

A Bayesian Semi-parametric Modelling Approach for Area Level Small Area Studies

Calcutta Statistical Association Bulletin

76(1) 78–95, 2024

© 2023 Calcutta Statistical Association, Kolkata

Article reuse guidelines:

in.sagepub.com/journals-permissions-india

DOI: 10.1177/00080683231198606

journals.sagepub.com/home/csa

**Marten Thompson¹ and Snigdhanu Chatterjee¹**

Abstract

We present a new semiparametric extension of the Fay-Herriot model, termed the *agnostic Fay-Herriot* model (AGFH), in which the sampling-level model is expressed in terms of an unknown general function $g(\cdot)$. Thus, the AGFH model can express any distribution in the sampling model since the choice of $g(\cdot)$ is extremely broad. We propose a Bayesian modelling scheme for AGFH where the unknown function $g(\cdot)$ is assigned a Gaussian Process prior. Using a Metropolis within Gibbs sampling Markov Chain Monte Carlo scheme, we study the performance of the AGFH model, along with that of a hierarchical Bayesian extension of the Fay-Herriot model. Our analysis shows that the AGFH is an excellent modelling alternative when the sampling distribution is non-Normal, especially in the case where the sampling distribution is bounded. It is also the best choice when the sampling variance is high. However, the hierarchical Bayesian framework and the traditional empirical Bayesian framework can be good modelling alternatives when the signal-to-noise ratio is high, and there are computational constraints.

Keywords

Agnostic Fay-Herriot model, metropolis within Gibbs, hierarchical Bayes

AMS Subject Classification: 62D05; 62F15

1. Introduction

In area-level small area studies, the Fay-Herriot model Fay III and Herriot^[9] is a commonly used framework and is given by the following two-layered structure:

1. **Sampling model:** Conditional on the unknown and unobserved independent *area-level effects* $\{\theta_i : i = 1, 2, \dots, M\}$ on M small areas, the observed data $\mathbf{Y} = (Y_1, \dots, Y_M)^T$ are themselves

¹ University of Minnesota, Minneapolis, USA

Corresponding author:

Snigdhanu Chatterjee, University of Minnesota, School of Statistics, Ford Hall, 224 Church Street SE, Minneapolis MN 55455, USA.

E-mail: chatt019@umn.edu

independent random variables. The i -th area observation Y_i follows a Normal distribution with mean θ_i and precision D_i with the latter typically assumed to be known.

2. **Linking model:** The unobserved area-level effect for the i -th area, θ_i , has mean $x_i^\top \beta$, determined by the independent variable $x_i \in \mathbb{R}^p$ and unknown β parameter, and variance λ .

In the above models and throughout this article, all vectors are column vectors, and the notation $a^\top$ denotes the transpose of (vector or matrix) a . The $M \times p$ matrix of auxiliary variables $\mathbf{X}$ whose i -th row is given by $x_i \in \mathbb{R}^p$ is assumed to be full column rank. Note that some authors parameterize the sampling model in terms of (known) variance terms; we use the equivalent precision specification for notational ease when manipulating this likelihood. *Thus, in this article, the notation D_i represents the **precision** of the sampling variance, and not its variance.*

Discussions, including theoretical as well a variety of applications of the above framework may be found in the book Rao and Molina^[22] and in the papers Jiang et al.^[12]; Das et al.^[5]; Pfeffermann and Glickman^[18]; Datta et al.^[8]; Rao^[20]; Jiang and Lahiri^[11]; Chatterjee et al.^[3]; Li and Lahiri^[15]; Salvati et al.^[24]; Datta et al.^[6]; Pfeffermann^[17]; Yoshimori and Lahiri^[30, 31]; Molina et al.^[16]; Sugawara and Kubokawa^[25]; Rao^[21].

Notationally, the Fay-Herriot model may be written as

$$Y_i \mid \theta_i \sim N(\theta_i, D_i^{-1}), \text{ independently, and} \quad (1.1)$$

$$\theta_i \sim N(x_i^\top \beta, \lambda), \text{ independently, for } i = 1, \dots, M. \quad (1.2)$$

Thus, we have the marginal distribution specification

$$Y_i \sim N(x_i^\top \beta, \lambda + D_i^{-1}), \text{ independently, for } i = 1, \dots, M.$$

Some routine frequentist computations can be readily performed on this model, and we can first obtain an ordinary least squares (OLS) estimator of β as follows:

$$\hat{\beta}_{OLS} = \left[\sum_{i=1}^M x_i x_i^\top \right]^{-1} \left[\sum_{i=1}^M x_i Y_i \right].$$

Using this, we may obtain the moment-based estimator of λ as in Rao and Molina^[22], that of $\hat{\lambda} = \max(0, \tilde{\lambda})$ where

$$\tilde{\lambda} = \frac{1}{M-p} \left[\sum_{i=1}^M (Y_i - x_i^\top \hat{\beta}_{OLS})^2 - \sum_{i=1}^M D_i^{-1} (1 - h_{ii}) \right], \text{ and}$$

$$h_{ii} = x_i^\top \left(\sum_{i=1}^M x_i x_i^\top \right)^{-1} x_i.$$

Works on profile and residual likelihoods, and adjustments thereof, have made improvements in estimating λ over the moment-based estimator. Specifically, we also consider the adjusted residual likelihood estimator introduced by Yoshimori and Lahiri^[30].

The central problem in small area statistics is to predict the θ_i 's. One natural predictor is the empirical version of the *Best Linear Unbiased Predictor* (BLUP). First, define the quantities

$$\gamma_i = \lambda / (D_i^{-1} + \lambda), \quad \text{and}$$

$$\tilde{\beta} = \left[\sum_{i=1}^M x_i x_i^\top / (D_i^{-1} + \lambda) \right]^{-1} \left[\sum_{i=1}^M x_i Y_i / (D_i^{-1} + \lambda) \right],$$

and their empirical versions, with estimates $\hat{\lambda}$ plugged in

$$\hat{\gamma}_i = \hat{\lambda} / (D_i^{-1} + \hat{\lambda}), \quad \text{and}$$

$$\hat{\beta} = \left[\sum_{i=1}^M x_i x_i^\top / (D_i^{-1} + \hat{\lambda}) \right]^{-1} \left[\sum_{i=1}^M x_i Y_i / (D_i^{-1} + \hat{\lambda}) \right].$$

Based on these, we follow Rao and Molina^[22] and define the BLUP and its empirical version EBLUP as follows:

$$\begin{aligned} \text{BLUP: } \tilde{\theta}_i &= x_i^\top \tilde{\beta} + \gamma_i (Y_i - x_i^\top \tilde{\beta}) = \gamma_i Y_i + (1 - \gamma_i) x_i^\top \tilde{\beta}, \quad \text{and} \\ \text{EBLUP: } \hat{\theta}_i &= x_i^\top \hat{\beta} + \hat{\gamma}_i (Y_i - x_i^\top \hat{\beta}) = \hat{\gamma}_i Y_i + (1 - \hat{\gamma}_i) x_i^\top \hat{\beta}, \quad i = 1, \dots, M. \end{aligned}$$

The above frequentist predictors are straightforward and widely used but seem to depend critically on the parametric assumptions given in (1.1) being correct. However in real applications, the Fay-Herriot and other area or unit level small area models are used on observed data that may be bounded in nature or strictly positive. Examples of such cases are proportions and variables like income or lifespan. While (1.1) is often justifiable when large samples are available, this may not always be the case in small area contexts. At the very least, the robustness of the EBLUP-based prediction technique against distributional misspecifications needs to be studied.

In this article, we present a semiparametric extension to the sampling distribution component of the Fay-Herriot model. The semiparametric extension is designed to encapsulate any distributional choice, consequently, we call this extended model the *agnostic Fay-Herriot* model (AGFH often hereafter). We present a Bayesian modelling technique using a Gaussian Process prior for the unknown functional component of the AGFH model and some theoretical properties of the proposed model. Additionally, we detail a Bayesian computational approach to estimating the AGFH model, which involves a Metropolis within Gibbs Markov Chain Monte Carlo procedure.

We compare the performance of the AGFH model with that of three other techniques in a variety of numeric experiments. We study four different choices of sampling distributions, including the traditional Fay-Herriot framework and cases where the observations are non-negative or bounded. The nature of the sampling distribution is the signal the AGFH learns amidst the noise of the linking model's variability, and multiple signal-to-noise ratio conditions are studied. The rival techniques we consider include the EBLUP-based approach outlined above, and a hierarchical Bayesian (HB) technique based on the traditional Fay-Herriot model (1.1)(1.2). In addition, we also study the performance of a regression-based predictor that ignores the multi-level dependency present in small area models but is straightforward to implement. In many of the cases we consider, especially where

the sampling distribution is bounded or where the sampling variance is high, the AGFH registers a lower Mean Squared Prediction Error (MSPE) value compared to other techniques. However, there are interesting and subtle details about the robustness aspect of the EBLUP-based prediction or hierarchical Bayesian prediction, which we discuss later in the article.

To ensure fair comparison, we always restrict the first two moments of the observable data $\{Y_1, \dots, Y_M\}$ to adhere to the Fay-Herriot model specifications. That is, we always impose the following conditions:

$$\mathbb{E}(Y_i) = x_i^T \beta, \quad (1.3)$$

$$\mathbb{V}(Y_i) = D_i^{-1} + \lambda. \quad (1.4)$$

Thus, based on the first two moments, the traditional frequentist approach outlined above remains a reasonable and valid technique to use. Using the moment conditions (1.3) and (1.4) is important to ensure a level playing field for the traditional EBLUP-based approach; the HB approach and the semi-parametric modelling approach easily adapts to alternate sets of conditions.

In this article, we report results only for the case where the *sampling distribution* is potentially misspecified. If instead only the linking model were misspecified, then the analysis and results are similar to the ones reported here, and we do not include them to reduce redundancy. When both the sampling and the linking models are misspecified, we need considerably more technical assumptions and mathematical details to ensure identifiability and address theoretical and algorithmic challenges; these will be presented in a future article. Note that for the hierarchical Bayesian study, we are interested in misspecification in the Fay-Herriot model itself, thus we do not address Bayesian robustness questions where robustness with respect to prior specifications is studied. Our preliminary studies on Bayesian robustness, for reasonable choices of priors, do not produce predictions that are fundamentally dissimilar to the results reported here. We will also address Bayesian robustness issues in a future publication.

Robustness in small area problems has been studied from several other perspectives earlier, but to the best of our knowledge not from the distributional misspecification and Bayesian computational viewpoint that we present in this article. A comprehensive recent review of existing perspectives is available in Jiang and Rao^[13]. A major initial work in small area robustness studies is Lahiri and Rao^[14], where it was shown that for the problem of estimating the Mean Squared Prediction Error (MSPE), the normality-based Prasad-Rao MSPE estimator remains second-order unbiased under non-normality of the latent linking model when a simple method-of-moments estimator (e.g., Rao and Molina^[22] Section 6.1.2) is used for the variance component and the sampling error distribution is normal. However, in Chen et al.^[4], it was established that the normality-based MSPE estimator is no longer second-order unbiased when the sampling error distribution is non-normal or when the Fay-Herriot moment method is used to estimate the variance component, even when the sampling error distribution is normal. The robust estimation perspective of MSPE is presented in Wu and Jiang^[29]. Robustness with respect to data issues is studied in Chatterjee^[2]. Robustness in small area models have also been explored from the perspective of having more broad modelling assumptions in place of (1.1). In Datta and Mandal^[7], a general model with uncertain linking model was considered. More recently, Chakraborty et al.^[11] proposed an extension of the traditional Fay-Herriot model where the linking model is given by a mixture of two Normal distributions. In Ghosh et al.^[10], Student's t-distribution and a hierarchical Bayesian modelling was used in studying robustness.

We discuss the hierarchical Bayesian analysis of the traditional Fay-Herriot model in Section 2. Following this, we present the semiparametric extension of the Fay-Herriot model, the AGFH, in Section 3. In Section 3 we discuss some theoretical properties of this model as well. Following this, in Section 4 we present the methodological details of the Bayesian analysis in the AGFH model. In Section 5, we report result from numerical simulation experiments, and compare performances of different small area predictors under two different scenarios. Our concluding remarks are collected in Section 6.

2. Hierarchical Bayesian Fay-Herriot Model

For a hierarchical Bayesian study, we augment (1.1) and (1.2) with priors on the parameters β and λ . We use a uninformative prior on β , and an inverse Gamma prior on λ . Notationally, the prior is specified by

$$f(\beta, \lambda) \propto 1 \cdot f(\lambda) = \text{IG}(a, b),$$

where the notation $\text{IG}(a, b)$ stands for *inverse Gamma* distribution with hyper-parameters $a > 0$ and $b > 0$. The probability density function of this is given by

$$f(\lambda; a, b) = \frac{b^a}{\Gamma(a)} (1/\lambda)^{a+1} \exp\{-b/\lambda\}, \quad \lambda > 0.$$

Let us use the notation $\boldsymbol{\theta} = (\theta_1, \dots, \theta_M)^\top \in \mathbb{R}^M$, and $\mathbf{Y} = (Y_1, \dots, Y_M)^\top \in \mathbb{R}^M$. Hierarchical Bayesian approaches in small-area models have been attempted in Polettini^[19], Wanjoya et al.^[28] and several other papers. In the current context, the unknown (hyper-)parameters are $\boldsymbol{\theta}, \lambda, \beta$.

We set up a Gibbs sampling scheme for the Markov Chain Monte Carlo procedure to approximate the posterior distribution of these parameters. Note that conditional on $\boldsymbol{\theta}$, the parameters β and λ are independent of $\mathbf{Y}$. Define

$$\begin{aligned} \beta^*(\boldsymbol{\theta}) &= \left[\sum_{i=1}^M x_i x_i^\top \right]^{-1} \left[\sum_{i=1}^M x_i \theta_i \right], \\ \gamma_i(\lambda) &= \lambda / (D_i^{-1} + \lambda) \quad \text{and} \\ \theta_i^*(\beta, \lambda) &= \gamma_i Y_i + (1 - \gamma_i) x_i^\top \beta. \end{aligned}$$

Then the conditional distributions for the Gibbs sampling algorithm are

$$\beta \mid \boldsymbol{\theta}, \lambda, \mathbf{Y} \sim N_p \left(\beta^*, \lambda \left(\sum_{i=1}^M x_i x_i^\top \right)^{-1} \right), \quad (2.1)$$

$$\lambda \mid \beta, \boldsymbol{\theta}, \mathbf{Y} \sim \text{IG} \left(a + M/2, b + \frac{1}{2} \sum_{i=1}^M (\theta_i - x_i^\top \beta)^2 \right), \quad (2.2)$$

$$\theta_i \mid \beta, \lambda, \mathbf{Y} \sim N(\theta_i^*(\beta, \lambda), \gamma_i D_i), \quad i = 1, \dots, M. \quad (2.3)$$

3. Semi-parametric Fay-Herriot Model Extension

In this section, we propose an extension of the Fay-Herriot model (1.1)–(1.2) using a semiparametric framework. We use the term *agnostic Fay-Herriot model* (AGFH in short often hereafter) to denote this model as it relaxes the distributional assumptions on the sampling level in the traditional Fay-Herriot model.

Suppose we have an unknown function $g : \mathbb{R} \rightarrow \mathbb{R}$ satisfying the following conditions for some $K > 0$:

$$\int_{u=-\infty}^{\infty} \exp\left\{-\frac{u^2}{2} + g(u)\right\} du = K, \quad (3.1)$$

$$\int_{u=-\infty}^{\infty} u \exp\left\{-\frac{u^2}{2} + g(u)\right\} du = 0, \quad (3.2)$$

$$K^{-1} \int_{u=-\infty}^{\infty} u^2 \exp\left\{-\frac{u^2}{2} + g(u)\right\} du = 1. \quad (3.3)$$

Without loss of generality, we may assume that $K = 1$ since any constant may be easily absorbed in $g(\cdot)$, but we will retain the notation K for clarity of presentation of the theoretical developments below.

Condition (3.1) ensures that $K^{-1} \exp\left\{-\frac{u^2}{2} + g(u)\right\}$ is a probability density function. Let U be the generic notation for a random variable that has this density function, and let $\{U_i : i = 1, \dots, M\}$ be iid copies of U . Note that conditions (3.2) and (3.3) imply that $\mathbb{E}(U) = 0$ and $\mathbb{V}(U) = 1$.

Let $\{Z_i : i = 1, \dots, M\}$ be independent standard Normal random variables that are independent of $\{U_i\}$. The latent variables driving the agnostic Fay-Herriot model (AGFH) are

$$\theta_i = x_i^\top \beta + \lambda^{1/2} Z_i, \quad i = 1, \dots, M. \quad (3.4)$$

Conditional on θ_i , the observed data are realizations of the random variable

$$Y_i = \theta_i + D_i^{-1/2} U_i, \quad i = 1, \dots, M. \quad (3.5)$$

Notice that (3.4) presents a generalization of the linking model (1.2). If $g(\cdot) \equiv 0$, then (3.4) is identical to (1.2) and we thus recover the classical Fay-Herriot model. However, for other choices of $g(\cdot)$ function, we obtain realizations of an area-level small area model with a non-Gaussian sampling error structure. The above AGFH model is semi-parametric in nature, owing to the presence of both finite-dimensional parameters β and λ and an infinite-dimensional parameter $g(\cdot)$.

3.1. Some Theoretical Preliminaries

We present below some probabilistic and theoretical properties related to the AGFH model which will be useful for methodological steps outlined later in this article. For notational simplicity, in several of the algebraic steps below we will drop the subscript i when there is no source of confusion.

For any $y \in \mathbb{R}$ we have

$$\begin{aligned}
 & \mathbb{P}[Y \leq y | \beta, \lambda, g, \theta, x] \\
 &= \mathbb{P}[\theta + D^{-1/2}U \leq y | \beta, \lambda, g, \theta, x] \\
 &= \mathbb{P}[U \leq D^{1/2}(y - \theta) | \beta, \lambda, g, \theta, x] \\
 &= K^{-1} \int_{u=-\infty}^{D^{1/2}(y-\theta)} \exp\left\{-\frac{u^2}{2} + g(u)\right\} du \\
 &= K^{-1} D^{1/2} \int_{t=-\infty}^y \exp\left\{-\frac{1}{2}D(t - \theta)^2 + g(D^{1/2}(t - \theta))\right\} dt.
 \end{aligned}$$

using $t = \theta + D^{-1/2}u$. Thus, the probability density function of Y_i , conditional on $\beta, \lambda, g, \theta_i, x_i$ is

$$p(y | \beta, \lambda, g, \theta_i, x_i) = K^{-1} D_i^{1/2} \exp\left\{-\frac{1}{2}D_i(y - \theta_i)^2 + g(D_i^{1/2}(y - \theta_i))\right\} \quad \text{for } y \in \mathbb{R}.$$

Thus, the AGFH model can be expressed more compactly using the following conditions:

$$\log p(Y_i | \beta, \lambda, g, \theta_i, x_i) \propto -\frac{1}{2}D_i(Y_i - \theta_i)^2 + g(D_i^{1/2}(Y_i - \theta_i)), \quad (3.6)$$

$$\theta_i | \beta, \lambda, g, x_i \sim N(x_i^\top \beta, \lambda). \quad (3.7)$$

Note that in principle, the AGFH model can express many distributions in the sampling model, since the choice of $g(\cdot)$ is extremely broad. The only conditions on $g(\cdot)$ are (3.1)–(3.3), and innumerable functions satisfy these conditions. Note that in Section 4 we consider a Bayesian analysis of the AGFH model with a Gaussian Process (GP) prior on $g(\cdot)$. This further restricts the set of functions in which $g(\cdot)$ belongs.

Let us now compute the first two moments of Y_i , conditional on $\beta, \lambda, g, \theta_i, x_i$. For the first moment, we have

$$\begin{aligned}
 & D^{1/2} K^{-1} \int_{y=-\infty}^{\infty} y \exp\left\{-\frac{1}{2}D(y - \theta)^2 + g(D^{1/2}(y - \theta))\right\} dy \\
 &= K^{-1} \int_{u=-\infty}^{\infty} (\theta + D^{-1/2}u) \exp\left\{-\frac{u^2}{2} + g(u)\right\} du \\
 &= \theta + K^{-1} D^{-1/2} \int_{u=-\infty}^{\infty} u \exp\left\{-\frac{u^2}{2} + g(u)\right\} du \\
 &= \theta,
 \end{aligned}$$

using (3.2). For the second non-central moment we have

$$\begin{aligned}
 & D^{1/2} K^{-1} \int_{y=-\infty}^{\infty} y^2 \exp\left\{-\frac{1}{2} D(y - \theta)^2 + g(D^{1/2}(y - \theta))\right\} dy \\
 &= K^{-1} \int_{u=-\infty}^{\infty} (\theta + D^{-1/2}u)^2 \exp\left\{-\frac{u^2}{2} + g(u)\right\} du \\
 &= \theta^2 + D^{-1} K^{-1} \int_{u=-\infty}^{\infty} u^2 \exp\left\{-\frac{u^2}{2} + g(u)\right\} du \\
 &= \theta^2 + D^{-1},
 \end{aligned}$$

using (3.1), (3.2), and (3.3). So, the conditional variance of Y_i is D_i^{-1} . These two moments match the conditional mean and variance of the traditional Fay-Herriot model given in (1.1).

The marginal mean and variance of Y_i follow from the laws of total expectation and variance. First,

$$\begin{aligned}
 \mathbb{E}(Y) &= \mathbb{E}(\mathbb{E}(Y|\theta)) \\
 &= \mathbb{E}(\theta) = x^\top \beta.
 \end{aligned} \tag{3.8}$$

Similarly

$$\begin{aligned}
 \mathbb{V}(Y) &= \mathbb{E}(\mathbb{V}(Y|\theta)) + \mathbb{V}(\mathbb{E}(Y|\theta)) \\
 &= \mathbb{E}(D^{-1}) + \mathbb{V}(\theta) = D^{-1} + \lambda
 \end{aligned} \tag{3.9}$$

Thus, the AGFH also matches the FH in terms of the marginal mean and variance expressions (1.3) (1.4) for the observed data.

Next, define $\tilde{U}$ as follows

$$\tilde{U}_i = D_i^{1/2} (Y_i - x_i^\top \beta).$$

Rearranging (3.5) yields

$$\begin{aligned}
 \tilde{U}_i &= D_i^{1/2} (Y_i - \theta_i) + D_i^{1/2} (\theta_i - x_i^\top \beta) \\
 &= U_i + \tilde{Z}_i,
 \end{aligned}$$

where $\{\tilde{Z}_i \sim N(0, D_i \lambda)\}$ are an independent sequence of random variables independent of $\{U_i\}$. Let us compute the probability density function of $\tilde{U}_i$. For any $u \in \mathbb{R}$, we have

$$\begin{aligned}
 \mathbb{P}[\tilde{U} \leq u] &= \mathbb{P}[U + \tilde{Z} \leq u] \\
 &= \mathbb{E}\mathbb{P}[U \leq u - \tilde{Z} | \tilde{Z}]
 \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} K^{-1} \int_{t=-\infty}^{u-\tilde{Z}} \exp\left\{-\frac{t^2}{2} + g(t)\right\} dt \\
&= \mathbb{E} K^{-1} \int_{v=-\infty}^u \exp\left\{-\frac{(v-\tilde{Z})^2}{2} + g(v-\tilde{Z})\right\} dv \\
&= (2\pi D\lambda)^{-1/2} K^{-1} \int_{s=-\infty}^{\infty} \int_{v=-\infty}^u \exp\left\{-\frac{(v-s)^2}{2} + g(v-s) - \frac{s^2}{2D\lambda}\right\} dv ds \\
&\propto \int_{v=-\infty}^u \left[\int_{s=-\infty}^{\infty} \exp\left\{-\frac{(v-s)^2}{2} + g(v-s) - \frac{s^2}{2D\lambda}\right\} ds \right] dv.
\end{aligned}$$

Thus, the density of $\tilde{U}_i$ is

$$f_i(v) \propto \int_{s=-\infty}^{\infty} \exp\left\{-\frac{(v-s)^2}{2} - \frac{s^2}{2D_i\lambda} + g(v-s)\right\} ds \quad (3.10)$$

3.2. Frequentist Parameter Estimation in AGFH

Note that the finite dimensional parameters β and λ , defined in equations (1.3) and (1.4), are also present in the AGFH model in moment equations (3.8) and (3.9). Consequently, we can use equations (3.8) and (3.9) to estimate these parameters in the AGFH model. Additionally, define the statistics

$$\hat{U}_i = D_i^{1/2} (Y_i - x_i^\top \hat{\beta}), \quad i = 1, \dots, M.$$

These in turn, can be used (3.10) to estimate $g(\cdot)$ nonparametrically. However, using this frequentist semiparametric estimation procedure is computationally cumbersome owing to the fact that (3.10) is an intricate function of the infinite dimensional parameter $g(\cdot)$. In a future publication, we will report on approaches where $g(\cdot)$ is approximated using basis functions on an appropriate Hilbert space.

4. Bayesian Approach for Analysing the AGFH Model

The AGFH model lends itself readily to Bayesian analysis, which we study below. On the finite dimensional parameters β and λ , we assume the same prior that we used earlier for hierarchical Bayesian analysis in the original Fay-Herriot model, as given in Section 2 above. Additionally, we assume that the function $g(\cdot)$ is a random function that has a *Gaussian Process prior*, with a covariance function indexed by hyperparameters denoted by α . Notationally, the prior for the AGFH model is given by

$$\begin{aligned}
(\beta, \lambda) &\sim 1 \cdot \text{IG}(a, b), \quad \text{and independently} \\
g \mid \alpha &\sim \mathcal{GP}(0, \kappa(\cdot, \cdot; \alpha)).
\end{aligned} \quad (4.1)$$

Here, $\kappa(\cdot, \cdot; \alpha)$ is a kernel function on $\mathbb{R}$ with parameter α . Thus, (4.1) implies that for any positive integer $n \in \{1, 2, \dots\}$ and any collection of points $\{t_1, \dots, t_n\} \subset \mathbb{R}$, the evaluation of the function $g(\cdot)$ at these points, that is, the vector $\mathbf{g} = (g(t_1), g(t_2), \dots, g(t_n))^\top$, follows an n -dimensional Gaussian distribution with mean zero, and the (i, j) -th entry of the variance-covariance matrix given by $\kappa(t_i, t_j; \alpha)$. As long as $\kappa(\cdot, \cdot; \alpha)$ is a positive definite function, the fact that this formulation satisfies the Kolmogorov consistency conditions and indeed defines a valid stochastic process is discussed in many standard references related to Gaussian Processes, see for example Rasmussen and Williams^[23]. For simplicity in presentation, in this article we consider the radial basis kernel function, given by

$$\kappa(x, y; \alpha) = \alpha_0 \exp\{-\alpha_1^{-1}(x - y)^2\}, \quad x, y \in \mathbb{R}$$

and we fix the hyperparameter $\alpha = (\alpha_0, \alpha_1)^\top = (0.1, 0.1)$. While the Bayesian computations we performed do not appear to be sensitive to reasonable choices of kernel functions or hyperparameters, additional details pertaining to robustness of Gaussian Process formulation will be studied in future works. Note however, our model assumptions above ensure that the posterior distribution is proper, we omit the algebraic details of computations relating to this matter.

4.1. Metropolis-Hastings Within Gibbs Computational Approach

The unknown quantities in the AGFH framework are β, λ, θ , and $g(\cdot)$, and the additional random quantity are the observables $\mathbf{Y}$. Similar to the hierarchical Bayesian framework for the traditional Fay-Herriot model, notice that conditional on θ and $g(\cdot)$, the parameters β and λ are independent of $\mathbf{Y}$. Consequently, we note that

$$p(\beta \mid \theta, \lambda, g, \alpha, \mathbf{Y}) = N_p \left(\beta^*, \lambda \left(\sum_{i=1}^M x_i x_i^\top \right)^{-1} \right). \quad (4.2)$$

Similarly,

$$\begin{aligned} p(\lambda \mid \theta, \beta, g, \alpha, \mathbf{Y}) &\propto p(\mathbf{Y} \mid \theta, g, \alpha) p(\theta \mid \beta, \lambda) p(\beta, \lambda) \\ &\propto p(\theta \mid \beta, \lambda) p(\beta, \lambda) \\ &\propto \frac{1}{\lambda^{M/2}} \exp \left(- \sum_{i=1}^M \frac{(\theta_i - x_i^\top \beta)^2}{2\lambda} \right) \frac{1}{\lambda^{a-1}} \exp \left(\frac{b}{\lambda} \right) \\ &\propto \text{IG} \left(a + M/2, b + \frac{1}{2} \sum_{i=1}^M (\theta_i - x_i^\top \beta)^2 \right). \end{aligned} \quad (4.3)$$

Thus, conditional on θ and β , the distribution of λ does not depend on $g(\cdot)$, α , and $\mathbf{Y}$. Similarly, conditional on θ and λ , the distribution of β does not depend on $g(\cdot)$, α , and $\mathbf{Y}$.

We individually sample the θ_i according the following MH likelihood.

$$p(\theta_i \mid \beta, \lambda, g, Y_i) \propto p(Y_i \mid \theta_i, g)p(\theta_i \mid \beta, \lambda)p(\beta, \lambda)p(g) \quad (4.4)$$

$$\exp\left(-\frac{D_i(Y_i - \theta_i)^2}{2} + g(D_i^{1/2}(Y_i - \theta_i)) - \frac{(\theta_i - x_i\beta)^2}{2\lambda}\right) \quad (4.5)$$

Note that at this point β , λ , and g are fixed within each step, and consequently their likelihoods do not enter into the computations.

Finally, $g(\cdot)$ is updated by treating a kernel density estimate of the residuals between Y_i and the current Gibbs estimate $\theta_i^{(h)}$ as observations of its target function. Recalling (3.5), the residuals

$$U_i^{(h)} = \sqrt{D_i}(Y_i - \theta_i^{(h)})$$

approximate draws from (3.1). At each sampling iteration, let $\{w_i^{(h)}\}_{i=1}^M$ be a KDE estimate of the density of $\{U_i^{(h)}\}_{i=1}^M$. Define

$$v_i^{(h)} = \log(w_i^{(h)}) + (U_i^{(h)})^2/2. \quad (4.6)$$

Then, treat $\{v_i^{(h)}\}_{i=1}^M$ as observations of $g(\cdot)$; when $g(\cdot)$ is required at other steps during sampling, perform the typical Gaussian Process posterior analysis. The entire AGFH sampler is described in Algorithm 1.

Algorithm 1: Agnostic Fay-Herriot Sampler

Given: Y, X, D , desired number of samples H .

Initialize $\beta^{(0)}, \lambda^{(0)}, \theta^{(0)}$.

for $h = 1$ **to** H **do**

$\beta^{(h)} \mid \theta^{(h-1)}, \lambda^{(h-1)}$ sampled according to (4.2).

$\theta^{(h)} \mid \beta^{(h)}, \lambda^{(h-1)}$ sampled with Metropolis-Hastings using (4.5).

$\lambda^{(h)} \mid \beta^{(h)}, \theta^{(h)}$ sampled according to (4.3).

 Update g with values in (4.6).

Return: Samples $(\beta^{(1)}, \dots, \beta^{(H)}), (\theta^{(1)}, \dots, \theta^{(H)}), (\lambda^{(1)}, \dots, \lambda^{(H)})$, and final g .

5. Results from Simulation Experiments

The potential advantage of the agnostic Fay-Herriot model is its ability to accommodate non-Normal error structures in the Fay-Herriot sampling model. Doing so may improve the sampling and estimation of the small area means θ_i . Comparing the HB Gibbs conditional distribution for θ_i (2.3) to that of (4.5) for AGFH, $g(\cdot)$ manipulates the latter's likelihood based on learned behavior of the sampling error. Similarly, frequentist and empirical Bayes methods assume Normal sampling errors.

We explore that here in the presence of Gamma, symmetric Beta, and asymmetric Beta errors. Doing so also provides an opportunity to investigate the robustness extant estimation methods exhibit in the presence of these errors. In all cases, sampling errors are transformed to satisfy the moment conditions (1.3) and (1.4).

Three estimation methods are explored in addition to the AGFH model. First, the the standard hierarchical Bayes (HB) Gibbs sampler from Section 2 provides a Bayesian baseline to the AGFH. Both the HB and AGFH samplers use the prior $\lambda \sim \text{IG}(1, 1)$ and were run 10^4 steps, thinning by a factor of one tenth to produce a thousands samples. Both Bayesian methods estimate θ with *maximum a posteriori* estimates using the final hundred samples. Note that the added complexity of the AGFH model incurs a computational cost resulting in runtimes approximately 60 times longer than the HB sampler. Additionally, two frequentist estimators are used: first, a linear regression of $\mathbf{Y} \sim \mathbf{X}\beta$ where $\hat{\theta}_i = x_i \hat{\beta}$ and $\hat{\beta}$ is an OLS estimate, referred to as LM in tables and figures. Second, the traditional EBLUP predictor discussed in Section 1 with the adjusted REML estimate of λ from Yoshimori and Lahiri^[30], referred to as Adj. REML. The same analysis was also conducted using REML and adjusted profile likelihood estimates of λ , and produced nearly identical results. Specific details are available in the supplementary materials.

The following two examples highlight the impact the relative size each source of variability—sampling variability through D and latent variability through λ —has on each estimation methodology. In both examples, $M = 100$ small area units are considered. Accurate estimation of $\theta = \{\theta_i\}_{i=1}^{100}$ is most important and will be stressed throughout. Additional figures and results are presented in the supplementary materials. Our simulation studies support the general intuition that the presence of $g(\cdot)$ in AGFH's distribution for θ_i (as opposed to the lack of such flexibility in HB or frequentist methods) allows the learned distribution of sampling errors to improve the sampling of θ_i . When the distributional assumptions corresponding to hierarchical Bayesian or empirical Bayes methods are satisfied, then these methods may marginally outperform the AGFH, but even under such circumstances the AGFH remains competitive. This exhibits the robustness of the AGFH approach.

5.1. High Sampling Variability

We first consider the case where the sampling variability $D \in \{0.1, 0.01\}$ dominates the latent effect variability $\lambda = 0.5$. In this case, the sampling errors provide ample signal for the AGFH to estimate their (potentially non-Normal) distribution, and Figure 1 depicts typical cases. In all four scenarios, the estimated sampling error density (green) captures the true error distribution.

An accurate representation of the sampling errors by the AGFH model appears to engender more accurate estimation of θ . Tables 1 and 2 summarizes the mean squared prediction error (MSPE) for each estimation methodology's estimation of θ , averaged across 30 independent repeated analyses (standard deviations in parenthesis). The former pertains to a scenario in which $p = 1$, $\beta \equiv 1$, and the latter $p = 3$, $\beta \sim \text{Unif}(-5, 5)$.

For the most extreme departure from normality (the two bimodal Beta scenarios, which still respect the moment conditions), AGFH is the most competitive method. Otherwise, extant methods are more performant, most notably the HB approach. A more detailed representation of the relationship among MSPEs is shown in Figure 2 for the $D = 0.01$, $\lambda = 0.5$ case (additional cases in the supplementary materials). The AGFH model outperforms the other methods in almost every repeated analysis under symmetric and asymmetric Beta errors.

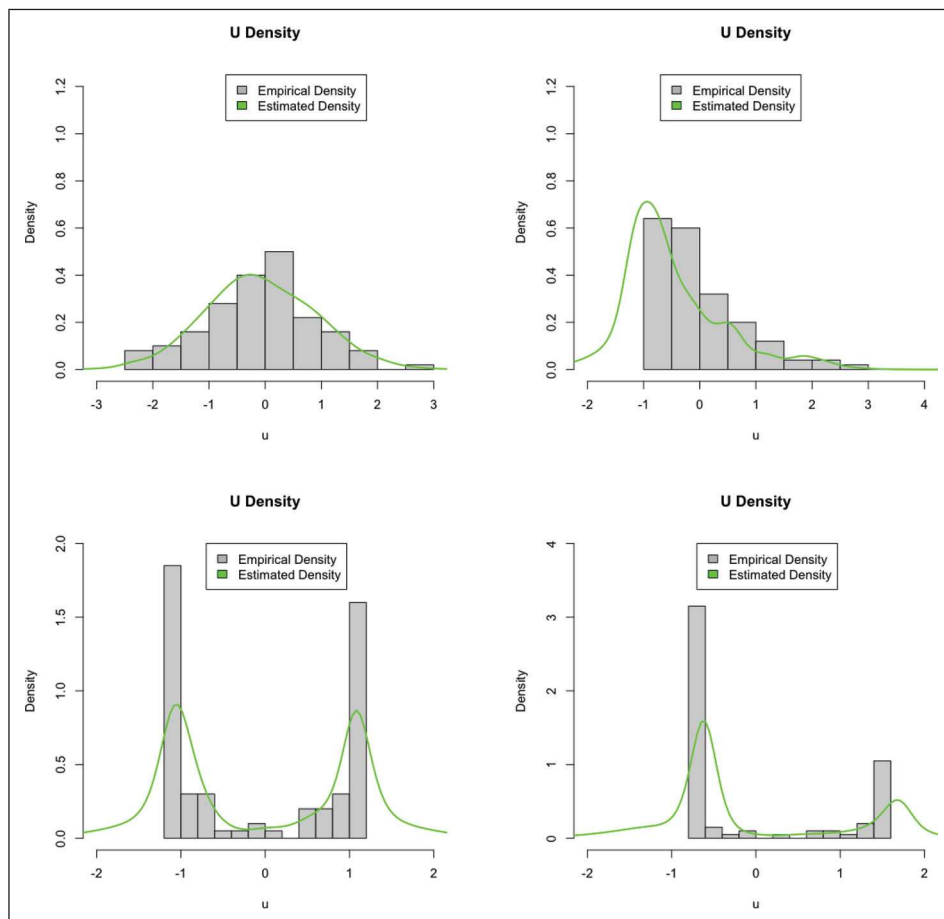


Figure 1. Empirical Sampling Error Distributions and their Estimates by AGFH. Here, $p = 3, \beta \sim \text{Unif}(-5, 5)$, $D = 0.01$, and $\lambda = 0.5$. From Top Left: Normal Errors, Gamma Errors, Symmetric Beta Errors, Asymmetric Beta Errors.

It is clear from Figure 1 that the AGFH model can accurately capture the shape of the sampling distribution. This is also reflected in its generally superior or competitive performance in terms of the MSPE, as noted in Tables 1 and 2. Under the Normality assumption (traditional Fay-Herriot model) in Table 1, the AGFH model's performance is almost identical to that of the traditional EBLUP-based approach, while the hierarchical Bayesian approach performs marginally better. When the sampling distribution is highly non-Normal, as in the case of the Beta-distribution based cases under study, the AGFH clearly dominates. In general, using ordinary linear regression does not result in good prediction in small area models as demonstrated by the third column in Tables 1 and 2. However, when the linking model's mean ($x_i^T \beta$) dominates as in the case of Table 2, and the sampling distribution is Gaussian, then linear regression can be competitive.

Table 1. MSPE of $\hat{\theta}_i$ for Each Method of Estimation across Four Data-generating Scenarios: Normal Errors, Gamma Errors, Beta (symmetric), and Beta (Asymmetric). Here, $p = 1$, $\beta \equiv 1$, and $\lambda = 0.5$.

	D	AGFH	HB	LM	Adj REML
Normal	0.1	0.63 (0.15)	0.60 (0.14)	0.70 (0.20)	0.63 (0.17)
	0.01	2.13 (2.45)	1.62 (1.47)	2.59 (1.95)	1.96 (1.60)
Gamma($\frac{1}{2}, 10$)	0.1	0.679 (0.55)	0.77 (0.55)	0.683 (0.21)	0.90 (0.72)
	0.01	3.46 (6.70)	2.30 (3.50)	2.44 (1.79)	4.70 (7.11)
Beta($\frac{1}{8}, \frac{1}{8}$)	0.1	0.50 (0.17)	0.59 (0.23)	0.71 (0.22)	0.59 (0.22)
	0.01	1.47 (1.53)	1.67 (1.75)	2.83 (2.04)	1.79 (1.99)
Beta($\frac{1}{12}, \frac{1}{6}$)	0.1	0.47 (0.10)	0.60 (0.14)	0.69 (0.18)	0.61 (0.17)
	0.01	1.19 (0.84)	1.57 (1.28)	2.31 (1.74)	1.80 (1.68)

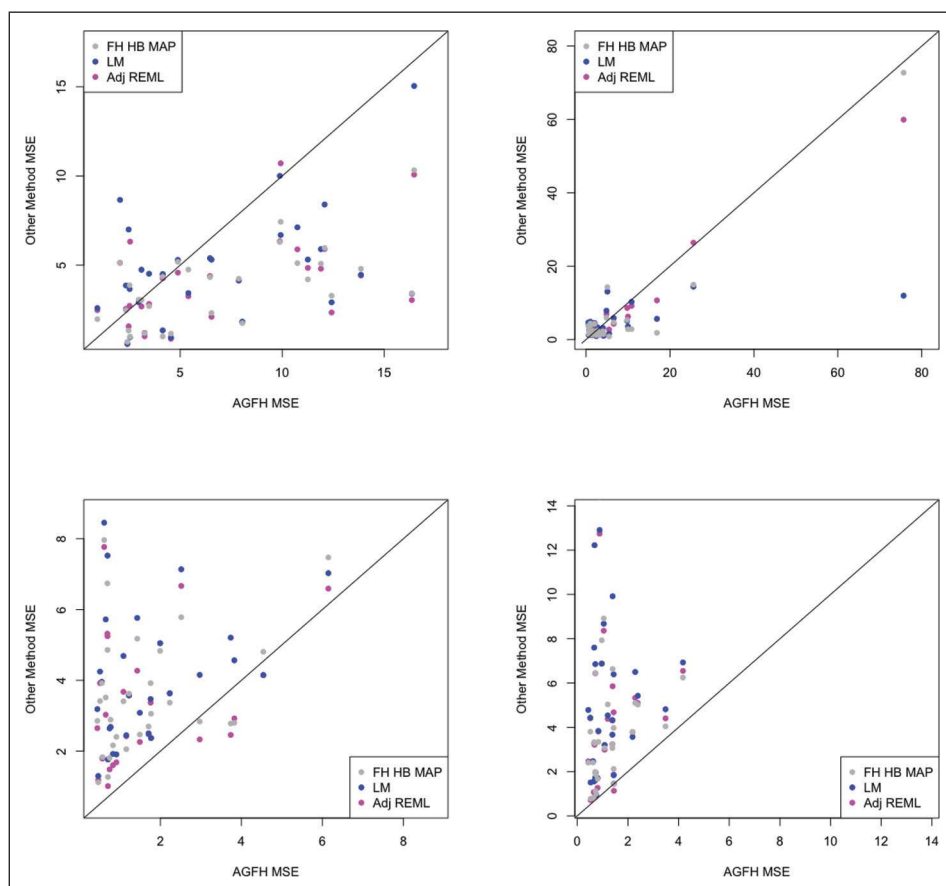


Figure 2. Comparison Between AGFH MSPEs and those of the Other Estimating Values, with a Diagonal Line Drawn along Parity. Points Lying above the Line Indicate AGFH is Favorable; Points Below Parity Favor Other Methods. Here, $p = 3$, $\beta \sim \text{Unif}(-5, 5)$, $D = 0.01$, and $\lambda = 0.5$. From Top Left: Normal Errors, Gamma Errors, Symmetric Beta Errors, Asymmetric Beta Errors.

Table 2. MSPE of $\hat{\theta}_i$ for Each Method of Estimation across Four Data-Generating Scenarios: Normal Errors, Gamma Errors, Beta (Symmetric), and Beta (Asymmetric). Here, $p = 3$, $\beta \sim \text{Unif}(-5, 5)$, and $\lambda = 0.5$.

	D	AGFH	HB	LM	Adj REML
Normal	0.1	1.00 (0.28)	0.77 (0.27)	0.88 (0.31)	0.80 (0.34)
	0.01	6.83 (4.57)	3.71 (2.17)	4.74 (3.08)	3.97 (2.37)
Gamma($\frac{1}{2}, 10$)	0.1	0.97 (0.95)	0.95 (0.84)	0.89 (0.36)	1.04 (0.94)
	0.01	4.59 (5.54)	3.46 (3.37)	4.11 (3.43)	4.54 (5.26)
Beta($\frac{1}{8}, \frac{1}{8}$)	0.1	0.65 (0.15)	0.75 (0.18)	0.83 (0.20)	0.74 (0.17)
	0.01	1.62 (1.38)	3.59 (1.71)	4.19 (2.14)	3.38 (1.69)
Beta($\frac{1}{12}, \frac{1}{6}$)	0.1	0.61 (0.14)	0.72 (0.22)	0.82 (0.26)	0.71 (0.23)
	0.01	1.25 (0.87)	3.94 (2.87)	4.81 (3.18)	3.88 (2.60)

Table 3. MSPE of $\hat{\theta}_i$ for Each Method of Estimation across Four Data-Generating Scenarios: Normal Errors, Gamma Errors, Beta (Symmetric), and Beta (Asymmetric). Here, $p = 1$, $\beta \equiv 1$, and $\lambda = 5$.

	D	AGFH	HB	LM	Adj REML
Normal	0.1	4.14 (0.50)	3.83 (0.45)	5.15 (0.53)	3.51 (0.40)
	0.01	6.67 (2.71)	6.00 (1.55)	7.04 (2.02)	6.28 (1.66)
Gamma($\frac{1}{2}, 10$)	0.1	3.71 (0.79)	4.16 (0.82)	5.07 (0.74)	3.84 (0.89)
	0.01	7.73 (6.45)	6.91 (4.35)	6.83 (2.12)	9.01 (7.24)
Beta($\frac{1}{8}, \frac{1}{8}$)	0.1	4.08 (0.71)	3.65 (0.70)	4.94 (0.71)	3.37 (0.54)
	0.01	5.67 (1.76)	5.87 (2.00)	7.07 (2.21)	5.90 (2.24)
Beta($\frac{1}{12}, \frac{1}{6}$)	0.1	4.05 (0.7)	3.79 (0.55)	5.25 (0.70)	3.55 (0.47)
	0.01	5.56 (0.95)	6.08 (1.36)	6.87 (1.82)	6.14 (1.67)

5.2. High Latent Variability

Here, we consider the case where $\lambda = 5$, thus there is considerable additional variability in the unobserved θ_i values. We leave the linking mean, $x_i^\top \beta$, and the sampling-level precisions, D_i 's, unchanged. Thus in this framework, there is considerable additional variability, and the signal-to-noise ratio is lower (greater λ relative the variability of the sampling distribution). In this case, the AGFH performs the best among the competing methods when $D = 0.01$ and the sampling errors are highly non-normal (see Tables 3, 4). When sampling-level variances are low and the precision is set at $D = 0.1$, the traditional EBLUP-based approach eclipses the other techniques in most cases.

6. Concluding Remarks

In this article, we present a new semiparametric extension of the Fay-Herriot model, termed the *agnostic Fay-Herriot model* (AGFH). Here, the sampling-level model is expressed in terms of an unknown general function $g(\cdot)$. Thus in principle, the AGFH model can express **any** distribution in the sampling model since the choice of $g(\cdot)$ is extremely broad. We proposed a Bayesian modelling scheme for the AGFH where the unknown function $g(\cdot)$ is assigned a Gaussian Process prior. This

Table 4. MSPE of $\hat{\theta}_i$ for Each Method of Estimation across Four Data-Generating Scenarios: Normal Errors, Gamma Errors, Beta (Symmetric), and Beta (Asymmetric). Here, $p = 3$, $\beta \sim \text{Unif}(-5, 5)$, and $\lambda = 5$.

	D	AGFH	HB	LM	Adj REML
Normal	0.1	4.48 (0.76)	4.43 (0.67)	5.13 (0.73)	3.89 (0.66)
	0.01	10.48 (4.01)	7.75 (2.64)	8.79 (3.10)	8.02 (3.43)
Gamma($\frac{1}{2}, 10$)	0.1	4.29 (1.2)	4.63 (0.95)	5.28 (0.68)	4.11 (1.06)
	0.01	10.43 (11.26)	9.21 (8.13)	8.95 (3.57)	10.42 (9.37)
Beta($\frac{1}{8}, \frac{1}{8}$)	0.1	4.78 (0.66)	4.45 (0.65)	5.20 (0.60)	3.73 (0.51)
	0.01	6.30 (1.14)	7.54 (1.81)	8.30 (2.01)	7.40 (1.74)
Beta($\frac{1}{12}, \frac{1}{6}$)	0.1	4.47 (0.94)	4.20 (0.97)	5.17 (0.76)	3.64 (0.67)
	0.01	6.43 (1.57)	7.46 (2.34)	8.15 (2.57)	7.14 (2.30)

choice naturally restricts the scope of the function $g(\cdot)$. Using a Metropolis within Gibbs sampling Markov Chain Monte Carlo scheme, we studied the performance of the AGFH model in relation to three other models. Among these three, we have the hierarchical Bayesian model for the traditional Fay-Herriot model (1.1)(1.2) as one of the competitors, thus studying the performance of hierarchical Bayesian models as well in this article. The other competitions are the popular EBLUP-based prediction method, and the use of linear regression-based prediction. We constrain the modelling and simulation frameworks to adhere to the moment conditions (1.3) and (1.4) so that the EBLUP-based and regression-based predictions remain plausible.

It can be seen that in general the AGFH method performs very well. It clearly outperforms rival techniques when the sampling-level precision is small (i.e., variance is high) and when the sampling distribution is bounded. The hierarchical Bayesian method also generally performs well. When the signal-to-noise ratio is weaker, the EBLUP-based method is often very competitive. The linear regression-based method is generally not very competitive unless the mean of the linking model, $x_i^\top \beta$, dominated its variance. If the moment conditions (1.3) and (1.4) are not imposed, the EBLUP-based and regression-based method are not competitive at all.

Overall, our simulation results show that when the moment conditions (1.3) and (1.4) are imposed, the hierarchical Bayesian technique and even the empirical Bayesian technique can yield very reasonable predictions even from misspecified models. This suggests strong robustness properties of these techniques. While the AGFH framework often produces better numeric results, it also requires considerable additional computations. Our additional simulations studies suggest that as sample size M increases, the performance of the AGFH model improves, as expected. Additional details, including a case study application, may be found in Thompson^[26], and the R-package titled `agfh` linked to this work is also available Thompson and Chatterjee^[27].

Instead of generalizing the sampling model, we could have taken that as some known parametric distribution and considered the linking model to be completely unknown and modeled that using the AGFH. The modelling strategy for this case is similar to the one presented above, involving a Gaussian Process prior and Bayesian techniques. We will report results for this case in a future article. However, the case where both the sampling distribution and the linking distribution are unknown is more complicated. There, identifiability conditions and other technical restrictions need to be imposed. In general, the Bayesian framework we use with the AGFH model can handle known constraints on the sampling and linking models, and can be extremely flexible owing to the versatility

of Gaussian Processes in approximating broad class of unknown functions. However, modelling flexibility should be balanced with computational requirements and data size availability.

In the simulation results presented above, we could have considerably improved the performance of the AGFH model by assuming certain properties of the sampling distribution. For example, when the distribution is Gamma, the performance of the AGFH model can be improved by assuming the knowledge that the sampling errors are bounded below. In order to use such knowledge for the hierarchical Bayesian or empirical Bayesian model, we would need the precise knowledge about the sampling distribution.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is partially supported by the US National Science Foundation (NSF) under grants 1939916, 1939956.

References

1. Chakraborty A, Datta GS, Mandal A. A two-component normal mixture alternative to the fay-herriot model. *Stat Transit new series* 2016; 17: 67–90.
2. Chatterjee S. On modifications to linking variance estimators in the Fay-Herriot model that induce robustness. *Stat Appl* 2018; 16: 289–303.
3. Chatterjee S, Lahiri P, Li H. Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models. *Ann Stat* 2008; 36: 1221–1245.
4. Chen S, Lahiri P, Rao JNK. Robust mean squared prediction error estimators of eblup of a small area total under the fay-herriot model. *Statistics Canada International Symposium Series: Proceedings* 2000. URL <https://www150.statcan.gc.ca/n1/en/catalogue/11-522-X200600110393>.
5. Das K, Jiang J, Rao JNK. Mean squared error of empirical predictor. *Ann Stat* 2004; 32: 818–840.
6. Datta GS, Hall P, Mandal A. Model selection by testing for the presence of small-area effects, and application to area-level data. *J Amer Stat Assoc* 2011; 106: 362–374.
7. Datta GS, Mandal A. Small area estimation with uncertain random effects. *J Amer Stat Assoc* 2015; 110: 1735–1744.
8. Datta GS, Rao JNK, Smith DD. On measuring the variability of small area estimators under a basic area level model. *Biometrika* 2005; 92: 183–196.
9. Fay III RE, Herriot RA. Estimates of income for small places: an application of James-Stein procedures to census data. *J Amer Stat Assoc* 1979; 74: 269–277.
10. Ghosh M, Myung J, Moura FA. Robust bayesian small area estimation. *Surv Meth* 2018; 44: 101–116.
11. Jiang J, Lahiri P. Mixed model prediction and small area estimation (with discussion). *Test* 2006; 15: 1–96.
12. Jiang J, Lahiri P, Wan SM. A unified jackknife theory for empirical best prediction with M-estimation. *Ann Stat* 2002; 30: 1782–1810.

13. Jiang J, Rao JS. Robust small area estimation: an overview. *Ann Rev Stat Appl* 2020; 7: 337–360.
14. Lahiri P, Rao J. Robust estimation of mean squared error of small area estimators. *J Amer Stat Assoc* 1995; 90: 758–766.
15. Li H, Lahiri P. An adjusted maximum likelihood method for solving small area estimation problems. *J Multivar Anal* 2010; 101: 882–892.
16. Molina I, Rao JNK, Datta GS. Small area estimation under a Fay–Herriot model with preliminary testing for the presence of random area effects. *Surv Meth* 2015; 41: 1–19.
17. Pfeiffermann D. New important developments in small area estimation. *Stat Sci* 2013; 28: 40–68.
18. Pfeiffermann D, Glickman H. Mean square error approximation in small area estimation by use of parametric and nonparametric bootstrap. *ASA Sec Surv Res Meth Proce* 2004: 4167–4178.
19. Polettini S. A generalised semiparametric bayesian fay–herriot model for small area estimation shrinking both means and variances. *Bayesian Anal* 2017; 12: 729–752.
20. Rao J. Inferential issues in small area estimation: some new developments. *Stat Transit* 2005; 7: 513–526.
21. Rao JNK. Inferential issues in model-based small area estimation: some new developments. *Stat Transit* 2015; 16: 491–510.
22. Rao JNK, Molina I. *Small area estimation*. John Wiley & Sons 2015.
23. Rasmussen CE, Williams CK. *Gaussian processes for machine learning*, volume 1. Springer 2006.
24. Salvati N, Tzavidis N, Pratesi M, Chambers R. Small area estimation via M-quantile geographically weighted regression. *Test* 2012; 21: 1–28.
25. Sugasawa S, Kubokawa T. Parametric transformed Fay–Herriot model for small area estimation. *J Multivar Anal* 2015; 139: 295–311.
26. Thompson M. *Gaussian processes in semi-parametric models*. PhD Thesis, University of Minnesota 2023.
27. Thompson M, Chatterjee S. AGFH: Agnostic Fay-Herriot model for small area statistics 2023. URL <https://cran.r-project.org/web/packages/agfh/index.html>.
28. Wanjoya A, Torelli N, Datta G. Small area estimation: an application of a flexible Fay-Herriot method. *J Agri Sci Tech* 2012; 14: 76–86.
29. Wu P, Jiang J. Robust estimation of mean squared prediction error in small-area estimation. *Canadian J Stat* 2021; 49: 362–396.
30. Yoshimori M, Lahiri P. A new adjusted maximum likelihood method for the Fay–Herriot small area model. *J Multivar Anal* 2014a; 124: 281–294.
31. Yoshimori M, Lahiri P. A second-order efficient empirical Bayes confidence interval. *Ann Stat* 2014b; 42: 1233–1261.