productions and environmental safety. National statistical offices (NSO's) around the world collect relevant data and produce these statistics. Many nations and international organizations recognize the need for these statistics at the national level as well as at su-bnational/sub-population levels. These sub-populations may be geographic (states, counties or districts), demographic (gender, race, age) or cross-classification of geographic and demographic factors (state level poverty rates for the school-age children).

The NSOs and international organizations, for example, the World Bank, rely on appropriate data to produce relevant statistics. Since national censuses are carried out every five or ten years, these data will fail to capture the current state of the population when the last census gets outdated. Decennial or quinquennial censuses are expensive. To gather timely and less expensive data the NSOs conduct carefully planned sample surveys to collect data from only a fraction of the population. It is well-documented in the statistics literature that carefully planned surveys with reasonably large samples can be as accurate as a census.

Even if a nation may be doing well overall, often various segments of the nation may not be doing as well. While any functioning government that cares to serve its people requires accurate data for the entire nation, it also needs reliable disaggregated data for various segments of the nation. For example, the U.S. government has mandated it by law to produce timely and accurate disaggregated statistics measuring income, employment and health service for various demographic groups at the county or state level. The European Union and the United Nations have many programs that require accurate poverty and income information for many geographic/demographic sub-populations. Production of reliable, disaggregated statistics is known as small area estimation in survey sampling.

Sample surveys are generally designed to provide useful data in estimating various characteristics of a population of interest. Sample sizes are so chosen to ensure that traditional design-based estimators are adequately accurate. Sample size is usually the key thing, and when it comes to estimating a sub-population characteristic, based solely on the part of the original sample which is in the sub-population, the sub-sample may be small or empty. The version of the national level design-based direct estimate from the sub-sample for a sub-population, if it has enough sample to be computed, may be highly variable, or may be non-existent due to lack of sample. Sub-populations with low or no sample size to produce reliable direct estimates are known as small areas. Due to limited resources, a survey, by design, may not allocate any sample to many sub-populations. For example, the American Community Survey (ACS) is conducted to produce reliable statistics for nearly three thousand U.S. counties. However, the ACS usually samples about one-third of the counties, resulting in many non-sampled small areas. Post-surveys some sub-populations may also be defined for the current need, and there may not be any units selected from these sub-populations. Again, resource constraints do not permit selection of new sample to transform unreliable or unavailable small area estimates to reliable ones. To increase the accuracy of inadequate direct estimates of small areas (or to produce estimates for non-sampled areas), statistical methods advocate model-based approach to enable borrowing information from direct estimates of other domains and other data sources. In many applications, other related surveys and administrative data provide useful covariates. A model-based estimate of an area is produced by suitably shrinking a direct estimate (if available) to a synthetic estimate of a regression function based on auxiliary variables.

In small area estimation if unit-level data are available, a unit-level small area model by Battese *et al.* (1988) is often recommended for modeling. However, in many applications to protect confidentiality of the respondents the organization conducting the survey releases only summary data at the area-level for the areas sampled. In this setup, Fay and Herriot (1979) introduced an area-level model. This popular model is known in small area estimation as the Fay-Herriot model. In this model estimating the small area mean $\theta_i$ for a small area $i$, if its direct estimator $Y_i$ is available, it is called a *sampled area*. We assume that $Y_i$ is unbiased for $\theta_i$. No direct estimator is available for an *unsampled area*.

Fay and Herriot (1979) proposed a linking model for all $m$ small area means $\theta_i$ based on a multiple linear regression of the $\theta_i$'s on some available suitable covariates $\mathbf{x}_i$. For a sampled area the model-based estimator of $\theta_i$ is obtained by shrinking its direct estimator $Y_i$ to the synthetic regression estimator $\mathbf{x}_i^T\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is an estimator of the regression coefficient $\boldsymbol{\beta}$ in the regression mean function $\mathbf{x}_i^T\boldsymbol{\beta}$. If an area is unsampled, synthetic estimator $\mathbf{x}_i^T\hat{\boldsymbol{\beta}}$ is the small area estimator of $\theta_i$.

In small area estimation a population is partitioned into $m$ sub-populations, and a survey design samples $m - m_1$ sub-populations and does not sample the other $m_1$ sub-populations (sometimes $m_1 = 0$ but for the ACS it is positive). The Fay-Herriot model described above uses the $m - m_1$ direct estimators and covariate $\mathbf{x}_i$ from all $m$ areas to estimate $\theta_i, i$th sub-population mean, $i = 1, \ldots, m$.

From cost and administrative considerations a survey, by design, may merge $t_1$ sub-populations and select a sample from this combined bigger sub-population. Suppose a direct estimator $S_1$ from this sample estimates $\eta_1$, where, for example, $\eta_1$ may represent the total employment or total healthcare expenditure, then it is equal to the sum of $\theta_i$'s for these $t_1$ sub-populations. In general, we assume that $\eta_1$ is a known linear combination of the $t_1$ $\theta_i$'s. Similarly, $t_2$ other sub-populations may be merged for sampling, and a direct estimator $S_2$ from a sample from this merged sub-populations may be formed which estimates the corresponding population characteristic $\eta_2$. Again, we assume that $\eta_2$ is a known linear combination of $t_2$ $\theta_i$'s. In this way, an estimator $S_r$ is obtained which is an unbiased estimator of $\eta_r$, where $\eta_r$ is a known linear combination of $\theta_i$'s. This setup is the motivation of the problem that we will consider here. We assume that we have an $r \times 1$ vector of estimators $\mathbf{S}$ with its associated variance-covariance matrix $\mathbf{D}_S$. We assume that $\mathbf{S}$ is an unbiased estimator of $\mathbf{C}\boldsymbol{\theta}$ for an $r \times m$ known matrix $\mathbf{C}$. We assume that rank of $\mathbf{C}$ is $r$ and that $\mathbf{D}_S$ is a known, positive definite (p.d.) matrix. If $r = m - m_1$ and each of the rows of $\mathbf{C}$ has all elements 0 and one element 1 (first element in the first row, the second in the second row, etc.), then $\eta_1 = \theta_1,\ \eta_2 = \theta_2$, etc. and we get the traditional Fay-Herriot setup (cf. Fay and Herriot (1979)).

Alternatively, in the Fay-Herriot setup, suppose an area $i$ is an union of $n_i$ sub-areas and we are interested in estimating the sub-area mean $\theta_{ij}$ based on available covariates $\mathbf{x}_{ij}$ from that area. The $i$th area mean $\eta_i$ is a known linear combination of the sub-area means $\theta_{ij}$'s. A direct estimator $Y_i$ is available for $\eta_i$ but there are no direct estimators of $\theta_{ij}$ for the sub-areas. Our goal is to estimate the $\theta_{ij}$'s based on the survey estimates $Y_i$'s and the sub-area covariates $\mathbf{x}_{ij}$'s. To address this problem we are expanding the scope of the traditional Fay-Herriot model. Note that there is no direct estimate of $\theta_{ij}$. We use $\boldsymbol{\theta}_i$ to denote the vector $(\theta_{i1}, \ldots, \theta_{in_i})^T$ and use the traditional independent Fay-Herriot linking model where $\theta_{ij} \overset{ind}{\sim} N(\mu_{ij}, \sigma^2), j = 1, \ldots, n_i, i = 1, \ldots, m$, where for simplicity of presentation we assume that $\mu_{ij}$'s and $\sigma^2$ known. Suppose the survey estimator $Y_i$ is normally distributed variance $D_i$ and mean $\sum_{j=1}^{n_i} c_{ij}\theta_{ij}$, the coefficients $c_{ij}$'s are known. Under this setup, simple algebra shows that (if we invoke a Bayesian setup), the posterior mean of $\theta_{ij}$ is $\tilde{\theta}_{ij} = \mu_{ij} + \{\sigma^2 c_{ij}/(D_i + \sigma^2 \sum_{j=1}^{n_i} c_{ij}^2)\}(Y_i - \sum_{j=1}^{n_i} c_{ij}\mu_{ij})$, and the posterior variance

$$\tilde{\sigma}_{ij}^2 = \frac{\sigma^2\{D_i + \sigma^2 \sum_{k\neq j} c_{ik}^2\}}{D_i + \sigma^2 \sum_{j=1}^{n_i} c_{ij}^2}. \tag{1}$$

This result makes sense. Since a $\theta_{ij}$ appears only in the distribution of $Y_i$ and since all the $\theta_{ij}$'s are independent, it follows that $Y_i|\theta_{ij} \sim N(c_{ij}(\theta_{ij} - \mu_{ij}) + \sum_{k=1}^{n_i} c_{ik}\mu_{ik}, D_i + \sigma^2 \sum_{k\neq j} c_{ik}^2)$ and $\theta_{ij} \sim N(\mu_{ij}, \sigma^2)$. These two distributions imply that $\theta_{ij}|y_i \sim N(\tilde{\theta}_{ij}, \tilde{\sigma}_{ij}^2)$. If $r = [m/2]$, and $n_i = 2$ for $i = 1, \ldots, r$, and $c_{i1} = c_{i2} = 1$, then $\tilde{\theta}_{i1} = \tilde{\theta}_{i2} + \mu_{i1} - \mu_{i2}$, and $\tilde{\sigma}_{i1}^2 = \tilde{\sigma}_{i2}^2$. If $n_i = 1$ and $c_{i1} = 1$, the above expressions for the posterior mean and the variance for $\theta_{i1}$ will reduce to the results from the regular independent Fay-Herriot model.

For a comprehensive literature on small area estimation we refer to Rao and Molina (2015) who documented the need for reliable small area statistics in many applications in agriculture, education, healthcare, economy and industry. Here is an outline of the article. In Section 2 we presented a generalized Fay-Herriot model for aggregated small area statistics. We introduced the hierarchical Bayes (HB) model as well as the distribution of Fay-Herriot linking model error under various spatial models. In Subsection 2.1, we introduced the neighborhood matrix, an important element in spatial modeling. We outlined some useful properties of the eigenvalues of this matrix and those of a couple of other matrices defined from this matrix. In Section 3, we presented a set of sufficient conditions to ensure the propriety of all the posterior distributions that result from the class of HB models and a class of noninformative improper prior pdf's introduced the last section. We illustrated our novel ideas in Section 4 to the estimation of four-person households median incomes of the forty-nine contiguous states of the US. Section 5 reviews the importance of the proposed methodology. Finally, Section 6 presents detailed arguments to prove the propriety of the posterior pdf's for a couple of spatial models, and how these arguments can be modified for the remaining models.

## 2. A generalized Fay-Herriot model for aggregated statistics

As it was described in Section 1, the aggregated statistics $\mathbf{S}$ is assumed to be an unbiased estimator of $\mathbf{C}\boldsymbol{\theta}$. We present below an extended version of the popular Fay-Herriot model to draw inference for $\boldsymbol{\theta}$ based on the aggregated data $\mathbf{S}$. The $r \times m$ matrix $\mathbf{C}$ is an appropriate known matrix, described further in Remark 1.

**The HB model:**

(a) $\mathbf{S}|\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2, \rho \sim N(\mathbf{C}\boldsymbol{\theta}, \mathbf{D}_S)$,

(b) $\boldsymbol{\theta}|\boldsymbol{\beta}, \sigma^2, \rho \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{\Omega}^{-1}(\rho))$,

(c) The prior pdf for $\boldsymbol{\beta}, \sigma^2$ and $\rho$ is

$$\pi(\boldsymbol{\beta}, \sigma^2, \rho) = \pi(\boldsymbol{\beta}) \times g(\sigma^2) \times h(\rho), \tag{2}$$

where $\pi(\boldsymbol{\beta})$ is a bounded positive function corresponding to a prior pdf (may be improper), $g(\sigma^2)$ is an appropriate (may also be improper) prior, and $h(\rho)$ is a proper pdf for $\rho$ defined on an appropriate finite interval.

For the model above, $\mathbf{D}_S$ is a known p.d. matrix. Also, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m]^T$ is an $m \times p$ matrix of covariates, with rank $p$. The regression coefficient $\boldsymbol{\beta}$ is a $p \times 1$ vector. For a particular model in (b), namely, the independent Fay-Herriot model, the matrix $\boldsymbol{\Omega} = \mathbf{I}_m$ is free from $\rho$. In this case, the model being free from $\rho$, a prior for $\rho$ is not required. However, we can use any proper prior for $\rho$ and the posterior pdf of $\rho$ will be the same as the prior pdf. An improper uniform prior $\pi(\boldsymbol{\beta}) = 1$ is extensively used in the Bayesian literature (see, for example, Berger (1985) and Ghosh (1992)).

**Remark 1:** The part (b) of the above hierarchical model is known as the *linking model* (see Rao and Molina (2015)). In order for the sampling and the linking models in the above hierarchical model to be capable of producing inference for $\boldsymbol{\beta}$ under the frequentist setup (without part (c) for prior specification), the matrix $\mathbf{C}$ needs to have certain structure. In particular, the row space of $\mathbf{CX}$ must be the same as that of $\mathbf{X}$. It is equivalent to $rank(\mathbf{CX}) = rank(\mathbf{X})$, the estimability requirement of $\boldsymbol{\beta}$ based on the design matrix for $\mathbf{S}$. It implies that $r \geq rank(\mathbf{C}) \geq rank(\mathbf{CX}) = rank(\mathbf{X}) = p$ is a necessary condition on $r$.

The part (b) of the above hierarchical model implies a representation for the $i$th component of $\boldsymbol{\theta}$, which is given by

$$\theta_i = \mathbf{x}_i^T\boldsymbol{\beta} + v_i, \ i = 1, \ldots, m, \tag{3}$$

where the $v_i$'s are also called random effects in mixed linear model. This decomposition implies that the random effects vector $\mathbf{v} = (v_1, \ldots, v_m)^T$ is normally distributed with mean vector $\mathbf{0}$ and variance-covariance matrix $\sigma^2\boldsymbol{\Omega}^{-1}(\rho)$. We appropriately choose various forms of $m \times m$ the p.d. matrix $\boldsymbol{\Omega}$ to specify a class of models for $\boldsymbol{\theta}$. For the independent Fay-Herriot model, $\boldsymbol{\Omega} = \mathbf{I}_m$ which means that the $\theta_i$'s are independently distributed. It is not unreasonable to anticipate that if effective covariates are available, they can capture most of the variability of the $\theta_i$'s. Any unexplained variation among the $\theta_i$'s will be modeled by the random effects, and across small areas these random effects will not have any particular pattern. This variability may be modeled through $\boldsymbol{\Omega} = \mathbf{I}_m$.

While the independent Fay-Herriot model is the default model, in a recent paper Chung and Datta (2022) showed that in the absence of good covariates some spatially-dependent models for the random effects vector improve the prediction of $\theta_i$'s. In our case when a majority of small areas have no direct estimators, and only a few (no less than $p$) aggregated statistics are available that estimate some linear combinations of the small area mean vector $\boldsymbol{\theta}$, importance of both effective covariates and good linking models explaining the dependence of the components of $\boldsymbol{\theta}$ cannot be overemphasized.

### 2.1. A neighborhood matrix for spatial models with some useful results

For the Fay-Herriot model, Chung and Datta (2022) considered four different spatial models for the random effects and showed that many of these models yielded better predictions of small area means for the non-sampled areas than the independent Fay-Herriot model. They considered four spatial models which are determined by suitable structure of the matrix $\mathbf{\Omega}$. These $\mathbf{\Omega}$ matrices depend on a special matrix $\mathbf{W}$, known as neighborhood or *incidence* matrix.

The incidence matrix $\mathbf{W}$ is determined by the neighborhood structure of the small areas. This matrix is an $m \times m$ non-null, symmetric, square matrix. All the diagonal elements of this matrix are zero. In the popular version if two areas $i$ and $j$ are neighbors, then $W_{ij} = 1$, and it is zero otherwise. We now introduce additional matrices derived from $\mathbf{W}$ and describe some properties of these matrices that would be useful in exploration of our spatial models. For $i = 1, \ldots, m$, we define the $i$th row sum of $\mathbf{W}$ by $W_{i\cdot}$. We assume that $W_{i\cdot} \geq 1$ for all $i$, and we define the diagonal matrix $\mathbf{L} = \text{diag}(W_{1\cdot}, \ldots, W_{m\cdot})$. Using $\mathbf{L}$ and $\mathbf{W}$, we define two more matrices: $\widetilde{\mathbf{W}} = \mathbf{L}^{-1}\mathbf{W}$ and $\mathbf{R} = \mathbf{L} - \mathbf{W}$. The matrices $\mathbf{W}, \widetilde{\mathbf{W}}$ are non-null. Each matrix must have at least one nonzero eigenvalue. Since $\text{tr}(\mathbf{W}) = 0 = \text{tr}(\widetilde{\mathbf{W}})$, all the eigenvalues of each matrix sum to zero. Since $\mathbf{W}$ is symmetric, all its eigenvalues are real. Let $\mathbf{L}^{-1/2}$ be a diagonal matrix such that the $i$th diagonal is $W_{i\cdot}^{-1/2}$. Then all the eigenvalues of the matrix $\mathbf{L}^{-1/2}\mathbf{W}\mathbf{L}^{-1/2}$ will be real. Further, these eigenvalues are the same as the eigenvalues of $\widetilde{\mathbf{W}}$. Hence, for both $\mathbf{W}$, $\widetilde{\mathbf{W}}$, the smallest eigenvalue must be negative and the largest must be positive.

Suppose $\tilde{\lambda}_i, i = 1, \ldots, m$, are the eigenvalues of $\widetilde{\mathbf{W}}$, which are real. We can order them as $\tilde{\lambda}_m \leq \cdots \leq \cdots \leq \tilde{\lambda}_1$. Note that the elements of the matrix $\widetilde{\mathbf{W}}$ are nonnegative, diagonals are zero, and each row of the matrix sums to 1. It is a stochastic matrix. That ensures that at least one of its eigenvalues is 1, and all other eigenvalues must be between $-1$ and 1. Thus, $-1 \leq \tilde{\lambda}_m < 0 < \tilde{\lambda}_1 = 1$.

Similarly, if $\lambda_i$'s are the eigenvalues of $\mathbf{W}$, then these are finite and real. With the smallest, $\lambda_m$, and the largest $\lambda_1$, we get $-\infty < \lambda_m < 0 < \lambda_1 < \infty$.

We consider four spatially dependent random effects models with variance-covariance matrix $\sigma^2\mathbf{\Omega}(\rho)^{-1}$, defined through their associated p.d. "precision" matrices, depending on a spatial parameter $\rho$. These models are simultaneous autoregressive (SAR), conditional autoregressive (CAR), simple CAR (SCAR) and Leroux CAR (LCAR). For these models we have

$$\text{SAR: } \mathbf{\Omega}_2(\rho) = (\mathbf{I}_m - \rho\widetilde{\mathbf{W}})^{\mathrm{T}}(\mathbf{I}_m - \rho\widetilde{\mathbf{W}}), \quad \rho \in (-1, 1), \tag{4}$$

$$\text{SCAR: } \mathbf{\Omega}_3(\rho) = \mathbf{I}_m - \rho\mathbf{W}, \quad \rho \in (\lambda_m^{-1}, \lambda_1^{-1}), \tag{5}$$

$$\text{CAR: } \mathbf{\Omega}_4(\rho) = \mathbf{L} - \rho\mathbf{W}, \quad \rho \in (-1, 1), \tag{6}$$

$$\text{LCAR: } \mathbf{\Omega}_5(\rho) = \rho\mathbf{R} + (1 - \rho)\mathbf{I}_m, \quad \rho \in [0, 1). \tag{7}$$

For all the models the ranges of the parameter $\rho$ are defined above so that the $\mathbf{\Omega}$ matrices are p.d. Even though we have used the same notation $\sigma^2, \rho$ for the scale and the spatial parameters in all four models (see stage (b) of the HB model), neither they admit the same interpretations nor a combination of their values signifies equal variability and spatial strength of dependence across the models. Finally, the SAR, SCAR and LCAR models include the traditional independent Fay-Herriot linking model as a special case.

## 3. The posterior distribution of the small area mean vector

We carry out inference for $\boldsymbol{\theta}$ by conditioning on $\mathbf{S} = \mathbf{s}$ from the HB model given in Section 2. Our approach is computing-based, we will use the Monte Carlo method to generate multiple copies of sample of $\boldsymbol{\theta}$ from its posterior pdf. We use the Hamiltonian Monte Carlo (HMC) algorithm to sample the posterior distribution, and we implement this algorithm using the `RStan` software package (see Stan Development Team (2018)). The samples for $\boldsymbol{\theta}$ from its posterior distribution will be meaningful provided the posterior distribution, $\pi(\boldsymbol{\theta}|\mathbf{s})$, is proper. In the Theorem below we provide a set of sufficient conditions for the propriety of $\pi(\boldsymbol{\theta}|\mathbf{s})$.

We now describe conditions for propriety of the posterior distributions under various spatial small area models given in (4)–(7). Let $I(\cdot)$ be the indicator function taking the value 1 when its argument is true and 0 otherwise. We first provide general conditions for the posterior propriety of the proposed models.

**Theorem 1:** For all the HB spatial models given above, and equations (2), and (4)–(7), the posterior probability density functions are proper if the following conditions hold for some positive constant $N > 0$:

(a) $\int_0^\infty g(\sigma^2) I(\sigma^2 \leq N) d\sigma^2 < \infty$.

(b) $\int_0^\infty (\sigma^2)^{-(r-p)/2} g(\sigma^2) I(\sigma^2 > N) d\sigma^2 < \infty$.

If $g(\cdot)$ is a proper pdf, then (a) holds true automatically, and (b) is satisfied if $r \geq p$. We explained earlier the obvious necessity of the condition $r \geq p$ since at least $p$ summary statistics are needed to estimate $p$ components of $\boldsymbol{\beta}$ when no substantive information about $\boldsymbol{\beta}$ is available. Note that for all the spatial models we have the conditions (a) and (b) for propriety of the respective posterior distribution. Under the popular family of noninformative priors

$$\pi(\boldsymbol{\beta}, \sigma^2, \rho) \propto (\sigma^2)^{-\alpha} I(l < \rho < u), \quad \boldsymbol{\beta} \in \mathbb{R}^p, \ \sigma^2 > 0, \tag{8}$$

the posterior pdfs are proper under the following conditions.

**Corollary 1.1:** For any of the HB spatial models given in (4)–(7) and with the prior in (8), the posterior pdf is proper as long as $\alpha < 1$ and $r > p + 2 - 2\alpha$.

For the uniform prior with $\alpha = 0$ (which is used in this paper), the propriety of the posterior distributions for models (4)–(7) are guaranteed as long as $r > p + 2$. We prove the Theorem in Section 6. The Corollary follows easily from the Theorem.

## 4. An illustration to a data from the current population survey

We illustrate our method to estimation of 1989 four-person family median incomes for the U.S. forty-eight mainland states and the Washington, DC. We consider this application for two reasons. First, Chung and Datta (2022) used this application and applied the independent Fay-Herriot model and four spatial models to estimate the true median incomes, $\theta_i$'s, based on forty-nine direct estimates for these states coming from an annual supplement of the Current Population Survey (CPS). Second, a reliable set of values of these incomes are available from a large sample from the 1990 Census. Many SAE experts, for example, Ghosh *et al.* (1996) treat these values as "true values" or "gold standards" and assess accuracy of various sets of estimates against these values. The Census Bureau annually supplied accurate estimates of median incomes for states to the U.S. Department of Health and Human Service (HHS) agency that needed these estimates to implement a federal welfare program. The annual state-level estimates of these parameters from the CPS data are less reliable due to their large sampling standard deviations. To produce more reliable estimates the U.S. Census Bureau considered model-based small area estimation by using effective auxiliary data from other sources.

In our illustration for the four spatial and the independent Fay-Herriot model, we consider two types of mean functions, specified by the regression function $\mathbf{x}_i^T\boldsymbol{\beta}$. The most effective regression function involves both the covariates $x_1$ and $x_2$ that are introduced above. It has been found that $x_2$ has more predictive power in predicting $\theta_i$,s than $x_1$. Here, $x_1$ is a weaker covariate and $x_2$ is a stronger covariate. We consider two regression functions: one with both the covariates (all covariates, $k = 1$), and the other with $x_1$ (the weaker covariate, $k = 2$).

Based on use of data types, we have full data case ($Y_i$'s available for all areas, $F$) and aggregated data case (based on $S_j$'s, $A$). Within each mean function and data type, we have fitted five versions of the Fay-Herriot model, resulting in a combination of 20 models and 20 sets of predictions of the $\theta_i$'s.

Our goal is to estimate $\theta_i$, the true 1989 four-person family median income of the $i$th state, $i = 1, \ldots, 49$, excluding Alaska and Hawaii. From the 1990 CPS we get $Y_i$, the direct estimate of $\theta_i$. The Census Bureau statistician Bob Fay found out that the corresponding 1980 Census median income figure ($x_{i1}$), and an adjusted 1980 Census median income $x_{i2}$, adjusted by per capita income data from 1979 and 1989, are two powerful covariates for prediction of $\theta_i$. The CPS data also provided $D_i$, the sampling variance of $Y_i$. In our illustration we create a set of aggregated statistics $\mathbf{S}$ by grouping 49 states into 25 "super-areas", 24 groups of two states, and one lone state. In our illustration, we create required aggregated statistics by calculating $S_i = Y_{2i-1} + Y_{2i}, D_{Si} = D_{2i-1} + D_{2i}$, $i = 1, \ldots, 24$, and $S_{25} = Y_{49}$, $D_{S25} = D_{49}$. We apply five versions of the Fay-Herriot model mentioned above to this data and compare results from each of these models with the similar results presented by Chung and Datta (2022). We will also compare the five proposed models among themselves in terms of their prediction accuracy when we have only aggregated data but no data for the individual states.

## 4.1. Four-person family median income estimation with all covariates

We have twenty different settings formed by combination of five types of model variance matrices in the Fay-Herriot model, two linear regressions and two data types for the response. In our Bayesian analysis for these twenty settings we used uniform prior for the regression and variance parameters that appear in the corresponding model. For each setting we used `Rstan` to generate 24000 representative, nearly independent, Monte Carlo samples of all the parameters from the respective posterior distribution. Based on the posterior samples for the $j$th model error variance type, $k$th mean function type, and the $T$th data type, we computed Bayes estimate of $\theta_i$, denoted by $\hat{\theta}_{T,j,k,i}$. We also compute the posterior standard deviation $\sigma_{T,j,k,i}$ associated with $\hat{\theta}_{T,j,k,i}$. We also computed summary and relative frequency histograms of the spatial parameters $\rho$ for the models that have this parameter $(j = 2, 3, 4, 5)$.

We use $g_i$ as the gold standard for $\theta_i$ from the 1990 Census to empirically evaluate performance of $\hat{\theta}_{T,j,k,i}, i = 1, \ldots, m$, we compute for each set of predictions based on data type $T$, the empirical mean squared error $eMSPE_{T,j,k} = \sum_{i=1}^{49}(\hat{\theta}_{T,j,k,i} - g_i)^2/49$ for $j = 1, \ldots, 5$, $k = 1$ ($k = 2$ is considered in Subsection 4.2). These values for $k = 1$ are presented in the second column of Table 1 (for aggregated data), and Table 2 (for full data). We also computed average posterior standard deviations $\bar{\sigma}_{T,j,k} = \sum_{i=1}^{49} \sigma_{T,j,k,i}/49$, $T = A, F$. These values are given in the sixth column of the tables we created. Additionally, within each model, using appropriate posterior quantiles, we constructed 95% central credible interval for each $\theta_i$. Using these intervals and the gold standard values we calculated empirical coverage rates of these intervals by computing the fraction of the 49 intervals that included the $g_i$ values (presented in the fifth column). We also presented in the fourth column average length of these intervals.

In the absence of a direct estimate for a small area, a synthetic estimate based on the estimated regression function and covariates from that area is a reasonable alternative. In our case where we only have access to aggregated statistics based on data from multiple areas, we do not typically have direct estimates for any areas. In this scenario, synthetic estimates for all the areas may appear to be appealing. A synthetic estimate of $\theta_i$ for a typical model is $\hat{\theta}_{syn,T,j,k,i} = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{T,j,k}$, where $T = A, F$, $j = 1, \ldots, 5$, and $k = 1, 2$. Here, $\hat{\boldsymbol{\beta}}_{T,j,k}$ is a Bayes estimator of $\boldsymbol{\beta}$ under the $T, j, k$th setting. We note that all these results corresponding to $T = F$ for full data were obtained by Chung and Datta (2022).

At the early stage of small area estimation— pre-dating use of random effects or hierarchical models— practitioners used synthetic estimates. For the synthetic estimates we computed empirical MSPE by averaging the squares differences of the estimates from the gold standard values, $g_i$. We present these measures, represented as “syn MSPE” in the eighth column of the tables. Under any Bayesian model, the accuracy of corresponding synthetic estimates are evaluated by the posterior root mean squared error of each estimate. Averages of these values are reported as synthetic average root posterior mean squared error (syn ARPME) in the last column. Synthetic estimators usually tend to be biased, particularly if the regression function is an inadequate fit for the $\theta_i$’s, but they have smaller variances. In the case of poor model fit, the bias term of the synthetic estimator usually gets elevated, and the variance may fail to compensate for the larger bias, resulting in a large posterior MSE of synthetic estimator.

Both Table 1 and Table 2 show that the synthetic estimates for each model have smaller empirical MSPE than for their Bayesian counterparts. This is rather unusual unless the covariates are very effective, which appears to be the case here. However, the Bayes estimates have average posterior standard deviations in column 6 which are smaller than the average root posterior means squared error of their synthetic estimate counterparts,

**Table 1: Aggregated data with all covariates**

| | eMSPE | eMSPE-PI | AL | CP | APSD | APSD-PI | syn MSPE | syn ARPME |
|---|---|---|---|---|---|---|---|---|
| FH | 4.02 | - | 14.30 | 0.9509 | 3.15 | - | 2.10 | 3.47 |
| SAR | 3.87 | 3.70 % | 14.21 | 0.9507 | 3.15 | -0.07 % | 2.12 | 3.70 |
| SCAR | 4.10 | -2.03 % | 14.26 | 0.9510 | 3.12 | 0.88 % | 2.10 | 3.45 |
| CAR | 4.41 | -9.72 % | 14.62 | 0.9471 | 3.19 | -1.37 % | 2.13 | 3.56 |
| LCAR | 3.37 | 16.09 % | 13.74 | 0.9489 | 3.09 | 1.82 % | 2.42 | 3.98 |

empirical Mean squared prediction error (eMSPE), average posterior standard deviation (APSD), and respective percentage improvements (PI) of spatial models over the independent FH model for Bayes predictor of $\boldsymbol{\theta}$ and synthetic estimator $\mathbf{X}^T\hat{\boldsymbol{\beta}}$, and also average length (AL), coverage probability (CP).

**Table 2: Full data from forty-nine states with all covariates**

| | eMSPE | eMSPE-PI | AL | CP | APSD | APSD-PI | syn MSPE | syn ARPME |
|---|---|---|---|---|---|---|---|---|
| FH | 2.88 | - | 7.63 | 0.9592 | 1.93 | - | 1.86 | 2.57 |
| SAR | 2.61 | 9.55 % | 7.58 | 0.9592 | 1.94 | 0.34% | 2.00 | 2.74 |
| SCAR | 3.03 | -5.14 % | 7.66 | 0.9592 | 1.95 | -0.91% | 1.86 | 2.57 |
| CAR | 2.64 | 8.47 % | 7.48 | 0.9592 | 1.91 | 1.24% | 1.98 | 2.63 |
| LCAR | 2.47 | 14.50 % | 7.31 | 0.9592 | 1.85 | 4.19% | 2.37 | 3.02 |

**Table 3: Posterior mean/mode (standard deviation) of $\rho$ for various models ans data types.**

| Data type | Covariate included | SAR | SCAR | CAR | LCAR |
|---|---|---|---|---|---|
| Aggregated | $x_1$,$x_2$ | -0.10 /-0.22(0.44) | -0.09 /-0.09(0.14) | -0.20 /-0.25(0.59) | 0.47 / 0.22(0.28) |
| data | $x_1$ | 0.43 / 0.68(0.39) | 0.02 / 0.17(0.13) | 0.41 / 0.95(0.54) | 0.71 / 0.98(0.24) |
| Full | $x_1$, $x_2$ | 0.10 / 0.38(0.48) | -0.06 / 0.11(0.14) | 0.21 / 0.98(0.55) | 0.57 / 0.78(0.27) |
| data | $x_1$ | 0.76 / 0.83(0.14) | 0.14 / 0.18(0.04) | 0.93 / 0.98(0.09) | 0.85 / 0.98(0.13) |

reported in column 9. Actually, by being the Bayes estimates, they will have smaller posterior means squared error. Since few applications provide any set of gold standards to compare estimates against, it is important to compare various estimates in terms of their variability or concentration.

Even in the presence of powerful predictors Table 1 showed that only the LCAR model emerged to be the best among the five sets (including the independent Fay-Herriot) in terms of eMSPE, AL and APSD. Other spatial models turned out to be less competitive or inferior to the independent FH model. If we turn to Table 2, even when we have direct estimates from all 49 states, the LCAR model still turned out to be the best of the five models in terms of the same measures. Among the other models, the SAR and CAR models also improved over the independent FH model.

To assess efficiency loss due to data compression through aggregation, we compare the results of Table 1 with those of Table 2. Across models the percentages increase in eMSPE's for the aggregated data over their counterparts for the full data, respectively, are 40, 48, 35, 67 and 36; the two smaller of the increases are for the SCAR and the LCAR models. We note that in both the tables that all the CP's are practically at the target 95%. When we compare the average length of the credible intervals, the percentage increases for the aggregated data across models over their counterparts for the full data, respectively, are 87, 87, 86, 95 and 88; this time, the smaller of the increases are for the models other than the

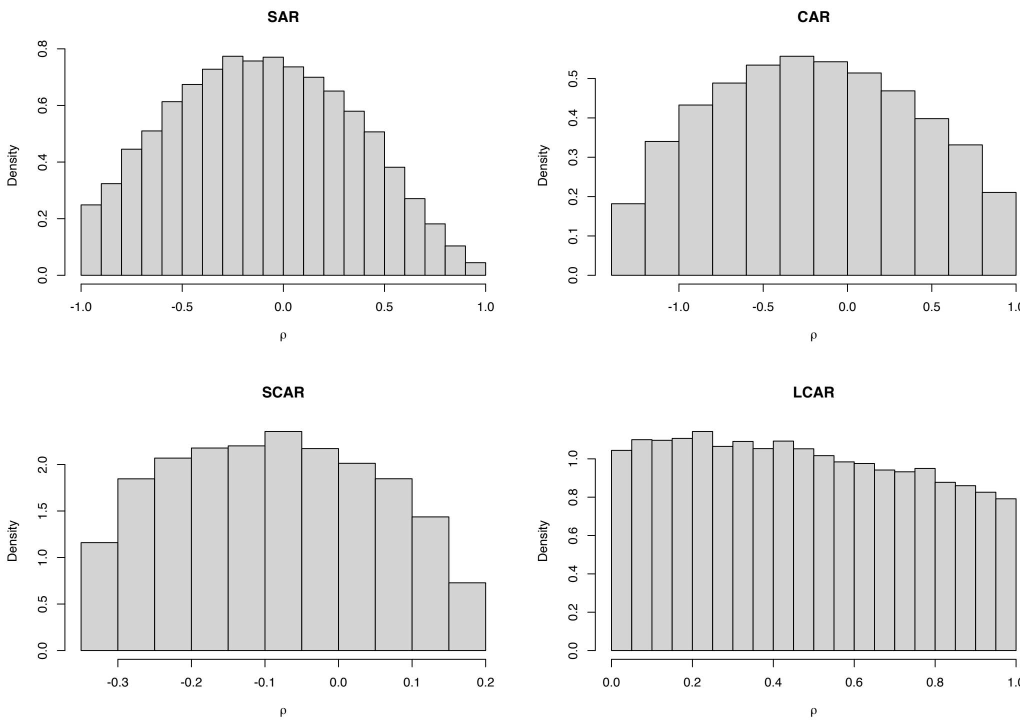


**Figure 1: Posterior relative frequency histogram of $\rho$ for aggregated data with all covariates**

CAR. Among these spatial models, the LCAR model produced the smallest eMSPE and the AL values. Finally, the rows corresponding to $x_1, x_2$ in Table 3 show that for both data types the spatial parameter only for the LCAR model appears to be the one that is more likely to be non-zero. The same conclusion emerges about the spatial models from the posterior relative frequency histograms of $\rho$ presented in Figures 1 and 2.

## 4.2. Four-person family median income estimation with the weaker covariate

Research shows that spatial random effects models tend to have better predictive power than a corresponding independent Fay-Herriot model when no effective covariates are available, see, for example, Chung and Datta (2022) and Vogt *et al.* (2023). For the median income estimation problem based on direct estimates from all 49 mainland states Chung and Datta (2022) showed that in the absence of any covariates some of the spatial models do better than the independent Fay-Herriot model. Usually, the SAR or the LCAR model provides the best prediction. In this section, we plan to investigate based on modeling of only aggregated statistics if any of the spatial models would be better than the independent Fay-Herriot model.

We note from Table 4 and Table 5 that for both aggregated data and full data cases in the absence of powerful predictors of $\theta_i$'s, all the spatial models provide better predictions than the independent FH model when compared in terms of eMSPE, AL and APSD. In this setting with low quality predictor, all synthetic estimators of $\theta_i$'s have bigger average MSPE's than their Bayesian counterparts (see columns 2 and 8). In this case, the LCAR is the best

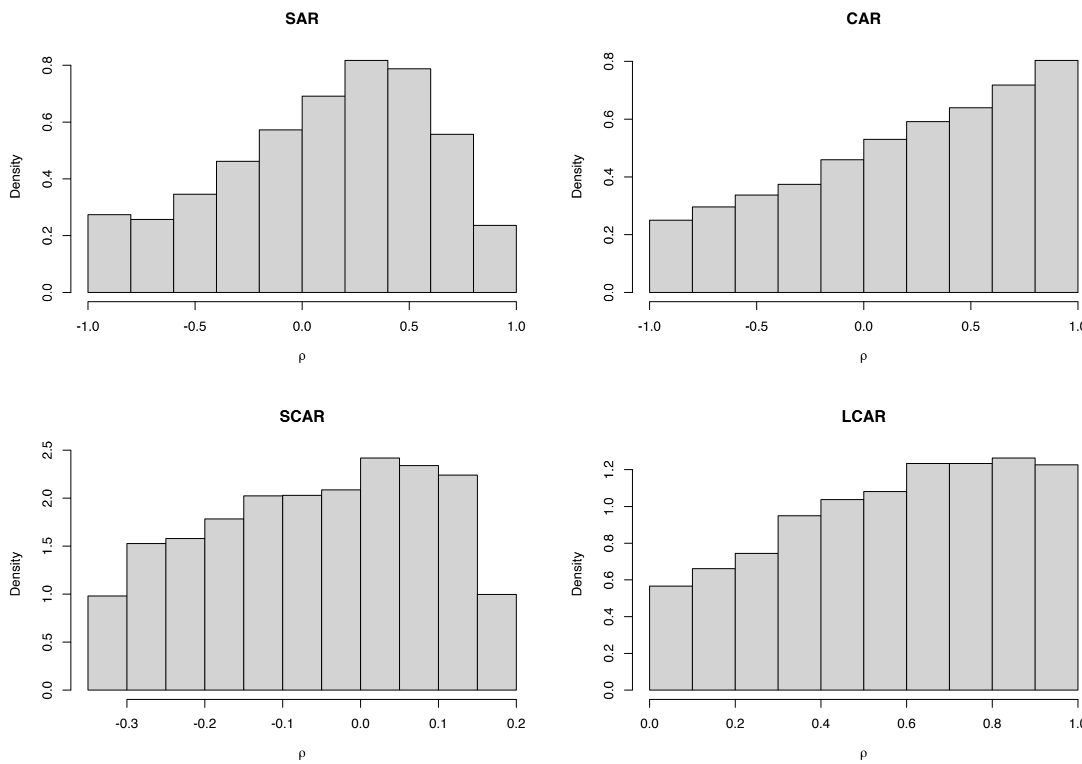


**Figure 2: Posterior relative frequency histogram of $\rho$ with all covariates and 49 states data**

spatial model across both data settings. Results from Table 4 and those corresponding to the $x_1$ rows in Table 3 show that the LCAR is the best of the spatial models and the spatial parameter of this model appeared most likely to be different from zero. Moreover, from the two relative frequency histograms of $\rho$ in Figure 3 and Figure 4 it is obvious that for the LCAR model 95% highest posterior density credible intervals of $\rho$ will not include the zero value. For the full data case we also note from the last row of Table 3 and Figure 4 that the respective spatial parameter in all the spatial models appears very likely to be different from zero.

Before concluding Subsection 4.1 and Subsection 4.2, from a quick look at the APSD's for the independent FH model reported in the first row of Table 1, Table 2, Table 4 and Table 5, we found out that the average of the posterior variances of the $\theta_i$'s under the aggregated data setting is nearly three times that quantity under the full data case. This substantial increase in the posterior variances of $\theta_i$ under the aggregated data setting compared to the full data setting can reasonably be explained by the expression of the posterior variance of $\theta_i$ in equation (1) under the assumption of known model parameters.

**Table 4: Aggregated data with a weaker covariate**

| | eMSPE | eMSPE-PI | AL | CP | APSD | APSD-PI | syn MSPE | syn ARPME |
|---|---|---|---|---|---|---|---|---|
| FH | 11.78 | - | 20.47 | 0.9513 | 4.02 | - | 14.44 | 4.77 |
| SAR | 6.76 | 42.64% | 17.53 | 0.9507 | 3.77 | 6.09% | 14.29 | 6.63 |
| SCAR | 10.60 | 10.06% | 19.88 | 0.9513 | 3.99 | 0.75% | 14.73 | 4.84 |
| CAR | 8.52 | 27.67% | 18.58 | 0.9476 | 3.85 | 4.08% | 14.17 | 5.06 |
| LCAR | 6.03 | 48.80% | 16.39 | 0.9480 | 3.48 | 13.31% | 14.20 | 6.55 |

**Table 5: Full data from forty-nine states with a weaker covariate**

| | eMSPE | eMSPE-PI | AL | CP | APSD | APSD-PI | syn MSPE | syn ARPME |
|---|---|---|---|---|---|---|---|---|
| FH | 7.27 | - | 9.09 | 0.9388 | 2.31 | - | 14.45 | 4.04 |
| SAR | 4.34 | 40.22% | 7.73 | 0.9796 | 1.98 | 14.25% | 14.61 | 7.65 |
| SCAR | 5.62 | 22.62% | 8.75 | 0.9592 | 2.22 | 3.52% | 15.36 | 4.27 |
| CAR | 4.62 | 36.35% | 7.84 | 0.9388 | 2.01 | 12.97% | 14.70 | 5.06 |
| LCAR | 4.54 | 37.51% | 7.77 | 0.9592 | 1.97 | 14.36% | 14.67 | 6.17 |

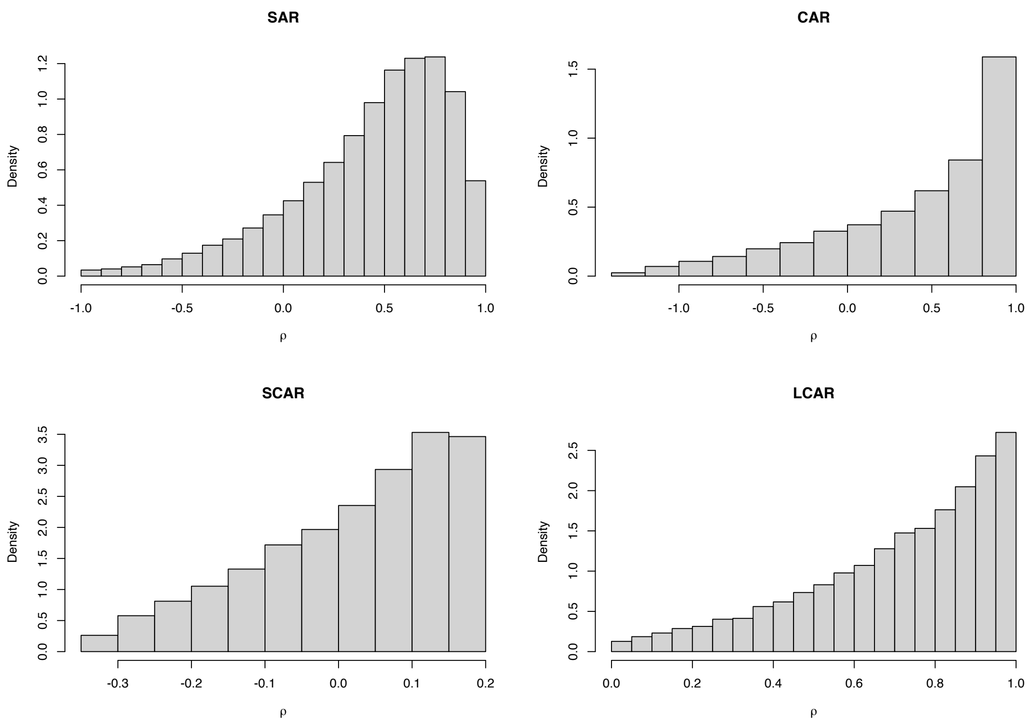


**Figure 3: Posterior relative frequency histogram of $\rho$ for aggregated data and a weaker covariate**

## 5. Importance of the study

This study addresses an important problem in area-level small area estimation when most or all of the small areas do not have a direct estimates for $\theta_i$'s. Such data can not be had due to not having a survey that collects data from the individual areas. Due to administrative or budgetary considerations, a survey may do stratified sampling where each stratum is formed by merging multiple targeted small areas. If the goal is to estimating

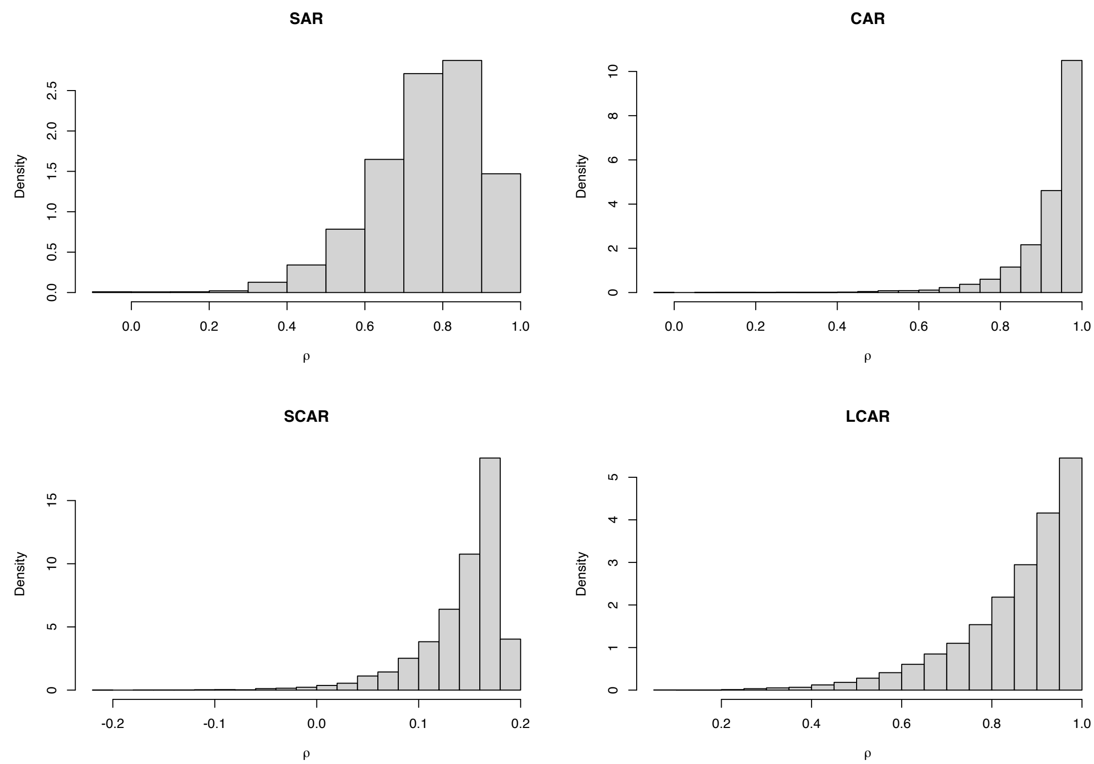


**Figure 4: Posterior relative frequency histogram of $\rho$ with a weaker covariate and 49 states data**

total agricultural productions or total employments for the strata, our study shows that the stratified means can be leveraged to reliably estimate the means for the original small areas by integrating the strata level means of a response variable with area-level data from covariates that have good predictive power to predict the small area means for the response.

For the setup we are considering here, the success of a generalization of the Fay-Herriot model depends on the availability of effective predictor variables for the response variable. In the absence of effective covariates, from the studies by Chung and Datta (2022) and Vogt *et al.* (2023), it is known that various spatial alternatives to the independent Fay-Herriot model produce significantly better predictions by accounting for the spatial variation of the small area means. Even when no substantial spatial variation exists among the means, the spatial models make marginally better predictions than the independent FH model without sacrificing model fit. We demonstrated the usefulness and the strength of our proposed method by applying this to an application that has been important to both the HHS Department and the Census Bureau of the United States.

## 6. Proof of propriety of the posterior pdfs

We know that our vector of aggregated statistics $\mathbf{S}$ is $r \times 1$ with $r \geq p$. We assume that

$$\mathbf{S}|\boldsymbol{\theta} \sim N(\mathbf{C}\boldsymbol{\theta}, \mathbf{D}_S), \tag{9}$$

where $\mathbf{C}$ is a known $r \times m$ matrix of rank $r$, $\boldsymbol{\theta}$ is an $m \times 1$ vector, and $\mathbf{D}_S$ is a known positive definite (p.d.) matrix of rank $r$.

Suppose the largest eigenvalue of $\mathbf{D}_S$ is $\delta$, which is finite and positive. Let $N(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the multivariate normal pdf with mean $\boldsymbol{\mu}$ and p.d. variance-covariance matrix $\boldsymbol{\Sigma}$ at $\mathbf{x}$. Since $\delta^{-1} > 0$ is the smallest eigenvalue of $\mathbf{D}_S^{-1}$, from the property of the minimum eigenvalue we get that

$$\begin{aligned}
&(\mathbf{s} - \mathbf{C}\boldsymbol{\theta})^T \mathbf{D}_S^{-1} (\mathbf{s} - \mathbf{C}\boldsymbol{\theta}) \geq \delta^{-1} (\mathbf{s} - \mathbf{C}\boldsymbol{\theta})^T (\mathbf{s} - \mathbf{C}\boldsymbol{\theta}) \\
&\Rightarrow N(\mathbf{s}|\mathbf{C}\boldsymbol{\theta}, \mathbf{D}_S) \leq K N(\mathbf{s}|\mathbf{C}\boldsymbol{\theta}, \delta \mathbf{I}_r),
\end{aligned} \tag{10}$$

where $K > 0$ is a generic known suitable constant, dependent on $\mathbf{D}_S$ but free from $\mathbf{s}$ or $\boldsymbol{\theta}$.

We can select a matrix $\mathbf{F}((m-r) \times m)$, dependent on $\mathbf{C}$ but known so that the $m \times m$ matrix $\mathbf{M} = (\mathbf{C}^T, \mathbf{F}^T)^T$ is non-singular. This implies that the rank of $\mathbf{F}$ is $m - r$. For an $(m - r) \times 1$ vector $\mathbf{h}_2$ note that

$$\int_{R^{m-r}} N(\mathbf{h}_2 | \mathbf{F}\boldsymbol{\theta}, \delta \mathbf{I}_{m-r}) d\mathbf{h}_2 = K < \infty, \tag{11}$$

where $K$ is a generic and positive constant. By (10)-(11) we get that

$$\begin{aligned}
N(\mathbf{s}|\mathbf{C}\boldsymbol{\theta}, \mathbf{D}_S) &\leq K N(\mathbf{s}|\mathbf{C}\boldsymbol{\theta}, \delta \mathbf{I}_r) \int_{R^{m-r}} N(\mathbf{h}_2 | \mathbf{F}\boldsymbol{\theta}, \delta \mathbf{I}_{m-r}) d\mathbf{h}_2 \\
&= K \int_{R^{m-r}} N(\mathbf{h} | \mathbf{M}\boldsymbol{\theta}, \delta \mathbf{I}_m) d\mathbf{h}_2,
\end{aligned} \tag{12}$$

where $\mathbf{h} = (\mathbf{s}^T, \mathbf{h}_2^T)^T$ is an $m \times 1$ vector.

Let $\mathbf{M}^{-1} = \mathbf{B}$. Let $k$ be the smallest eigenvalue of the p.d. matrix $\mathbf{M}^T \mathbf{M}$. Using $\mathbf{h} - \mathbf{M}\boldsymbol{\theta} = \mathbf{M}(\mathbf{B}\mathbf{h} - \boldsymbol{\theta})$ we get

$$\begin{aligned}
(\mathbf{h} - \mathbf{M}\boldsymbol{\theta})^T (\mathbf{h} - \mathbf{M}\boldsymbol{\theta}) &= (\mathbf{B}\mathbf{h} - \boldsymbol{\theta})^T \mathbf{M}^T \mathbf{M} (\mathbf{B}\mathbf{h} - \boldsymbol{\theta}) \\
&\geq k (\mathbf{B}\mathbf{h} - \boldsymbol{\theta})^T (\mathbf{B}\mathbf{h} - \boldsymbol{\theta}).
\end{aligned}$$

From the above, using $k > 0$, we get that

$$N(\mathbf{h} | \mathbf{M}\boldsymbol{\theta}, \delta \mathbf{I}_m) \leq K N(\mathbf{B}\mathbf{h} | \boldsymbol{\theta}, \delta k^{-1} \mathbf{I}_m). \tag{13}$$

By (12)-(13), writing $\delta k^{-1} = \delta^*$, we get

$$\begin{aligned}
N(\mathbf{s}|\mathbf{C}\boldsymbol{\theta}, \mathbf{D}_S) &\leq K \int_{R^{m-r}} N(\mathbf{h} | \mathbf{M}\boldsymbol{\theta}, \delta \mathbf{I}_m) d\mathbf{h}_2 \\
&\leq K \int_{R^{m-r}} N(\mathbf{B}\mathbf{h} | \boldsymbol{\theta}, \delta^* \mathbf{I}_m) d\mathbf{h}_2.
\end{aligned} \tag{14}$$

Recall that for the class of spatial models, the *linking model* is given by

$$\boldsymbol{\theta} | \boldsymbol{\beta}, \sigma^2, \rho \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{\Omega}^{-1}), \tag{15}$$

where $\mathbf{X}$ is a known $m \times p$ matrix of covariates of rank $p$, and $\boldsymbol{\Omega}$ is an $m \times m$ p.d. matrix that depends on a parameter $\rho$ which varies on a known finite interval.

Let $f_{\mathbf{S}}(\mathbf{s}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Omega}) = \int_{R^m} N(\mathbf{s}|\mathbf{C}\boldsymbol{\theta}, \mathbf{D}_S) N(\boldsymbol{\theta}|\mathbf{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{\Omega}^{-1}) d\boldsymbol{\theta}$ be the pdf of $\mathbf{S}$ given $\boldsymbol{\beta}, \sigma^2, \rho$. Then from (14) we get

$$\begin{aligned} f_{\mathbf{S}}(\mathbf{s}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Omega}) &\leq K \int_{R^m} \int_{R^{m-r}} N(\mathbf{B}\mathbf{h}|\boldsymbol{\theta}, \delta^* \mathbf{I}_m) N(\boldsymbol{\theta}|\mathbf{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{\Omega}^{-1}) d\mathbf{h}_2 d\boldsymbol{\theta} \\ &= K \int_{R^{m-r}} N(\mathbf{B}\mathbf{h}|\mathbf{X}\boldsymbol{\beta}, \delta^*\mathbf{I}_m + \sigma^2\boldsymbol{\Omega}^{-1}) d\mathbf{h}_2 \\ &= K \int_{R^{m-r}} N(\mathbf{B}\mathbf{h} - \mathbf{X}\boldsymbol{\beta}|\mathbf{0}, \delta^*\mathbf{I}_m + \sigma^2\boldsymbol{\Omega}^{-1}) d\mathbf{h}_2. \end{aligned} \tag{16}$$

Now, $\mathbf{B}\mathbf{h} - \mathbf{X}\boldsymbol{\beta} = \mathbf{B}(\mathbf{h} - \mathbf{M}\mathbf{X}\boldsymbol{\beta})$. Let $\mathbf{d} = (\mathbf{h}_1^T, \mathbf{0}^T)^T$ and

$$\mathbf{G} = \begin{bmatrix} \mathbf{C}\mathbf{X} & \mathbf{0} \\ \mathbf{F}\mathbf{X} & -\mathbf{I}_{m-r} \end{bmatrix}.$$

Then, we have

$$\mathbf{h} - \mathbf{M}\mathbf{X}\boldsymbol{\beta} = \mathbf{d} - \mathbf{G}\boldsymbol{\phi}, \tag{17}$$

where $\boldsymbol{\phi} = (\boldsymbol{\beta}^T, \mathbf{h}_2^T)^T$ is a $(p + m - r) \times 1$ vector. Now, define submatrices $\mathbf{G}_1$ and $\mathbf{G}_2$ to introduce a column partition of the matrix $\mathbf{G}$, where $\mathbf{G}_1$ is given by the first $p$ columns of $\mathbf{G}$, and $\mathbf{G}_2$ is given by the last $m - r$ columns of $\mathbf{G}$. Columns of $\mathbf{G}_2$ are linearly independent. So $rank(\mathbf{G}_2) = m - r$. Also, since we require $\mathbf{C}\mathbf{X} \neq \mathbf{0}$, the columns of $\mathbf{G}_1$ cannot be linearly expressed by the columns of $\mathbf{G}_2$. However, $\mathbf{G}_1 = \mathbf{M}\mathbf{X}$ implies $rank(\mathbf{G}_1) = rank(\mathbf{X}) = p$. Hence, $rank(\mathbf{G}) = rank(\mathbf{G}_1) + rank(\mathbf{G}_2) = p + m - r$.

Let $\mathbf{B}\mathbf{d} = \mathbf{d}_*,\ \mathbf{B}\mathbf{G} = \mathbf{G}_*$. Then, by (17)

$$\mathbf{B}\mathbf{h} - \mathbf{X}\boldsymbol{\beta} = \mathbf{B}(\mathbf{h} - \mathbf{M}\mathbf{X}\boldsymbol{\beta}) = \mathbf{B}(\mathbf{d} - \mathbf{G}\boldsymbol{\phi}) = \mathbf{d}_* - \mathbf{G}_*\boldsymbol{\phi}. \tag{18}$$

Using (16) and (18) we get,

$$f_{\mathbf{S}}(\mathbf{s}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Omega}) \leq K \int_{R^{m-r}} N(\mathbf{d}_* - \mathbf{G}_*\boldsymbol{\phi}|\mathbf{0}, \delta^*\mathbf{I}_m + \sigma^2\boldsymbol{\Omega}^{-1}) d\mathbf{h}_2. \tag{19}$$

Further,

$$\begin{aligned} & N(\mathbf{d}_* - \mathbf{G}_*\boldsymbol{\phi}|\mathbf{0}, \delta^*\mathbf{I}_m + \sigma^2\boldsymbol{\Omega}^{-1}) = K|\delta^*\mathbf{I}_m + \sigma^2\boldsymbol{\Omega}^{-1}|^{-1/2} \\ \times \quad & \exp[-\frac{(\mathbf{d}_* - \mathbf{G}_*\boldsymbol{\phi})^T(\delta^*\mathbf{I}_m + \sigma^2\boldsymbol{\Omega}^{-1})^{-1}(\mathbf{d}_* - \mathbf{G}_*\boldsymbol{\phi})}{2}]. \end{aligned} \tag{20}$$

We consider four spatially dependent random effects models with variance-covariance matrix $\sigma^2\boldsymbol{\Omega}(\rho)^{-1}$, defined through their associated p.d. "precision" matrices, depending on a spatial parameter $\rho$: for all the models the parameter $\rho$ is defined on an appropriate *finite* interval so that the $\boldsymbol{\Omega}$ matrices are p.d.

To continue our propriety proof, for convenience of notation, we denote $\boldsymbol{\Omega}_k(\rho)$ by $\boldsymbol{\Omega}_k$, for $k = 1, \ldots, 5$. Here, $\boldsymbol{\Omega}_1(\rho) = \mathbf{I}_m$ is for the independent Fay-Herriot model. In the next two subsections we present detail arguments establishing the propriety of the posterior pdfs for the SCAR and the SAR models. Under the same conditions, similar arguments can be made for proving the propriety of the posterior pdfs for the CAR and the LCAR models; see also Appendices A.3 and A.4 of Chung and Datta (2022). Result for the independent model follows from the SAR or the SCAR model with $\rho = 0$.

Finally, suppose $\mathbf{C} = \begin{bmatrix}\mathbf{0} & \mathbf{I}_r\end{bmatrix}$, an $r \times m$ matrix. This is a special case of the general setup considered in this paper. This special case was considered in Chung and Datta (2022).

## 6.1. The propriety for the SCAR model

For the eigenvalues $\lambda_i$'s of $\mathbf{W}$, let $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_m)$ and $\mathbf{P}_W$ be an orthogonal matrix such that $\mathbf{W} = \mathbf{P}_W \boldsymbol{\Lambda} \mathbf{P}_W^T$. For the SCAR model, $\boldsymbol{\Omega} = \boldsymbol{\Omega}_3$ and

$$\boldsymbol{\Omega}_3^{-1} = \mathbf{P}_W [\mathbf{I}_m - \rho \boldsymbol{\Lambda}]^{-1} \mathbf{P}_W^T. \tag{21}$$

From this we get

$$(\delta^* \mathbf{I}_m + \sigma^2 \boldsymbol{\Omega}_3^{-1})^{-1} = \mathbf{P}_W [\delta^* \mathbf{I}_m + \sigma^2 \{\mathbf{I}_m - \rho \boldsymbol{\Lambda}\}^{-1}]^{-1} \mathbf{P}_W^T,$$

which implies that

$$\begin{aligned}
& (\mathbf{d}_* - \mathbf{G}_* \boldsymbol{\phi})^T (\delta^* \mathbf{I}_m + \sigma^2 \boldsymbol{\Omega}_3^{-1})^{-1} (\mathbf{d}_* - \mathbf{G}_* \boldsymbol{\phi}) \\
= \; & (\mathbf{d}_{**} - \mathbf{G}_{**} \boldsymbol{\phi})^T [\delta^* \mathbf{I}_m + \sigma^2 \{\mathbf{I}_m - \rho \boldsymbol{\Lambda}\}^{-1}]^{-1} (\mathbf{d}_{**} - \mathbf{G}_{**} \boldsymbol{\phi}) \\
= \; & \sum_{i=1}^{m} \frac{\{d_{**i} - \mathbf{g}_{**i}^T \boldsymbol{\phi}\}^2}{\delta^* + \sigma^2 (1 - \rho \lambda_i)^{-1}} 
\end{aligned} \tag{22}$$

where $\mathbf{d}_{**} = \mathbf{P}_W^T \mathbf{d}_*$, $\mathbf{G}_{**} = \mathbf{P}_W^T \mathbf{G}_*$, $d_{**i}$ is the $i$th element of $\mathbf{d}_{**}$ and $\mathbf{g}_{**i}^T$ is the $i$th row of $\mathbf{G}_{**}$.

Clearly,

$$rank(\mathbf{G}_{**}) = rank(\mathbf{G}_*) = rank(\mathbf{G}) = p + m - r = q \text{ (say)}.$$

We can select $q$ linearly independent rows of $\mathbf{G}_{**}$. By rearrangement of those rows we can assume that the first $q$ rows of $\mathbf{G}_{**}$ could be taken as linearly independent. Then from (22) we get

$$(\mathbf{d}_* - \mathbf{G}_* \boldsymbol{\phi})^T (\delta^* \mathbf{I}_m + \sigma^2 \boldsymbol{\Omega}_3^{-1})^{-1} (\mathbf{d}_* - \mathbf{G}_* \boldsymbol{\phi}) \geq \sum_{i=1}^{q} \frac{\{d_{**i} - \mathbf{g}_{**i}^T \boldsymbol{\phi}\}^2}{\delta^* + \sigma^2 (1 - \rho \lambda_i)^{-1}}.$$

Using this inequality in equations (19)-(20) we get

$$\begin{aligned}
&\int_{R^p} f_{\mathbf{S}}(\mathbf{s}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\Omega})\pi(\boldsymbol{\beta})d\boldsymbol{\beta} \leq K|\delta^*\mathbf{I}_m+\sigma^2\boldsymbol{\Omega}^{-1}|^{-1/2} \\
\times &\int\int \pi(\boldsymbol{\beta})\exp[-\frac{(\mathbf{d}_*-\mathbf{G}_*\boldsymbol{\phi})^T(\delta^*\mathbf{I}_m+\sigma^2\boldsymbol{\Omega}_3^{-1})^{-1}(\mathbf{d}_*-\mathbf{G}_*\boldsymbol{\phi})}{2}]d\mathbf{h}_2 d\boldsymbol{\beta} \\
\leq &\; K\prod_{i=1}^m\{\delta^*+\sigma^2(1-\rho\lambda_i)^{-1}\}^{-1/2} \\
&\times\int\int \pi(\boldsymbol{\beta})\exp[-\frac{1}{2}\sum_{i=1}^q\frac{\{d_{**i}-\mathbf{g}_{**i}^T\boldsymbol{\phi}\}^2}{\delta^*+\sigma^2(1-\rho\lambda_i)^{-1}}]d\boldsymbol{\beta}d\mathbf{h}_2 \\
\leq &\; K\prod_{i=1}^m\{\delta^*+\sigma^2(1-\rho\lambda_i)^{-1}\}^{-1/2} \\
\times &\int_{R^q}\exp[-\frac{1}{2}\sum_{i=1}^q\frac{\{d_{**i}-\mathbf{g}_{**i}^T\boldsymbol{\phi}\}^2}{\delta^*+\sigma^2(1-\rho\lambda_i)^{-1}}]d\boldsymbol{\phi} \\
= &\; K\prod_{i=q+1}^m\{\delta^*+\sigma^2(1-\rho\lambda_i)^{-1}\}^{-1/2}, \qquad (23)
\end{aligned}$$

where we assumed that $\pi(\boldsymbol{\beta})$ is bounded above, which is satisfied by a uniform prior on $R^p$.

Now, we notice that for any positive constant $N$

$$\begin{aligned}
\delta^*+\sigma^2(1-\rho\lambda_i)^{-1} &\geq \delta^* I(\sigma^2\leq N)+\sigma^2(1-\rho\lambda_i)^{-1}I(\sigma^2>N) \\
\Rightarrow \{\delta^*+\sigma^2(1-\rho\lambda_i)^{-1}\}^{-1/2} &\leq KI(\sigma^2\leq N)+(\sigma^2)^{-1/2}(1-\rho\lambda_i)^{1/2}I(\sigma^2>N). \qquad (24)
\end{aligned}$$

Since $\lambda_i$'s are finite and $\rho$ is integrated over a finite interval, it follows that $1-\rho\lambda_i$ is a finite positive quantity. Then, using $q=m-r+p$,

$$\begin{aligned}
\prod_{i=q+1}^m\{\delta^*+\sigma^2(1-\rho\lambda_i)^{-1}\}^{-1/2} &\leq K[I(\sigma^2\leq N)+(\sigma^2)^{-(m-q)/2}I(\sigma^2>N)] \\
&\leq K[I(\sigma^2\leq N)+(\sigma^2)^{-(r-p)/2}I(\sigma^2>N)]. \qquad (25)
\end{aligned}$$

Now using $\pi(\sigma^2,\rho)=g(\sigma^2)h(\rho)$ and that $h(\rho)$ is a pdf, we get

$$\begin{aligned}
&\int_0^\infty\int_l^u\prod_{i=q+1}^m\{\delta^*+\sigma^2(1-\rho\lambda_i)^{-1}\}^{-1/2}g(\sigma^2)h(\rho)d\rho d\sigma^2 \\
\leq &\; K\int_0^N g(\sigma^2)d\sigma^2+K\int_N^\infty g(\sigma^2)(\sigma^2)^{-(r-p)/2}d\sigma^2<\infty, \qquad (26)
\end{aligned}$$

by sufficient conditions (a) and (b) in Theorem 1.

In particular, if $g(\sigma^2)=(\sigma^2)^{-\alpha}$, $1-\alpha>0$ ensures (a), and $(r-p)/2+\alpha>1$ ensures (b). Equivalently, we need $\alpha<1$ and $r>p+2-2\alpha$.

### 6.2. The propriety of the posterior pdf under the SAR model

We now consider the SAR model. For this model

$$\mathbf{\Omega}_2(\rho) = (\mathbf{I}_m - \rho\widetilde{\mathbf{W}})^{\mathrm{T}}(\mathbf{I}_m - \rho\widetilde{\mathbf{W}}),\ -1 < \rho < 1.$$

Note that $\mathrm{tr}[\mathbf{\Omega}_2(\rho)] = m + \rho^2 \sum\sum \tilde{w}_{ij}^2 \leq 2m$. Let $\mathbf{W}_* = \mathbf{L}^{-1/2}\mathbf{W}\mathbf{L}^{-1/2}$. Again,

$$\mathbf{\Omega}_2(\rho) = \mathbf{L}^{1/2}(\mathbf{I} - \rho\mathbf{W}_*)\mathbf{L}^{-1}(\mathbf{I} - \rho\mathbf{W}_*)\mathbf{L}^{1/2}. \tag{27}$$

Let $\nu_1 \geq \cdots \geq \nu_m$ be the eigenvalues of $\mathbf{W}_*$. From our discussions in Subsection 2.1 all $\nu_i$'s are real. Moreover, $\nu_1 = 1$ and $|\nu_i| \leq 1$. Since $1 - \rho\nu_i$ are the eigenvalues of $\mathbf{I} - \rho\mathbf{W}_*$, for $-1 < \rho < 1$, these eigenvalues are all positive. Hence the matrix is p.d. Actually, for all $i$, $0 < 1 - \rho\nu_i < 2$.

Let $l_{(1)} = \min W_{i\cdot}$ and $l_{(m)} = \max W_{i\cdot}$. Note that $1 \leq l_{(1)} \leq l_{(m)} < m$. Define the matrix

$$\mathbf{H} = \delta^*\mathbf{L} + \sigma^2(\mathbf{I} - \rho\mathbf{W}_*)^{-1}\mathbf{L}(\mathbf{I} - \rho\mathbf{W}_*)^{-1}.$$

Since the matrix $\mathbf{L} - \mathbf{I}$ is n.n.d., the matrix $\mathbf{H} - \{\delta^*\mathbf{I} + \sigma^2(\mathbf{I} - \rho\mathbf{W}_*)^{-2}\}$ is n.n.d. It easily follows that

$$|\mathbf{H}| \geq |\delta^*\mathbf{I} + \sigma^2(\mathbf{I} - \rho\mathbf{W}_*)^{-2}| = \prod_{i=1}^{m}\{\delta^* + \sigma^2(1 - \rho\nu_i)^{-2}\}.$$

Let $\mathbf{\Sigma}_2 = \delta^*\mathbf{I} + \sigma^2\mathbf{\Omega}_2^{-1}$. Note that $\mathbf{\Sigma}_2 = \mathbf{L}^{-1/2}\mathbf{H}\mathbf{L}^{-1/2}$, and $|\mathbf{L}| < m^m$. Using these, and if we use $K$ to denote a suitable finite, positive and generic constant, not depending on any parameters, we get that

$$|\mathbf{\Sigma}_2|^{-1/2} \leq K \prod_{i=1}^{m}\{\delta^* + \sigma^2(1 - \rho\nu_i)^{-2}\}^{-1/2}. \tag{28}$$

Let $\mathbf{P}_{\mathbf{W}_*}$ be the matrix of eigenvectors of $\mathbf{W}_*$ such that $\mathbf{P}_{\mathbf{W}_*}^T\mathbf{W}_*\mathbf{P}_{\mathbf{W}_*} = \mathrm{diag}(\nu_1, \ldots, \nu_m) = \mathbf{N}_*$. Then,

$$\begin{aligned}
& (\mathbf{d}_* - \mathbf{G}_*\phi)^T(\delta^*\mathbf{I}_m + \sigma^2\mathbf{\Omega}_2^{-1})^{-1}(\mathbf{d}_* - \mathbf{G}_*\phi) \\
= \; & (\mathbf{L}^{1/2}\mathbf{d}_* - \mathbf{L}^{1/2}\mathbf{G}_*\phi)^T\mathbf{H}^{-1}(\mathbf{L}^{1/2}\mathbf{d}_* - \mathbf{L}^{1/2}\mathbf{G}_*\phi) \\
\geq \; & l_{(m)}^{-1}(\mathbf{r} - \mathbf{F}\boldsymbol{\phi})^T\{\delta^*\mathbf{I} + \sigma^2(\mathbf{I} - \rho\mathbf{W}_*)^{-2}\}^{-1}(\mathbf{r} - \mathbf{F}\boldsymbol{\phi}) \\
= \; & (\tilde{\mathbf{r}} - \tilde{\mathbf{S}}\boldsymbol{\phi})^T\{\delta^*\mathbf{I} + \sigma^2(\mathbf{I} - \rho\mathbf{N}_*)^{-2}\}^{-1}(\tilde{\mathbf{r}} - \tilde{\mathbf{S}}\boldsymbol{\phi}),
\end{aligned} \tag{29}$$

where $\mathbf{r} = \mathbf{L}^{1/2}\mathbf{d}_*,\ \mathbf{F} = \mathbf{L}^{1/2}\mathbf{G}_*$, $\tilde{\mathbf{r}} = l_{(m)}^{-1/2}\mathbf{P}_{\mathbf{W}_*}\mathbf{r}$, and $\tilde{\mathbf{S}} = l_{(m)}^{-1/2}\mathbf{P}_{\mathbf{W}_*}\mathbf{F}$.

Suppose $\{i_1, \ldots, i_q\}$ is a subset of $\{1, \ldots, m\}$ so that the matrix $\tilde{\mathbf{S}}_1$ formed by plucking the rows of $\tilde{\mathbf{S}}$ corresponding to the indices $\{i_1, \ldots, i_q\}$ is non-singular. Note that this matrix is determined by $\mathbf{W}$.

From (29) we get

$$(\mathbf{d}_* - \mathbf{G}_*\phi)^T(\delta^*\mathbf{I}_m + \sigma^2\mathbf{\Omega}_2^{-1})^{-1}(\mathbf{d}_* - \mathbf{G}_*\phi) \geq \sum_{j=1}^{q}\frac{(\tilde{r}_{i_j} - \tilde{\mathbf{s}}_{i_j}^T\boldsymbol{\phi})^2}{\delta^* + \sigma^2(1 - \rho\nu_{i_j})^{-2}}. \tag{30}$$

Using equations (29)-(30) we get

$$\begin{aligned}
&\int \pi(\beta) \exp[-\frac{(\mathbf{d}_* - \mathbf{G}_*\phi)^T(\delta^*\mathbf{I}_m + \sigma^2\boldsymbol{\Omega}_2^{-1})^{-1}(\mathbf{d}_* - \mathbf{G}_*\phi)}{2}]d\boldsymbol{\phi} \\
\leq\ & K \int \exp[-\frac{1}{2}\sum_{j=1}^{q} \frac{(\tilde{r}_{i_j} - \tilde{\mathbf{s}}_{i_j}^T\boldsymbol{\phi})^2}{\delta^* + \sigma^2(1-\rho\nu_{i_j})^{-2}}]d\boldsymbol{\phi} \\
=\ & K\prod_{i=1}^{q}\{\delta^* + \sigma^2(1-\rho\nu_{i_j})^{-2}\}^{1/2}. \qquad (31)
\end{aligned}$$

Hence we get

$$\begin{aligned}
&|\boldsymbol{\Sigma}_2|^{-1/2}\int \pi(\beta) \exp[-\frac{(\mathbf{d}_* - \mathbf{G}_*\phi)^T(\delta^*\mathbf{I}_m + \sigma^2\boldsymbol{\Omega}_2^{-1})^{-1}(\mathbf{d}_* - \mathbf{G}_*\phi)}{2}]d\boldsymbol{\phi} \\
\leq\ & K \prod_{i\notin\{i_1,\ldots,i_q\}} \{\delta_* + \sigma^2(1-\rho\nu_i)^{-2}\}^{1/2} \\
\leq\ & K[I(\sigma^2 \leq N) + I(\sigma^2 > N)(\sigma^2)^{-(m-q)/2}\prod_{i\notin\{i_1,\ldots,i_q\}}(1-\rho\nu_i)] \\
\leq\ & K[I(\sigma^2 \leq N) + I(\sigma^2 > N)(\sigma^2)^{-(r-p)/2}], \qquad (32)
\end{aligned}$$

where we use the facts that $-1 < \rho < 1$ and $-1 \leq \nu_i \leq 1$ to claim that $0 < 1 - \rho\nu_i < 2$ for all $i$. From equation (32) if we continue our proof along the lines of the proof for the SCAR model, we will get the propriety of the posterior pdf for the SAR model under the same conditions.

## Disclaimer and acknowledgments

This research was funded by a grant from the United States Department of State (SSJTIP18CA0015). This report is released to inform interested parties of ongoing research and to encourage discussion. The views, opinions, findings and conclusions expressed on statistical, methodological, technical, or operational issues herein are those of the authors and not those of the United States Department of State, the U.S. Census Bureau, Wells Fargo Bank, or the University of Georgia.

## References

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **83**, 28–36.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media.

Chung, H. C. and Datta, G. S. (2022). Bayesian spatial models for estimating means of sampled and non-sampled small areas. *Survey Methodology*, **48**, 463–489.

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

Ghosh, M. (1992). Hierarchical and empirical Bayes multivariate estimation. In Ghosh, M. and Pathak, P. K., editors, *Current Issues in Statistical Inference: Essays in honor of D. Basu*, pages 151–177. Institute of Mathematical Statistics.

Ghosh, M., Nangia, N., and Kim, D. H. (1996). Estimation of median income of four-person families: a Bayesian time series approach. *Journal of the American Statistical Association*, **91**, 1423–1431.

Rao, J. N. K. and Molina, I. (2015). *Small Area Estimation.* John Wiley & Sons.

Stan Development Team (2018). RStan: the R interface to Stan. R package version 2.17.3.

Vogt, M., Lahiri, P., and Munnich, R. (2023). Spatial prediction in small area estimation. *Statistics in Transition new series*, **24**.