Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Performance of hierarchical Bayes small area estimators using noninformative and informative priors with an application to the Canadian Labor Force Survey

by Yong You and Keven Bosa

Release date: December 23, 2025

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, the Agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Survey Methodology, December 2025
Vol. 51, No. 2, pp. 425-448
Statistics Canada, Catalogue No. 12-001-X

Performance of hierarchical Bayes small area estimators using noninformative and informative priors with an application to the Canadian Labor Force Survey

Yong You and Keven Bosa[1]

Abstract

In this paper, we study the performance of hierarchical Bayes (HB) small area estimators using noninformative and informative priors. We apply the Bayesian models of You and Chapman (2006) and You (2021) to the Canadian Labor Force Survey (LFS) data and evaluate the impact of the priors on the HB estimators. A Bayesian model comparison and simulation study are also conducted. Our results indicate that a correct informative prior can lead to very good results, and noninformative priors can also perform very well. Incorrect informative priors can lead to poor results in terms of large bias and large coefficient of variation (CV). Noninformative priors are recommended in practice for HB small area estimation unless correctly specified informative priors are available. Informative priors are particularly useful when the number of small areas is relatively small.

Key Words: Bias; CPO; Fay-Herriot model; Gibbs sampling; Model selection; Posterior predictive distribution.

1. Introduction

Sample surveys, for most purposes, are usually designed to provide reliable direct estimates for large areas by using area-specific sample data. These direct estimates frequently fail to provide reliable estimates for small areas due to small sample sizes. Direct survey estimates for small areas often have unsuitably large standard errors. To gain precision and reliability, it is necessary to borrow strength from related areas thus increasing the effective sample size to construct indirect estimates for the small areas. Explicit model-based methods that use supplementary data such as census and administrative data associated with the small areas have been widely used in practice to obtain reliable model-based estimates. For overviews and appraisals of models for small area estimation, see Rao and Molina (2015). In this paper, we study the performance of hierarchical Bayes (HB) small area estimators using area level models.

The advantage of using area level models is that the area level models take into account the survey design through the use of direct survey estimates and related design-based variance estimates. Area level models can be used in practice whenever the direct survey estimates and area level auxiliary variables are available. Among the area level models, the Fay-Herriot model (Fay and Herriot, 1979) is a widely used area level model for small area estimation. The Fay-Herriot model has two components, namely, a sampling model for the direct survey estimates and a linking model for the small area parameters of interest. The sampling model assumes that given the area-specific sample size $n_i > 1,$ for ith area, there exists a direct survey estimator $y_i,$ which is usually design unbiased, for the small area parameter $\theta_i,$ such that

$$y_i = \theta_i + e_i,\ i = 1, \ldots, m, \tag{1.1}$$

1. Yong You and Keven Bosa, Economic Statistics Methods Division, Statistics Canada, Ottawa, Ontario, Canada, K1A 0T6. E-mails: yong.you@statcan.gc.ca and keven.bosa@statcan.gc.ca.

where e_i is the sampling error associated with the direct estimator y_i and m is the number of small areas. It is customary in practice to assume that the e_i's are independent normal random variables with mean $E(e_i) = 0$ and sampling variance $\text{Var}(e_i) = \sigma_i^2$. The linking model assumes that the small area parameter of interest θ_i is related to area level auxiliary variable $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$ through a linear regression model

$$\theta_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i, i = 1, \ldots, m, \tag{1.2}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of regression coefficients, and the v_i's are area-specific random effects assumed to be independent and identically distributed (iid) with $E(v_i) = 0$ and $\text{Var}(v_i) = \sigma_v^2$. The assumption of normality is generally also included, even though it is more difficult to justify the assumption. The variance σ_v^2 is called model variance, and it is unknown and needs to be estimated from the data. The area level random effects v_i capture the unstructured heterogeneity among areas that are not explained by the sampling variance σ_i^2. Combining models (1.1) and (1.2) leads to a linear mixed area level model given as

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i + e_i, i = 1, \ldots, m. \tag{1.3}$$

Model (1.3) involves both design-based random errors e_i and model-based random effects v_i. For the Fay-Herriot model, the sampling variance σ_i^2 is assumed to be known. In practice, we must estimate the sampling variance σ_i^2. In some applications, the direct sampling variance estimate is used directly in the Fay-Herriot model, but this estimate is typically unstable for small areas. A smoothing or modeling approach is often employed to obtain a more stable estimate of the sampling variance. In this paper, we consider a modeling approach for σ_i^2 using the direct sampling variance estimator under a HB framework as follows.

Let s_i^2 denote the direct estimator for the sampling variance σ_i^2. We consider a commonly used model for s_i^2 as $d_i s_i^2 \sim \sigma_i^2 \chi_{d_i}^2$, where $d_i = n_i - 1$ and n_i is the sample size for the ith area. For example, suppose we have n_i observations from small area i and these observations are iid $N(\mu_i, \sigma^2)$. Let y_i be the sample mean of the n_i observations. Then $y_i \sim N(\mu_i, \sigma_i^2)$ with $\sigma_i^2 = \sigma^2/n_i$. Then we can obtain an estimator of σ_i^2 as $s_i^2 = s^2/n_i$, where s^2 is the sample variance of the n_i observations. Furthermore, y_i and s_i^2 are independent and $(n_i - 1)\, s_i^2 \sim \sigma_i^2 \chi_{n_i-1}^2$. You and Chapman (2006) considered a HB approach and combined the sampling variance model $d_i s_i^2 \sim \sigma_i^2 \chi_{d_i}^2$ with the small area model (1.3) to construct an integrated model. The integrated model borrows strength from small area estimates and sampling variance estimates simultaneously. The integrated HB modeling approach with $d_i s_i^2 \sim \sigma_i^2 \chi_{d_i}^2$ has been widely used in practice, for example, see You (2008a, 2021), Dass, Maiti, Ren and Sinha (2012), Sugasawa, Tamae and Kubokawa (2017), Ghosh, Myung and Moura (2018), and Hidiroglou, Beaumont and Yung (2019), among others. An advantage of the HB approach is that it is straightforward, and the inferences for parameters θ_i are based on the posterior samples. The HB approach automatically takes account of the uncertainties associated with the unknown parameters, but it requires the specification of prior distributions on model parameters $\boldsymbol{\beta}$ and σ_v^2 in the Fay-Herriot model.

In this paper, we study HB small area estimators using noninformative and informative priors for $\boldsymbol{\beta}$ and σ_v^2 through real data analysis and a simulation study. In Bayesian analysis, prior specification is important

and quite challenging. Generally, the priors of model parameters could be informative, weakly informative, and noninformative for the HB models. In this paper, we evaluate the impact of the choice of priors on small area estimates, particularly for the specification of the prior for the regression parameter $\boldsymbol{\beta}$ in the model. The process of choosing a prior distribution may involve certain steps or criteria, for example, using the context and purpose of the data analysis, knowledge of small area survey data, and criteria such as coherence, that is, the prior should be consistent with the logic and structure of the HB model and parameter space. The prior should be conjugate, that is, the prior should be mathematically compatible with the likelihood function of the model; compatible with the data so that it doesn't contradict or overwhelm the evidence from the sampling data. Also, the prior should reflect prior knowledge and beliefs about the parameter, based on relevant sources, census and administration data. In practice, various information and knowledge may be used to construct informative priors, particularly for the prior mean specification for the regression parameter. For example, Sakshaug, Wiśniowski, Ruiz and Blom (2019) and Wiśniowski, Sakshaug, Ruiz and Blom (2020) demonstrate that informative priors based on nonprobability data can lead to reductions in variances and mean squared errors for linear model coefficients. Assuming an ignorable probability sampling design, Sakshaug et al. (2019) and Wiśniowski et al. (2020) proposed and assessed through simulation studies a Bayesian approach to integrating data from both samples for the estimation of model parameters by specifying informative prior distributions for the model parameters: non-probability sample data are used to determine the prior mean, but data from both samples are used to determine the prior variance; i.e., the estimated bias of the non-probability sample estimator is used to inflate the prior variance. In general, the specification of prior distribution for model parameters is more from the application perspective, and in many cases depends on the specific application and the actual available data. For some examples and guidance, we cite some references such as Consonni, Fouskakis, Liseo and Ntzoufras (2018), Grzenda (2016), Lemoine (2019), Mikkola, Martin, Chandramouli, Hartmann, Abril Pla, Thomas, Pesonen, Corander, Vehtari, Kaski, Bürkner and Klami (2024) and Zondervan-Zwijnenburg, Peeters, Depaoli and Van de Schoot (2017).

For the specification of priors for variance components in the HB small area model, an inverse gamma (IG) prior is a proper prior and conditionally conjugate for the variance components. An IG prior is widely used in Bayesian literature (e.g., Gelman, Carlin, Stern and Rubin, 2004) and Bayesian software packages (e.g., WinBUGS, Lunn, Thomas, Best and Spiegelhalter, 2000). You (2023) showed the effective use of an IG prior in the HB small area estimation for the LFS application. Alternatively, flat priors can be used for the model variances. A flat prior is usually used as a noninformative prior in the literature (e.g., Gelman, 2006).

In this paper, we study the HB small area estimates based on different prior specifications, particularly for the prior of the regression parameter in the HB small area model. The paper is organized as follows: In Section 2, we present the HB small area models and the corresponding Gibbs sampling inference procedures. In Section 3, we compare the HB estimates based on different prior specifications through the application of real data analysis. In Section 4, we provide Bayesian model selection and model check analysis.

In Section 5, we conduct a simulation study to evaluate the HB estimators based on different prior specifications. And in Section 6, we compare and study the HB estimates for grouped small areas using informative and flat priors. Finally in Section 7, we offer some concluding remarks.

2. Hierarchical Bayes model specification and inference

In this section, we present the HB approach using the Gibbs sampling method for the Fay-Herriot model when the sampling variances are estimated. In particular, we consider two HB models for small area estimation using noninformative and informative priors, namely, the models proposed by You and Chapman (2006) and You (2021). The HB models with prior specifications are as follows.

HB Model 1: You-Chapman model (You and Chapman, 2006), denoted as YCM:

- $y_i \mid \theta_i, \sigma_i^2 \sim N(\theta_i, \sigma_i^2),\ i = 1, \ldots, m;$
- $d_i s_i^2 \mid \sigma_i^2 \sim \sigma_i^2 \chi_{d_i}^2,\ d_i = n_i - 1,\ i = 1, \ldots, m;$
- $\theta_i \mid \boldsymbol{\beta}, \sigma_v^2 \sim N(\mathbf{x}_i'\boldsymbol{\beta}, \sigma_v^2),\ i = 1, \ldots, m;$
- Prior for regression parameter $\boldsymbol{\beta}$:
  (1) A flat prior for $\boldsymbol{\beta}$: $\pi(\boldsymbol{\beta}) \propto 1$;
  (2) A normal informative prior for $\boldsymbol{\beta}$: $\pi(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0)$, where $\boldsymbol{\beta}_0$ is known and $\boldsymbol{\Phi}_0$ is a pre-specified variance covariance matrix;
- Prior distributions for variance components σ_v^2 and σ_i^2: inverse gamma priors are $\pi(\sigma_v^2) \sim \text{IG}(a_v, b_v)$, $\pi(\sigma_i^2) \sim \text{IG}(a_i, b_i)$, where a_v, b_v and a_i, b_i are pre-specified constants. In particular, a_i, b_i $(0 \le i \le m)$ are chosen to be very small constants, say, 0.00001, to reflect a vague knowledge of σ_i^2, as in You and Chapman (2006).

We are mainly interested in the estimation of the small area parameter θ_i. We use the Gibbs sampling approach to obtain the posterior estimate of θ_i. Let $\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^m \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \left(\sum_{i=1}^m \mathbf{x}_i \theta_i\right)$, and $\mathbf{V}_\beta = \sigma_v^2 \left(\sum_{i=1}^m \mathbf{x}_i \mathbf{x}_i'\right)^{-1}$, $\boldsymbol{\theta}' = (\theta_1, \ldots, \theta_m)$, then the Gibbs sampling full conditional distributions corresponding to the HB model 1 are given as:

- $\left[\theta_i \mid y_i, \boldsymbol{\beta}, \sigma_v^2\right] \sim \text{N}\left(\gamma_i y_i + (1-\gamma_i)\, \mathbf{x}_i'\boldsymbol{\beta}, \gamma_i \sigma_i^2\right)$, where $\gamma_i = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_i^2}$, for $i = 1, \ldots, m;$
- We have the following conditional distributions corresponding to the prior specification for $\boldsymbol{\beta}$:
  (1) $\left[\boldsymbol{\beta} \mid y_i, \boldsymbol{\theta}, \sigma_v^2\right] \sim N\left(\hat{\boldsymbol{\beta}}, \mathbf{V}_\beta\right)$, if $\pi(\boldsymbol{\beta}) \propto 1$;
  (2) $\left[\boldsymbol{\beta} \mid y_i, \boldsymbol{\theta}, \sigma_v^2\right] \sim N\left(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Psi}_\beta\right)$, if $\pi(\boldsymbol{\beta}) \sim N\left(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0\right)$, where $\tilde{\boldsymbol{\beta}} = \boldsymbol{\Psi}_\beta \left(\boldsymbol{\Phi}_0^{-1}\boldsymbol{\beta}_0 + \mathbf{V}_\beta^{-1}\hat{\boldsymbol{\beta}}\right)$ and $\boldsymbol{\Psi}_\beta = \left(\boldsymbol{\Phi}_0^{-1} + \mathbf{V}_\beta^{-1}\right)^{-1}$;
- $\left[\sigma_v^2 \mid y_i, \boldsymbol{\theta}, \boldsymbol{\beta}\right] \sim \text{IG}\left(a_v + \frac{m}{2}, b_v + \frac{1}{2}\sum_{i=1}^m \left(\theta_i - \mathbf{x}_i'\boldsymbol{\beta}\right)^2\right);$

- $\left[\sigma_i^2 \mid y_i, \boldsymbol{\theta}, s_i^2\right] \sim \text{IG}\left(a_i + \frac{n_i}{2}, b_i + \frac{1}{2}\left((y_i - \theta_i)^2 + (n_i - 1)\, s_i^2\right)\right)$.

The second HB model is similar to the YCM but with a log linear model for the sampling variance (You, 2021). We now present the model with prior specification as follows:

HB Model 2: You (2021) log-linear model, denoted as YLLM:

- $y_i \mid \theta_i, \sigma_i^2 \sim N\left(\theta_i, \sigma_i^2\right), i = 1, \ldots, m$;
- $d_i s_i^2 \mid \sigma_i^2 \sim \sigma_i^2 \chi_{d_i}^2,\ d_i = n_i - 1, i = 1, \ldots, m$;
- $\theta_i \mid \boldsymbol{\beta}, \sigma_v^2 \sim N\left(\mathbf{x}_i'\boldsymbol{\beta}, \sigma_v^2\right), i = 1, \ldots, m$;
- $\log(\sigma_i^2) \mid \boldsymbol{\delta}, \tau^2 \sim N\left(z_i'\delta, \tau^2\right),\ \boldsymbol{\delta}' = (\delta_1, \delta_2),\ \mathbf{z}_i' = (1, \log(n_i)), i = 1, \ldots, m$;
- Prior for regression parameter $\boldsymbol{\beta}$:
  (1) A flat prior for $\boldsymbol{\beta}$: $\pi(\boldsymbol{\beta}) \propto 1$;
  (2) A normal informative prior for $\boldsymbol{\beta}$: $\pi(\boldsymbol{\beta}) \sim N\left(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0\right)$, where $\boldsymbol{\beta}_0$ is known and $\boldsymbol{\Phi}_0$ is a pre-specified variance covariance matrix;
- Prior distribution for model variance σ_v^2: $\pi\left(\sigma_v^2\right) \sim \text{IG}\left(a_v, b_v\right)$, where a_v, b_v are pre-specified constants;
- Prior specification for hyperparameter $\boldsymbol{\delta}$ and τ^2: $\pi(\boldsymbol{\delta}) \propto 1$, $\pi\left(\tau^2\right) \sim \text{IG}\left(a_\tau, b_\tau\right)$, and a_τ, b_τ are set to be 0.00001.

The full conditional distributions and sampling procedure for the YLLM model are given as follows:

- $\left[\theta_i \mid y_i, \boldsymbol{\beta}, \sigma_i^2, \sigma_v^2\right] \sim N\left(\gamma_i y_i + (1 - \gamma_i)\mathbf{x}_i'\boldsymbol{\beta}, \gamma_i \sigma_i^2\right)$, where $\gamma_i = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_i^2}$, $i = 1, \ldots, m$;
- For $\boldsymbol{\beta}$, the following are the conditional distributions corresponding to the priors on $\boldsymbol{\beta}$:
  (1) $\left[\boldsymbol{\beta} \mid y_i, \boldsymbol{\theta}, \sigma_v^2\right] \sim N\left(\hat{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}}\right)$, if $\pi(\boldsymbol{\beta}) \propto 1$;
  (2) $\left[\boldsymbol{\beta} \mid y_i, \boldsymbol{\theta}, \sigma_v^2\right] \sim N\left(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Psi}_{\boldsymbol{\beta}}\right)$, if $\pi(\boldsymbol{\beta}) \sim N\left(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0\right)$, where $\tilde{\boldsymbol{\beta}} = \boldsymbol{\Psi}_{\boldsymbol{\beta}}\left(\boldsymbol{\Phi}_0^{-1}\boldsymbol{\beta}_0 + \mathbf{V}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}}\right)$ and $\boldsymbol{\Psi}_{\boldsymbol{\beta}} = \left(\boldsymbol{\Phi}_0^{-1} + \mathbf{V}_\beta^{-1}\right)^{-1}$;
- $\left[\sigma_v^2 \mid y_i, \boldsymbol{\theta}, \boldsymbol{\beta}, \sigma_i^2\right] \sim \text{IG}\left(a_v + \frac{m}{2}, b_v + \frac{1}{2}\sum_{i=1}^m (\theta_i - \mathbf{x}_i'\boldsymbol{\beta})^2\right)$;
- $\left[\sigma_i^2 \mid y_i, \boldsymbol{\theta}, \boldsymbol{\beta}, \sigma_v^2, \delta, \tau^2\right] \propto f\left(\sigma_i^2\right) \cdot h\left(\sigma_i^2\right)$, where $f\left(\sigma_i^2\right)$ and $h\left(\sigma_i^2\right)$ are $f\left(\sigma_i^2\right) \sim \text{IG}\left(\frac{d_i + 1}{2}, \frac{(y_i - \theta_i)^2 + d_i s_i^2}{2}\right)$, and $h\left(\sigma_i^2\right) = \exp\left(-\frac{(\log(\sigma_i^2) - z_i'\delta)^2}{2\tau^2}\right)$;
- $\left[\delta \mid y_i, \boldsymbol{\theta}, \boldsymbol{\beta}, \sigma_i^2, \sigma_v^2, \tau^2\right] \sim N_2\left(\left(\sum_{i=1}^m \mathbf{z}_i\mathbf{z}_i'\right)^{-1}\left(\sum_{i=1}^m \mathbf{z}_i \log(\sigma_i^2)\right), \tau^2\left(\sum_{i=1}^m \mathbf{z}_i\mathbf{z}_i'\right)^{-1}\right)$;
- $\left[\tau^2 \mid y_i, \boldsymbol{\theta}, \boldsymbol{\beta}, \sigma_i^2, \sigma_v^2, \boldsymbol{\delta}\right] \sim \text{IG}\left(a_\tau + \frac{m}{2}, b_\tau + \frac{1}{2}\sum_{i=1}^m \left(\log(\sigma_i^2) - \mathbf{z}_i'\boldsymbol{\delta}\right)^2\right)$.

As in You (2021), we use the Metropolis-Hastings (MH) rejection step (Chip and Greenberg, 1995) to update σ_i^2 in the Gibbs sampling procedure. At the iteration *k* in the Gibbs sampling procedure:

(1) Draw σ_i^{2*} from IG$\left(\frac{d_i+1}{2}, \frac{(y_i-\theta_i)^2+d_i s_i^2}{2}\right)$;

(2) Compute the acceptance probability $\alpha\left(\sigma_i^{2*}, \sigma_i^{2(k)}\right) = \min\left\{h\left(\sigma_i^{2*}\right)/h\left(\sigma_i^{2(k)}\right), 1\right\}$;

(3) Generate u from Uniform(0,1), if $u < \alpha\left(\sigma_i^{2*}, \sigma_i^{2(k)}\right)$, the candidate σ_i^{2*} is accepted, $\sigma_i^{2(k+1)} = \sigma_i^{2*}$; otherwise σ_i^{2*} is rejected, and set $\sigma_i^{2(k+1)} = \sigma_i^{2(k)}$.

The Metropolis-Hastings rejection step is quite efficient in the Gibbs sampling procedure, and there is no computational problem to update the σ_i^2. We monitor the Gibbs sampling procedure by checking the convergence and the MH rejection rate. The average rejection rate is quite low in our application as shown in Section 3. Parker, Holan and Janicki (2023) used a very similar approach that models the variances on the log scale by using a multivariate log-gamma distribution (Bradley, Holan and Wikle, 2018, 2020) to bypass the need for the MH sampling steps for all the variance components σ_i^2. Parker, Holan and Janicki (2023) completed their model by setting their model parameters to reflect relatively vague priors. It might be useful and interesting to investigate how informative and noninformative priors can be constructed using the multivariate log-gamma distribution. Other works that are closely related to the sampling variance modeling include, for example, Maiti, Ren and Sinha (2014), Sugasawa, Tamae and Kubokawa (2017).

We are interested in the estimation of the small area parameters θ_i and the corresponding coefficient of variations (CV) under different priors on the regression parameter $\boldsymbol{\beta}$ and model variance σ_v^2. In the next section, we will compare the HB estimates through the analysis of the Canadian LFS unemployment rate data.

3. LFS application

In this section, we apply the two HB models presented in Section 2 to the LFS unemployment rate estimation to evaluate the prior specifications. The Canadian LFS produces monthly estimates of unemployment rate at the national and provincial levels across Canada. The LFS also releases unemployment estimates for sub-provincial areas such as Census Metropolitan Areas (CMAs) and Census Agglomerations (CAs) across Canada. Details of the LFS methodology are given in the Methodology of the Canadian Labour Force Survey (2017). The estimated variances for the LFS are computed through the Rao-Wu bootstrap procedure. For some sub-provincial areas, the direct estimates are not reliable because the sample sizes in some areas are quite small. The small area estimation procedure, as applied to the LFS data, usually estimates unemployment rates for local sub-provincial areas such as CMA/CAs using small area models. These small area models and methods are discussed, for example, in Hidiroglou, Beaumont, and Yung (2019), Lesage, Beaumont and Bocci (2021), You, Rao and Gambino (2003), and You (2008a, 2021). We apply the HB models described in Section 2 to the 2011 monthly data of unemployment rate estimates at the CMA/CA level to demonstrate the use of noninformative and informative priors for the HB small area estimates. There are 65 CMA/CAs (areas) considered in our study. The sample size for the 65 CMA/CAs ranges from 58 to

3,773. There are 9 areas having sample size less than 120, and there are 17 areas having sample size larger than 1,000 including some large cities such as Toronto, Montreal, and Vancouver.

We use the 2011 monthly LFS unemployment rate data in our application study because a census was performed in May 2011. For May, we do have an estimate of the unemployment rate coming from the census which will be considered as our gold standard. The May 2011 small area estimates for many HB scenarios will be compared against census estimates in this section.

3.1 Prior specification

We first used the 2011 monthly LFS unemployment rate data based on 65 CMA/CAs across Canada to obtain the estimates of $\boldsymbol{\beta}$ and model variance σ_v^2. We apply Model 1, the You-Chapman (2006) model with noninformative priors, to the 2011 monthly unemployment rate data from January to December 2011, omitting the May 2011. Then we take the average of the posterior estimates and the posterior variance estimates of the model parameters using these 11 months of HB estimates to construct relatively stable estimates as the prior distribution for $\boldsymbol{\beta}$ and σ_v^2.

The priors are specified by the normal and inverse gamma (IG) distributions as follows:

- $\pi(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0)$, where $\boldsymbol{\beta}_0 = (0.0507, 0.4255)'$, and
$$\boldsymbol{\Phi}_0 = \begin{pmatrix} 2.070245e-05 & -0.0002 \\ -0.0002 & 0.00405 \end{pmatrix},$$
- $\pi(\sigma_v^2) \sim$ IG(17.42, 0.00272), then the prior mean of σ_v^2 is 0.0001662.

The prior specification obtained in such a way is for illustration purposes only, so as to have a prior for $\boldsymbol{\beta}$ as stable as possible and close to the true value of $\boldsymbol{\beta}$ in the model using the 2011 data. The prior $\pi(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0)$ is defined as the default informative prior or correct prior in our LFS application study. In practice, how to construct and define correct priors for the model parameters is actually not easy.

To compare the noninformative and informative priors, we consider the following four prior specifications and combinations:

- Prior case 1: Noninformative prior for $\boldsymbol{\beta}: \pi(\boldsymbol{\beta}) \propto 1$, and vague prior for $\sigma_v^2: \pi(\sigma_v^2) \sim$ IG(0.0001, 0.0001) by choosing $a_v = b_v = 0.0001$.
- Prior case 2: Noninformative prior for $\boldsymbol{\beta}: \pi(\boldsymbol{\beta}) \propto 1$, and informative prior for $\sigma_v^2: \pi(\sigma_v^2) \sim$ IG(17.42, 0.00272).
- Prior case 3: Informative prior for $\boldsymbol{\beta}: \pi(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0)$, and vague prior for $\sigma_v^2: \pi(\sigma_v^2) \sim$ IG(0.0001, 0.0001) by choosing $a_v = b_v = 0.0001$.
- Prior case 4: Informative prior for $\boldsymbol{\beta}: \pi(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0)$, and informative prior for $\sigma_v^2: \pi(\sigma_v^2) \sim$ IG(17.42, 0.00272).

In summary, priors for case 1 and case 2 are noninformative for $\boldsymbol{\beta}$, whereas priors for case 3 and case 4 are informative for $\boldsymbol{\beta}$. Note that for variance components, flat priors $\pi\left(\sigma_v^2\right) \propto 1$ and $\pi\left(\tau^2\right) \propto 1$ can also be used for the model variances σ_v^2 and τ^2 in the YCM and YLLM models. However, using flat prior on variance components in the YCM and YLLM models has no advantage over the use of IG prior with respect to the final HB estimates. The YCM and YLLM HB models using the IG prior can be very effective and have better model fit than the models using flat prior for the variance component, as shown in You (2023). Thus, in our study we use the vague IG prior IG(0.0001, 0.0001) and the informative IG prior for σ_v^2.

3.2 Application to LFS data

We applied the HB Model 1 (YCM) and Model 2 (YLLM) to the May 2011 LFS direct estimates of the unemployment rate for 65 CMA/CAs across Canada with the four different prior specifications for $\boldsymbol{\beta}$ and σ_v^2. A local area employment insurance monthly beneficiary rate is used as an auxiliary variable x_i in the HB models. More details of model specification based on the LFS data can be found, for example, in Hidiroglou, Beaumont, and Yung (2019), You (2021) and You and Hidiroglou (2023).

To implement the Gibbs sampling for both YCM and YLLM models, we have a "burn-in" length of $B = 2{,}000$ and a Gibbs sampling size of $G = 50{,}000$. For the YLLM, for various combinations of prior specifications, the average Metropolis-Hastings step rejection rate over 65 areas is around 8.12%. For the YLLM HB model, we take every 5th iteration after the "burn-in" period, thus we can reduce the auto-correlation which may result from the Metropolis-Hastings accept-rejection algorithm for σ_i^2 in the Gibbs sampling run. Therefore, for model YLLM, we have actual Gibbs sampling size of 10,000 to compute the posterior estimates. Convergence of the Gibbs sampling is monitored for the small area parameter θ_i, the regression parameter $\boldsymbol{\beta}$ and the variance component σ_v^2 in the model using the potential scale reduction factor (Gelman and Rubin, 1992; Gelman, Carlin, Stern and Rubin, 2004). We have computed the reduction factors for all the monitored parameters in the model in the Gibbs sampling. These factor values are all very close to 1 (less than 1.05), which suggests that the desired convergence for these parameters is achieved by the Gibbs sampler.

We compare the HB estimates with the May 2011 census value and compute the absolute relative error (ARE) and coefficient of variation (CV). The small area HB estimates are compared via the ARE of the direct and HB estimates with respect to the census estimates for each CMA/CA as follows:

$$\text{ARE}_i = \left| \frac{\theta_i^{\text{Census}} - \theta_i^{\text{Est}}}{\theta_i^{\text{Census}}} \right|,$$

where θ_i^{Est} is the direct or the HB estimate and θ_i^{Census} is the corresponding census value of the LFS unemployment rate. Then we obtain the average ARE over all the areas. Table 3.1 presents the results of the average ARE and average CV based on the YCM (Model 1) and Table 3.2 presents the results based on the YLLM (Model 2).

From Table 3.1, for the YCM model, the HB estimates have improved the direct estimates by reducing the ARE and CV. Using a noninformative prior for β, the ARE is 0.1348 under a vague prior for σ_v^2 and 0.1337 under an informative IG prior for σ_v^2. In addition, the corresponding CV is 0.1192 under a vague prior for σ_v^2 and 0.1221 under an informative IG prior for σ_v^2. Therefore, using an informative IG prior for σ_v^2 leads to a slightly smaller ARE but also leads to a larger CV. When using an informative normal prior for $\boldsymbol{\beta}$, the ARE is 0.1287 with CV equal to 0.1164 under a vague prior for σ_v^2, and the ARE is 0.1286 with CV equal to 0.1195 under an informative IG prior for σ_v^2. Thus, there is a small improvement of ARE when using the informative normal prior for $\boldsymbol{\beta}$. However, using the informative IG prior for σ_v^2 leads to a larger CV.

Table 3.1
Comparison of ARE and CV for May 2011 LFS using the YCM model

	Direct LFS	HB with Prior case 1	HB with Prior case 2	HB with Prior case 3	HB with Prior case 4
Average ARE	0.1739	0.1348	0.1337	0.1287	0.1286
Average CV	0.2117	0.1192	0.1221	0.1164	0.1195

Note: ARE = absolute relative error; CV = coefficient of variations; HB = hierarchical Bayes; LFS = Canadian Labor Force Survey; YCM = You-Chapman model.

From Table 3.2, the YLLM leads to better results than the YCM in terms of smaller ARE and CV for all the prior cases. Again, using the informative normal prior for $\boldsymbol{\beta}$ leads to smaller ARE for the HB small area estimates. Using the informative IG prior leads to slightly larger CV as the results in Table 3.1. From Table 3.1 and Table 3.2, the best result is obtained under YLLM using an informative normal prior for $\boldsymbol{\beta}$ and a vague prior for σ_v^2, with ARE equal to 0.1278 and a CV value of 0.1161.

Table 3.2
Comparison of ARE and CV for May 2011 LFS using the YLLM model

	Direct LFS	HB with Prior case 1	HB with Prior case 2	HB with Prior case 3	HB with Prior case 4
Average ARE	0.1739	0.1329	0.1325	0.1278	0.1277
Average CV	0.2117	0.1185	0.1217	0.1161	0.1192

Note: ARE = absolute relative error; CV = coefficient of variations; HB = hierarchical Bayes; LFS = Canadian Labor Force Survey; YLLM = You log-linear model.

Next, we will use the YLLM as the studying model and adjust the values for the informative normal prior to evaluate the corresponding impact on the HB small area estimates. We will use the vague prior IG(0.0001, 0.0001) for σ_v^2, as the informative IG prior IG(17.42, 0.00272) always leads to larger CVs for the HB estimators compared to the vague prior IG(0.0001, 0.0001). We use the following informative prior for $\boldsymbol{\beta}: \pi(\boldsymbol{\beta}) \sim N(a\boldsymbol{\beta}_0, b\boldsymbol{\Phi}_0)$, where a and b are constants set as a = 0.5, 1, 2, 3, 4 and b = 0.25, 1, 4, 9, 16, 25. When $a = b = 1$, we have the default informative prior $\pi(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0)$ for $\boldsymbol{\beta}$. When a and b take other values, we have other incorrect priors. We will evaluate the performance of HB small area estimates under the default and the different incorrect prior combinations. In particular, we enlarge the prior

variance for $\boldsymbol{\beta}$ by assigning larger values for b, so we can evaluate the performance of the HB estimator under different prior variances. When the prior variance is large enough, the HB estimates under the informative prior should be close to the HB estimates under the flat prior for $\boldsymbol{\beta}$. Table 3.3 presents the comparison of ARE under different priors using different values of a and b.

Table 3.3
Comparison of the average ARE and (CV) under different prior specifications

	$b = 0.25$	$b = 1$	$b = 4$	$b = 9$	$b = 16$	$b = 25$ (flat prior)
$a = 0.5$	0.1718 (0.2087)	0.1701 (0.1665)	0.1364 (0.1186)	0.1340 (0.1185)	0.1333 (0.1184)	0.1328 (0.1183)
$a = 1$	0.1267 (0.1146)	0.1278 (0.1161)	0.1321 (0.1181)	0.1326 (0.1182)	0.1327 (0.1182)	0.1326 (0.1182)
$a = 2$	0.1827 (0.1882)	0.1805 (0.1837)	0.1226 (0.1189)	0.1312 (0.1183)	0.1323 (0.1183)	0.1325 (0.1182)
$a = 3$	0.1835 (0.2014)	0.1815 (0.2051)	0.1279 (0.1224)	0.1286 (0.1185)	0.1318 (0.1184)	0.1322 (0.1183)
$a = 4$	0.1815 (0.2052)	0.1814 (0.2055)	0.1281 (0.1311)	0.1292 (0.1183)	0.1317 (0.1184)	0.1322 (0.1183)

Note: ARE = absolute relative error; CV = coefficient of variations.

At first, when $a = b = 1$, we have the default informative prior $\pi(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0)$ for $\boldsymbol{\beta}$. The ARE of the HB estimator is 0.1278 with average CV 0.1161. Under the noninformative prior for $\boldsymbol{\beta}$, the ARE is 0.1326 and CV is 0.1183 suggesting that the HB estimator performs slightly better using the default informative prior. If we keep $b = 1$ (variance unchanged) and adjust the values for a to 0.5 (decrease mean) or 2, 3 and 4 (increase mean), the ARE and the corresponding CV become large. For example, for $a = 2$, the ARE is equal to 0.1805 and the CV is 0.1837. Therefore, if we have wrong values for the prior mean, the HB estimator will have large ARE and large CV, and these wrong priors lead to worse HB results than the noninformative prior. If we increase the values for b from 1 to 9, 16 and 25, the ARE and CV values are all close to the results under the noninformative prior as expected. Thus, if we set a very large variance value for the prior variance, then the impact of the different prior mean value is negligible. As expected, the prior with very large variance is roughly equivalent to the noninformative flat prior for $\boldsymbol{\beta}$. The results in Table 3.3 indicate that correct specification of prior mean is very important. A wrong specification of the prior mean could lead to a large bias and CV. If we cannot ensure a correct mean value for the prior, a flat prior for $\boldsymbol{\beta}$ should be used.

4. Bayesian model selection and model check

In this section, we conduct Bayesian model selection and model check for the LFS application presented in Section 3 to evaluate the YLLM HB model with different priors. At first we use the so-called conditional predictive ordinate (CPO) to evaluate and select the best fit model with different prior specifications reported in Table 3.3. The CPOs are the observed likelihoods based on the cross-validation predictive density $f(y_i \mid y_{\text{obs}(i)})$ for each data point y_i, where $y_{\text{obs}(i)}$ denotes all observed data except y_i. We compute the CPO

values for each observed data point $y_{i,\text{obs}}$, that is, $\text{CPO}_i = f(y_{i,\text{obs}} \mid y_{\text{obs}(i)})$, and a larger value of CPO_i indicates a better model fit. For a more detailed formula on the CPO computation, see, for example, Chen, Shao and Ibrahim (2000), You and Rao (2000) and Rao and Molina (2015, page 346). For more applications of CPO, see for example, Gilks, Richardson and Spiegelhalter (1996), Molina, Nandram and Rao (2014) and You (2021). For Bayesian model choice between model A and model B, we can compute the CPOs for all data points under model A and model B, and obtain the ratio of CPO for each data point. If this ratio is greater than 1, then $y_{i,\text{obs}}$ supports model A. We used the YYLM model with the correct informative prior ($a = b = 1$) for $\boldsymbol{\beta}$ as the default model (model A), and computed the CPO ratios against the model with other prior specifications (different a and b values reported in Table 3.3). Then we count the number of the CPO ratios that are larger than 1. We have 65 CMA/CAs in our data analysis. The ratio number with more than 33 points indicates that the YLLM HB model with default correct prior provides better model fit.

Table 4.1 presents the #-number of CPO ratios larger than 1 over the 65 CMA/CAs for the YLLM with different prior specifications. Larger CPO ratio numbers (more than 33) will indicate that the correct prior specification provides better model fit than the wrong prior specifications. The results in Table 4.1 clearly shows that the YLLM with the default informative prior specification ($a = b = 1$) for $\boldsymbol{\beta}$ provides a much better model fit than the model with incorrect prior specifications, especially better than the model with the wrong prior mean specifications. For example, in column one ($b = 0.25$), for the YLLM with wrong prior mean specifications ($a = 0.5$, 2, 3 and 4), more than 50 out of 65 data points support the YLLM with a correct prior mean specification. For the model with a correct prior mean specification ($a = 1$), the model with a smaller prior variance ($b = 0.25$) provides a better model fit as expected, the corresponding #-number of CPO ratios is 28, which is the only number that is less than 33 reported in Table 4.1. We can also note from Table 4.1 that for the model with wrong prior mean specifications, when the prior variance is small, more data points support the model with a correct prior specification, as the number of ratios greater than 1 reported in columns 1 to 3 generally ranges from 40 to 50. When the prior variance is very large, for example, when $b = 25$, the number of CPO ratios greater than 1 is around 37 or 38, as shown in the last column of Table 4.1, which implies that the impact of a wrong prior mean becomes much smaller when the prior variance is very large. The model selection result in Table 4.1 shows that the data supports the models with correct prior mean specification, and the model selection result is very consistent with the results reported in Table 3.3.

Table 4.1
Summary of the # of CPO ratios larger than 1

	$b = 0.25$	$b = 1$	$b = 4$	$b = 9$	$b = 16$	$b = 25$ (flat prior)
$a = 0.5$	51	48	38	37	41	38
$a = 1$	28	-	35	37	36	38
$a = 2$	52	51	42	36	35	37
$a = 3$	52	52	50	42	38	37
$a = 4$	51	50	53	51	41	37

Note: CPO = conditional predictive ordinate.

Next, we compare the different prior specification models using Bayesian deviance information criterion (DIC) proposed by Spiegelhalter, Best, Carlin and van der Linde (2002). DIC is based on the deviance of the model $D(\theta)$, which is equal to minus twice the loglikelihood of the model. The DIC is usually computed as $\text{DIC} = D(\hat{\theta}) + 2\ P_D$, where $D(\hat{\theta})$ is the deviance of the model evaluated at the posterior mean of the model parameters, which summarizes the goodness of fit of the model, and P_D is the effective number of parameters, which captures the complexity of the model. P_D is defined as $P_D = \bar{D}(\theta) - D(\hat{\theta})$, and $\bar{D}(\theta)$ is the posterior mean of the deviance of the model. Thus, the DIC is defined as the summation of the goodness of fit of the model and the model complexity. Smaller values of DIC indicate a better model fit. Computation of DIC is relatively straightforward provided that the deviance $D(\theta)$ is available in a closed form, and P_D may be calculated after the Gibbs sampling run by taking the sample mean of the simulated values of $D(\theta)$ minus the plug-in estimate of the deviance $D(\hat{\theta})$. Table 4.2 presents the DIC values for the YLLM HB model with different prior specification combinations

Table 4.2
Comparison of DIC values for different prior specifications

	$b = 0.25$	$b = 1$	$b = 4$	$b = 9$	$b = 16$	$b = 25$ (flat prior)
$a = 0.5$	280.13	271.15	257.72	256.49	256.22	255.87
$a = 1$	254.26	254.52	255.37	255.45	255.69	255.83
$a = 2$	282.72	281.49	262.56	256.73	255.63	255.79
$a = 3$	285.67	285.91	285.33	264.01	257.39	256.21
$a = 4$	286.22	286.11	286.06	286.31	261.48	257.19

Note: DIC = deviance information criterion.

It is clear that the HB model with the default informative prior mean has the smallest DIC values, which clearly indicates that the YLLM model with the correct prior specification has better model fit. The models with correct prior specifications $a = 1$ and $b = 0.25$ or $b = 1$ perform the best as shown with the smallest DIC values. The models with wrong prior mean specifications ($a = 0.5, 2, 3, 4$) have large DIC values, which indicates relatively poor model fit, particularly when the prior variance is not large ($b = 0.25, 1, 4$). When the prior variance is large, for example, when $b = 16$ or 25, the DIC values become smaller and close to the DIC value of the model with correct prior specification as expected. The results in Table 4.2 show that the correct prior mean specification provides the best model fit, and when the prior variance is very large, the model fit is close to the model with correct prior or flat prior. The DIC comparison is consistent with the CPO comparison results in Table 4.1 and HB results reported in Table 3.3.

Finally, we evaluate the overall fit of the YLLM HB model with the correct prior specification using the posterior predictive p-value based on the posterior predictive distribution. The posterior predictive p-value is typically useful if we think of the current model as a plausible ending point with modifications to be made only if a substantial lack of fit is found, as noted in Sinharay and Stern (2003). Let y_{rep} denote the replicated observation under the model. The posterior predictive distribution of y_{rep} given the observed data y_{obs} is defined as

$$f(y_{\text{rep}} \mid y_{\text{obs}}) = \int f(y_{\text{rep}} \mid \theta) f(\theta \mid y_{\text{obs}}) d\theta.$$

In this approach, a test statistic $T(y, \theta)$ that depends on the data y and possibly the parameter θ can be defined and the observed value $T(y_{\text{obs}}, \theta \mid y_{\text{obs}})$ compared to the posterior predictive distribution of $T(y_{\text{rep}}, \theta \mid y_{\text{obs}})$ with any significant difference indicating a model failure. A lack of fit of the data with respect to the posterior predictive distribution can be measured by a p-value of the test quantity (Meng, 1994; Gelman, Meng and Stern, 1996). The posterior predictive p-value is defined as

$$p = P(T(y_{\text{rep}}, \theta) \geq T(y_{\text{obs}}, \theta) \mid y_{\text{obs}}).$$

This is a natural extension of the usual p-value in a Bayesian context. If a model fits the observed data, then the two values of the discrepancy measure are similar. In other words, if the given model adequately fits the observed data, then $T(y_{\text{obs}}, \theta \mid y_{\text{obs}})$ should be near the central part of the histogram of the $T(y_{\text{rep}}, \theta \mid y_{\text{obs}})$ values if y_{rep} is generated repeatedly from the posterior predictive distribution. Consequently, the posterior predictive p-value is expected to be near 0.5 if the model adequately fits the data. Extreme p-values (near 0 or 1) suggest poor fit. The posterior predictive p-value can be estimated as follows: Let θ^* represent a draw from the posterior distribution $f(\theta \mid y_{\text{obs}})$, and let y^*_{rep} represent a draw from $f(y_{\text{rep}} \mid \theta^*)$. Then marginally y^*_{rep} is a sample from the posterior predictive distribution $f(y_{\text{rep}} \mid y_{\text{obs}})$. Computing the p-value is relatively easy using the simulated values of θ^* from the Gibbs sampler. For each simulated value θ^*, we can simulate y^*_{rep} from the model and compute $T(y^*_{\text{rep}}, \theta^*)$ and $T(y_{\text{obs}}, \theta^*)$. Then the p-value is estimated by the proportion of times $T(y^*_{\text{rep}}, \theta^*)$ exceeds $T(y_{\text{obs}}, \theta^*)$.

To carry out the posterior predictive model checking, we need to specify a test quantity $T(y, \theta)$. In our LFS application, we consider two test quantities. The first one is $T_1 = \sum_{i=1}^{m} (y_i - \theta_i)^2 \big/ \text{var}(y_i \mid \theta)$, a general goodness-of-fit test statistic that resembles the classical χ^2 goodness-of-fit measure. T_1 is suggested in Gelman, Carlin, Stern and Rubin (2004) and has been widely used in practice for an omnibus measure of model fit. Following You (2008b) and You and Zhou (2011), we also consider another test quantity $T_2 = \big|\max(y_i) - \text{mean}(\theta_i)\big| - \big|\min(y_i) - \text{mean}(\theta_i)\big|$, which is particularly useful for HB area level model checking. For the YLLM with the correct prior specification (when $a = b = 1$) as presented in Table 3.3, we obtained an estimated average p-value of 0.428 using test quantity T_1 and a p-value of 0.477 using test quantity T_2. Both numbers are adequate to support the model fit. Thus, we have no indication of lack of overall model fit for the YLLM model.

5. Simulation study

In this section, we conduct a simulation study to evaluate the performance of the prior specification for the regression parameter $\boldsymbol{\beta}$. Following You and Hidiroglou (2023), we generate the data using the Fay-Herriot model with estimated sampling variances based on the LFS data studied in Section 3. We consider m = 65 areas in the simulation study. Let θ_i be the simulated true parameter of interest. The θ_i's are generated as $\theta_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i = \beta_0 + x_i\beta_1 + v_i$, where x_i is the LFS beneficiary rate of the month May 2011, β_0 = 0.0507 and β_1 = 0.42550, and v_i is generated from $N(0, \sigma_v^2)$, and σ_v^2 = 0.0001662. The values of

β_0, β_1 and σ_v^2 are obtained from the prior specification of Section 3. The direct estimate $\hat{\theta}_i$ is generated as $\hat{\theta}_i = \theta_i + e_i$, where $e_i = N(0, \sigma_i^2)$. The sampling variance σ_i^2 is the smoothed sampling variance obtained from the direct sampling variance estimates using the average smoothed sampling variance (ASM) of You and Hidiroglou (2023). The smoothed sampling variance σ_i^2 is treated as the true sampling variance in the simulation. The direct sampling variance estimate s_i^2 is then generated as $s_i^2 = (d_i)^{-1}\sigma_i^2\chi_{d_i}^2$, where $d = n_i - 1$, as in Rivest and Vandal (2002), Wang and Fuller (2003) and You (2021). We perform 5,000 simulation runs. For each run, we use the Gibbs sampling procedure to estimate the model parameters, and the Gibbs sampling procedure was run with 1,000 burn-in period and 2,000 more iterations for each simulation run.

We apply the Model 2 YLLM to the simulated data using noninformative and informative priors for **β**. The prior for σ_v^2 is the vague prior IG(0.0001, 0.0001). To compare the HB estimator under different priors across areas, we consider the average absolute relative bias (ARB) for the HB estimator $\hat{\theta}_i$ of the simulated small area mean θ_i defined as $\overline{\text{ARB}} = \left(\sum_{i=1}^{m} \text{ARB}_i\right) \Big/ m$, where

$$\text{ARB}_i = \left| \frac{1}{R} \sum_{r=1}^{R} \frac{(\hat{\theta}_i^{(r)} - \theta_i^{(r)})}{\theta_i^{(r)}} \right|,$$

and $\hat{\theta}_i^{(r)}$ is the HB estimate and $\theta_i^{(r)}$ is the generated mean based on the *r*th simulated sample, R = 5,000, m = 65.

We also compute the relative root mean squared error (RRMSE) for the HB estimator under different priors. Then we take the average of the RRMSE over areas. The average simulation RRMSE is computed as $\overline{\text{RRMSE}} = \left(\sum_{i=1}^{m} \text{RRMSE}_i\right) \Big/ m$, where $\text{RRMSE}_i = \left(\sum_{r=1}^{R} \text{RRMSE}_i^{(r)}\right) \Big/ R$, and

$$\text{RRMSE}_i^{(r)} = \frac{\sqrt{(\hat{\theta}_i^{(r)} - \theta_i^{(r)})^2}}{\theta_i^{(r)}}.$$

For the HB estimator, we can also compare the average estimated CV as $\overline{\text{CV}} = \left(\sum_{i=1}^{m} \text{CV}_i\right) \Big/ m$, where $\text{CV}_i = \left(\sum_{r=1}^{R} \text{CV}_i^{(r)}\right) \Big/ R$ and

$$\text{CV}_i^{(r)} = \sqrt{\text{Var}(\hat{\theta}_i^{(r)})} \Big/ \hat{\theta}_i^{(r)},$$

where $\text{var}(\hat{\theta}_i^{(r)})$ is the estimated posterior variance of the HB estimator $\hat{\theta}_i^{(r)}$ for the *i*th area.

In the simulation study, the priors for **β** are specified as a non-informative flat prior $\pi(\boldsymbol{\beta}) \propto 1$ and an informative prior $\pi(\boldsymbol{\beta}) \sim N(a\boldsymbol{\beta}_0, b\boldsymbol{\Phi}_0)$, where a = 0.5, 1, 2, 3, 4 and b = 1 and 16. When $a = 1$, the prior mean is correctly specified. When a is 0.5, 2, 3 or 4, the prior mean is mis-specified. When b = 16, the prior variance is very large and the prior should be roughly equivalent to a noninformative flat prior for **β**. Table 5.1 presents the comparison of the ARB, RRMSE and CV for the HB estimators based on different prior mean specifications (different a value) with the same prior variance specification $b = 1$ (regular prior variance). The ARB, RRMSE and CV of the direct estimator are also included in the table for benchmarking comparison. We also have included the average coverage rate $(\overline{\text{CR}})$ over the 65 small areas for the 95% Bayesian credible intervals of the Bayesian estimators for comparison.

Table 5.1
Comparison of average ARB, RRMSE, CV and CR for simulation study using different prior means, same prior variance $b\boldsymbol{\Phi}_0$ ($b = 1$)

	Direct estimate	Flat prior	Correct prior mean $a = 1$	Incorrect prior mean $a = 0.5$	Incorrect prior mean $a = 2$	Incorrect prior mean $a = 3$	Incorrect prior mean $a = 4$
Average $\overline{\text{ARB}}$	0.0067	0.0189	0.0190	0.0781	0.0653	0.0318	0.0213
Average $\overline{\text{RRMSE}}$	0.1822	0.1135	0.1115	0.1426	0.1778	0.1819	0.1826
Average $\overline{\text{CV}}$	0.3035	0.1365	0.1338	0.1961	0.2172	0.2484	0.2613
Average $\overline{\text{CR}}$	-	0.9445	0.9468	0.9307	0.9462	0.9471	0.9478

Note: ARB = absolute relative bias; CR = coverage rate; CV = coefficient of variation; RRMSE = relative root mean squared error.

When $\boldsymbol{\beta}$ has a flat prior, the average ARB is 0.0189, the average RRMSE is 0.1135 and the average CV is 0.1365 for the HB estimator, as reported in Table 5.1. When $\boldsymbol{\beta}$ has the default prior $\pi(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0)$ with $a = 1$ and $b = 1$, the average ARB is 0.0190, the average RRMSE is 0.1115 and the average CV is 0.1338. Thus, under the correct prior specification, the HB estimator has slightly better performance with very slightly smaller RRMSE and CV. But the average ARB is about the same under the flat prior and correct informative prior, as shown in columns 3 and 4 in Table 5.1. The direct estimator has large RRMSE and CV as shown in column 2 in Table 5.1. The average coverage rates for the Bayesian 95% credible intervals of the HB estimators using different priors are about the same around 95%, except for the prior mean $a = 0.5$, the average coverage rate is slightly lower at 93%.

If the prior has an incorrect prior mean specification $\pi(\boldsymbol{\beta}) \sim N(a\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0)$ with $a = 0.5$ or $a = 2$, 3 or 4, then the HB estimator will have either larger ARB and/or larger RRMSE and CV as shown in Table 5.1. For example, when $a = 0.5$, the prior mean is 0.5 β_0, then the average ARB is 0.0781, the average RRMSE is 0.1426 and the average CV is 0.1961. Thus, the ARB, RRMSE and CV values are obviously larger than the corresponding values under flat prior or correct prior. It is interesting to note that when the prior mean of $\boldsymbol{\beta}$ becomes large (for example, $a = 3$ or 4), then the estimate of model variance σ_v^2 becomes significantly larger. For example, the posterior estimate of σ_v^2 is 0.00018 when $a = 1$, 0.0215 when $a = 3$, and 0.0502 when $a = 4$. The HB estimator $\hat{\theta}_i$ puts much more weight on the direct estimate y_i and less weight on the regression estimate $\mathbf{x}_i'\boldsymbol{\beta}$ when the model variance is larger. Thus, the HB estimator performs more like the direct estimator under this situation as shown in the last two columns in Table 5.1 with a smaller average ARB but large RRMSE and large CV.

The results in Table 5.1 indicate that the HB estimator using flat prior or correct prior improves the direct estimator and achieves better results by reducing the average RRMSE and CV. The flat prior and correct prior perform very similarly in our simulation study. If the prior mean is not correctly specified, the HB estimator may have larger bias and CV, or the HB estimator may perform as the direct survey estimator as shown in Table 5.1. So, if we cannot determine the correct prior, the best option is to use the flat prior for the regression parameters in the HB small area models of You and Chapman (2006) and You (2021).

For comparison, we also compute the average ARB, RRMSE, CV and CR of the HB estimates when the prior variance is very large by setting the value for b as large as $b = 16$. Table 5.2 presents the comparison of the ARB, RRMSE and CV for the simulation study based on different prior mean specifications (different a value) with the same prior variance specification $b = 16$ (very large prior variance).

Table 5.2
Comparison of ARB, RRMSE, CV and CR for simulation study using different prior means, same prior variance $b\boldsymbol{\Phi}_0$ ($b = 16$)

	Direct estimate	Flat prior	Correct prior mean $a = 1$	Incorrect prior mean $a = 0.5$	Incorrect prior mean $a = 2$	Incorrect prior mean $a = 3$	Incorrect prior mean $a = 4$
Average $\overline{\text{ARB}}$	0.0067	0.0189	0.0188	0.0185	0.0195	0.0205	0.0217
Average $\overline{\text{RRMSE}}$	0.1822	0.1135	0.1135	0.1134	0.1136	0.1137	0.1137
Average $\overline{\text{CV}}$	0.3035	0.1365	0.1366	0.1364	0.1363	0.1361	0.1359
Average $\overline{\text{CR}}$	-	0.9445	0.9448	0.9445	0.9438	0.9441	0.9443

Note: ARB = absolute relative bias; CR = coverage rate; CV = coefficient of variation; RRMSE = relative root mean squared error.

As expected, when the prior variance is very large, all the average ARB, RRMSE, CV and CR of the HB estimators are respectively similar under different mean specifications and the flat prior. The simulation results in Table 5.2 show that use of a very large value for the prior variance is equivalent to the flat prior. As shown in Table 5.2, all the HB estimates have improved the direct estimates by substantially reducing the RRMSE and CV. The average coverage rates for the Bayesian intervals are almost all the same with a value around 94.45%. The simulation results also confirm the results in Table 3.3 whereby a prior with very large variance specification ($b = 16$ or 25) in the LFS application is roughly equivalent to the non-informative flat prior for $\boldsymbol{\beta}$.

6. Estimation for grouped small areas

In Sections 3 and 4, we studied the performance of the HB small area estimators using informative priors through the LFS application and a simulation study. In this section, we consider different area groups for the LFS application and simulation study. For the 65 CMA/CAs across Canada, we now divide these areas into five groups as follows: Area Group1: Eastern province areas, 1-14, including 14 CMA/CAs. Area Group2: Quebec areas, 15-23, including 9 CMA/CAs. Area Group3: Ontario areas, 24-46, including 23 CMA/CAs. Area Group4: Central province areas, 47-58, including 12 CMA/CAs
Area Group5: BC areas, 59-65, including 7 CMA/CAs

The purpose of using grouped areas is to study the performance of different priors when the number of small areas is relatively small. Now we apply the YLLM HB model to these five groups separately using an informative prior and a flat prior for $\boldsymbol{\beta}$, that is, informative prior $\pi(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0)$ and flat prior $\pi(\boldsymbol{\beta}) \propto 1$. We obtained the HB estimates and compared these estimates with the census values within each area group.

Table 6.1 reports the average ARE and CV for each area group for the HB estimator based on informative and flat priors.

Table 6.1
Comparison of ARE and CV for different area groups under separate area modeling

	Informative prior $\pi(\beta) \sim N(\beta_0, \Phi_0)$		Flat prior $\pi(\beta) \propto 1$	
	ARE	**CV**	**ARE**	**CV**
Area Group1 (14 areas)	0.1898	0.1142	0.2526	0.1274
Area Group2 (9 areas)	0.1133	0.0572	0.1279	0.0739
Area Group3 (23 areas)	0.1126	0.1107	0.0893	0.0989
Area Group4 (12 areas)	0.0915	0.0443	0.0937	0.0772
Area Group5 (7 areas)	0.0841	0.1219	0.0813	0.1688
Combined all areas (65 areas)	0.1226	0.0933	0.1297	0.1050

Note: ARB = absolute relative bias; CV = coefficient of variation.

For Area Group 1, the ARE under an informative prior is 0.1898, which is much smaller than the ARE of 0.2526 under a flat prior. The CV is 0.1142 under an informative prior, which is also smaller than the CV of 0.1274 under a flat prior. Thus, it is obvious that using the correct informative prior in the YLLM model leads to much better HB estimates for the 14 areas in the Group 1. For Group 2 which includes 9 areas in Quebec, both the ARE and CV are smaller under the informative prior as well. However, for Group 3 which includes 23 areas in Ontario, the ARE and CV are slightly larger under the informative prior than the corresponding values under flat prior. For Group 4, both the ARE and CV are smaller under the informative prior. For Group 5, the ARE is slightly larger under the informative prior, but the CV is much smaller. For the combined 65 areas, the ARE is 0.1226 under the informative prior and 0.1297 under the flat prior. The CV is 0.0933 under informative prior and 0.1050 under the flat prior. In summary, it is shown that using the informative prior for within group modeling is useful and leads to better results in terms of smaller ARE and CV for most area groups except for the Group 3 Ontario. When the number of areas is small, using the correct informative prior in the HB model can lead to better results, as the priors will be given more weight in the HB estimator. Therefore, when the number of modeling areas is small, the informative prior could be more useful.

We now evaluate the models using informative and flat priors for area groups by CPO ratio and DIC values as in Section 4. We compute the CPO ratio value of YLLM model using informative prior over YLLM model using flat prior. Table 6.2 presents the #-number of CPO ratios larger than 1 and the DIC values for YLLM model using informative and flat priors.

It is shown in Table 6.2 that for area groups with smaller number of areas both the CPO ratio and DIC value support the model using informative prior. For example, for Area Group 1 (14 areas), CPO ratio shows 9 out of 14 areas support the model using an informative prior, and the DIC value is also smaller (65.59) using the informative prior. For Area Group 3 (Ontario), both CPO and DIC support the model using a flat prior, which is consistent with the result reported in Table 6.1. For Ontario (23 areas), the flat prior is sufficient to achieve a better model fit in terms of CPO and DIC, and smaller ARE and CV as reported in

Table 6.1. Overall areas (65 areas), the CPO and DIC slightly support the model with an informative prior, with 38 areas out of 65 supporting the model with informative prior. The results from Table 6.1 and Table 6.2 indicate that the flat prior may be sufficient to achieve the same results as the correct informative prior when the number of areas is large enough. But when the number of areas is small (Area Groups 1,2,4,5), using the correct informative prior leads to better results and better model fit.

Table 6.2
# of CPO ratio larger than 1 and DIC values for area groups

	# of CPO ratio larger than 1	DIC Informative prior $\pi(\beta) \sim N(\beta_0, \Phi_0)$	DIC Flat prior $\pi(\beta) \propto 1$
Area Group1 (14 areas)	9	65.59	66.72
Area Group2 (9 areas)	7	25.41	27.84
Area Group3 (23 areas)	10	88.12	87.83
Area Group4 (12 areas)	8	30.61	32.47
Area Group5 (7 areas)	4	34.76	35.45
Combined all areas (65 areas)	38	254.52	255.95

Note: CPO = conditional predictive ordinate; DIC = deviance information criterion.

For further comparison and evaluation of the modeling within each group, we now use the results from Table 3.2 based on the modeling of all the 65 areas and compare with the census values within each area group as the comparison in Table 6.1. If the number of modeling areas is large, the sample data can provide a more reliable estimate and information, and therefore, we expect that using the flat prior or an informative prior will not make much of a difference. Table 6.3 provides the comparison results for each area group under the modeling of all areas. As expected, it is shown in Table 6.3 that the average ARE and CV values under informative prior and flat prior are similar and the difference is small for all the 5 area groups.

Table 6.3
Comparison of ARE and CV for different area groups under all area modeling

	Informative prior $\pi(\beta) \sim N(\beta_0, \Phi_0)$		Flat prior $\pi(\beta) \propto 1$	
	ARE	CV	ARE	CV
Area Group1 (14 areas)	0.1663	0.1104	0.1853	0.1226
Area Group2 (9 areas)	0.1166	0.1154	0.1206	0.1145
Area Group3 (23 areas)	0.1327	0.1153	0.1316	0.1153
Area Group4 (12 areas)	0.0995	0.1228	0.1041	0.1224
Area Group5 (7 areas)	0.0983	0.1188	0.0983	0.1198
Combined all areas (65 areas)	0.1278	0.1161	0.1329	0.1185

Note: ARE = absolute relative error; CV = coefficient of variation.

From the results presented in Table 6.1, Table 6.2 and Table 6.3, it is clear that when the number of small areas is relatively large, using the flat prior for the regression parameter can still provide efficient HB estimates for small area estimation. However, when the number of small areas is small, using the correct informative prior can lead to better HB estimates with smaller bias and smaller CV.

Finally, we also conducted a simulation study as in Section 5 within the five grouped areas. Table 6.4 presents the simulation results comparing average ARB, RRMSE and CV of the HB estimators within each grouped area using an informative prior $\pi(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0)$ and a noninformative flat prior $\pi(\boldsymbol{\beta}) \propto 1$ for $\boldsymbol{\beta}$. Table 6.4 demonstrates that the average ARB under the informative prior is slightly larger than the average ARB under the flat prior, but both $\overline{\text{ARB}}$s are all very small. The average RRMSE and CV under the informative prior are smaller than the corresponding values under the flat prior for all area groups. For example, for Area Group 1, the $\overline{\text{ARB}}$ is 0.0108 under the informative prior and 0.0104 under the flat prior, whereas the $\overline{\text{CV}}$ is 0.1288 under the informative prior and 0.1427 under the flat prior. The simulation results indicate that using the correct informative prior can improve the HB estimation by achieving smaller RRMSE and CV for grouped small area estimation.

Table 6.4
Comparison of ARB, RRMSE and CV for different area group simulation study

	Informative prior $\pi(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Phi}_0)$			**Flat prior** $\pi(\boldsymbol{\beta}) \propto 1$		
	$\overline{\text{ARB}}$	$\overline{\text{RRMSE}}$	$\overline{\text{CV}}$	$\overline{\text{ARB}}$	$\overline{\text{RRMSE}}$	$\overline{\text{CV}}$
Area Group1 (14 areas)	0.0108	0.0967	0.1288	0.0104	0.1091	0.1427
Area Group2 (9 areas)	0.0181	0.1199	0.1587	0.0163	0.1303	0.1751
Area Group3 (23 areas)	0.0183	0.1172	0.1447	0.0152	0.1224	0.1537
Area Group4 (12 areas)	0.0185	0.1202	0.1563	0.0146	0.1323	0.1732
Area Group5 (7 areas)	0.0203	0.1169	0.1594	0.0191	0.1335	0.1839

Note: ARB = absolute relative bias; CV = coefficient of variation; RRMSE = relative root mean squared error.

7. Concluding remarks

In this paper, we have studied the performance of the hierarchical Bayesian small area estimators using informative and noninformative flat priors for the regression parameters in the linking model of the area level small area HB models. Our real data analysis and simulation study indicate that both the default informative prior and flat prior for the regression parameters can perform very well when the number of small areas is relatively large. The flat prior $\pi(\boldsymbol{\beta}) \propto 1$ can be sufficient in general for the HB small area models when the number of small areas is large. This conclusion is also supported by the results of Bayesian model selection and Bayesian model check in the data analysis. Wrong specification of informative prior means can lead to large bias and large CV. When the number of small areas is relatively small, the use of correct informative prior in the HB modeling can be helpful to achieve better results in terms of smaller bias and smaller CV. For example, in the application of census undercoverage estimation studied in You and Rao (2002) and You, Rao and Dick (2004), the number of small areas (provinces) is very small, only 10 direct undercoverage estimates are available at province level across Canada, in which case the correct informative priors may be helpful in the HB models to achieve better model-based estimates and better model fit. Of course, how to construct a correct and proper informative prior is not easy in practice. We already provided some general guidance of choosing priors in the Introduction section. For future work of

LFS application, we plan to investigate and study some possible options of prior specification, for example, by borrowing information from previous monthly data using an iterative approach for the LFS data, that is, using the posterior estimates of previous month as the prior information for the current month or future month, and evaluate the corresponding HB estimates through real data analysis and simulation study.

Using historic data to construct priors for the model parameters can be useful and can provide direct information on the model parameters. Another way to make use of historic data is to borrow information over time using a time series model. For example, an autoregressive or random walk term can be added to the small area mean model (e.g., Rao and Yu, 1994; Datta, Lahiri, Maiti, and Lu, 1999). You, Rao and Gambino (2003) and You (2008a) used cross-sectional and time series models for the LFS data to improve the direct LFS estimates. It would be interesting to investigate and compare the prior specification and time series model for the use of historic data in the future work.

Benchmarking is another method to make use of external or prior data to improve the model-based estimates at the aggregated level. For Bayesian benchmarking methods, Zhang and Bryant (2020) provide a nice review of different Bayesian benchmarking methods in the literature and propose a flexible fully Bayesian approach to benchmarking that allows for a wide range of models and benchmarks. They revise the likelihood by multiplying it by a probability distribution that measures agreement with the benchmarks and outline Markov chain Monte Carlo methods to generate samples from benchmarked posterior distributions. More recently, Okonek and Wakefield (2024) critically review Bayesian benchmarking methods in the context of small area estimation, and they improve the method of Zhang and Bryant (2020) by proposing a novel approach in which posterior samples of small area characteristics from an un-benchmarked model can be combined with a rejection sampler or Metropolis-Hastings algorithm to produce benchmarked posterior distributions in a computationally efficient way. For the LFS application in our study, it would also be interesting to compare and evaluate the benchmarked HB estimates with the HB estimates based on different prior specifications.

Acknowledgements

We would like to thank three referees, the Associate Editor and the Editor for their constructive comments and suggestions so that we can improve the paper substantially and make the paper more interesting to read and applicable for practical users.

References

Bradley, J.R., Holan, S.H. and Wikle, C.K. (2018). Computationally efficient multivariate spatio-temporal models for high-dimensional count-valued data (with Discussion). *Bayesian Analysis*, 13, 253-310.

Bradley, J.R., Holan, S.H. and Wikle, C.K. (2020). Bayesian hierarchical models with conjugate full-conditional distributions for dependent data from the natural exponential family. *Journal of the American Statistical Association*, 115, 2037-2052.

Chen, M.-H., Shao, Q.-M. and Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer.

Chip, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49, 327-335.

Consonni, G., Fouskakis, D., Liseo, B. and Ntzoufras, I. (2018). Prior distributions for objective Bayesian analysis. *Bayesian Analysis*, 13, 627-679.

Dass, S.C., Maiti, T., Ren, H. and Sinha, S. (2012). Confidence interval estimation of small area parameters shrinking both means and variances. *Survey Methodology*, 38(2), 173-187. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2012002/article/11756-eng.pdf.

Datta, G.S., Lahiri, P., Maiti, T. and Lu, K.L. (1999). Hierarchical Bayes estimation of unemployment rates for the states of the U.S. *Journal of the American Statistical Association*, 94, 1074-1082.

Fay, R.E. and Herriot, R.A. (1979). Estimation of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 268-277.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515-533.

Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-472.

Gelman, A., Meng, X.L. and Stern, H.S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, 6, 733-807.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004). *Bayesian Data Analysis*. 2nd Edition, Chapman & Hall/CRC.

Ghosh, M., Myung, J. and Moura, F.A.S. (2018). Robust Bayesian small area estimation. *Survey Methodology*, 44(1), 101-115. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2018001/article/54959-eng.pdf.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC.

Grzenda, W. (2016). Informative versus non-informative prior distributions and their impact on the accuracy of Bayesian inference. *Statistics in Transition New Series*, 17, 763-780.

Hidiroglou, M.A., Beaumont, J.-F. and Yung, W. (2019). Development of a small area estimation system at Statistics Canada. *Survey Methodology*, 45(1), 101-126. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2019001/article/00009-eng.pdf.

Lemoine, N.P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, 128: 912-928. DOI: 10.1111/oik.05985.

Lesage, E., Beaumont, J.-F. and Bocci, C. (2021). Two local diagnostics to evaluate the efficiency of the empirical best predictor under the Fay-Herriot model. *Survey Methodology*, 47(2), 279-297. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2021002/article/00001-eng.pdf.

Lunn, D.J., Thomas, A., Best, N. and Spiegelhalter, D.J. (2000). WinBUGS – A Bayesian modeling framework: Concepts, structure and extensibility. *Statistics and Computing*, 10, 325-337.

Maiti, T., Ren, H. and Sinha, S. (2014). Prediction error of small area predictors shrinking both means and variances. *Scandinavian Journal of Statistics*, 41, 775-790.

Meng, X.L. (1994). Posterior predictive *p* value. *The Annals of Statistics*, 22, 1142-1160.

Methodology of the Canadian Labour Force Survey (2017). Statistics Canada: 71-526-X, ISBN 978-0-660-24068-8. Available at: https://www150.statcan.gc.ca/n1/pub/71-526-x/71-526-x2017001-eng.htm.

Mikkola, P., Martin, O.A., Chandramouli, S., Hartmann, M., Abril Pla, O., Thomas, O., Pesonen, H., Corander, J., Vehtari, A., Kaski, S., Bürkner, P.C. and Klami, A. (2024). Prior knowledge elicitation: The past, present, and future. *Bayesian Analysis*, 19, 1129-1161.

Molina, I., Nandram, B. and Rao, J.N.K. (2014). Small area estimation of general parameters with application to poverty indicators: A hierarchical Bayes approach. *Annals of Applied Statistics*, 8, 852-885.

Okonek, T. and Wakefield, J. (2024). A computationally efficient approach to fully Bayesian benchmarking. *Journal of Official Statistics*, 40, 283-316.

Parker, P.A., Holan, S.H. and Janicki, R. (2023). Conjugate modeling approaches for small area estimation with heteroscedastic structure. *Journal of Survey Statistics and Methodology*, smad002, https://doi.org/10.1093/jssam/smad002.

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

Rao, J.N.K. and Yu, M. (1994). Small area estimation by combining time series and cross-sectional data. *The Canadian Journal of Statistics*, 22, 511-528.

Rivest, L.P. and Vandal, N. (2002). Mean squared error estimation for small areas when the small area variances are estimated. *Proceedings of the International Conference on Recent Advances in Survey Sampling*, July 10-13, 2002, Ottawa, Canada.

Sakshaug, J.W., Wiśniowski, A., Ruiz, D.A.P. and Blom, A.G. (2019). Supplementing small probability samples with nonprobability samples: A Bayesian approach. *Journal of Official Statistics*, 35, 653-681.

Sinharay, S. and Stern, H.S. (2003). Posterior predictive modelchecking in hierarchical models. *Journal of Statistical Planning and Inference*, 111, 209-221.

Spiegelhalter, D.J., Best, N., Carlin, B.P. and van der Linde, A. (2002). Bayesain measures of model complexity and fit. *Journal of Royal Statistical Society*, B, 64, 583-639.

Sugasawa, S., Tamae, H. and Kubokawa, T. (2017). Bayesian estimators for small area models shrinking both means and variances. *Scandinavian Journal of Statistics*, 44, 150-167.

Wang, J. and Fuller, W.A. (2003). The mean square error of small area predictors constructed with estimated area variances. *Journal of the American Statistical Association*, 98, 716-723.

Wiśniowski, A., Sakshaug, J.W., Ruiz, D.A.P. and Blom, A.G. (2020). Integrating probability and nonprobability samples for survey inference. *Journal of Survey Statistics and Methodology*, 8, 120-147.

You, Y. (2008a). An integrated modeling approach to unemployment rate estimation for sub-provincial areas of Canada. *Survey Methodology*, 34(1), 19-27. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2008001/article/10614-eng.pdf.

You, Y. (2008b). Small area estimation using area level models with model checking and applications. *Proceedings of Survey Methods Section*, *Statistical Society of Canada*.

You, Y. (2021). Small area estimation using Fay-Herriot area level model with sampling variance smoothing and modeling. *Survey Methodology*, 47(2), 361-370. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2021002/article/00007-eng.pdf.

You, Y. (2023). An empirical study of hierarchical Bayes small area estimators using different priors for model variances. *Statistics in Transition New Series*, 24, 139-148.

You, Y. and Chapman, B. (2006). Small area estimation using area level models and estimated sampling variances. *Survey Methodology*, 32(1), 97-103. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2006001/article/9263-eng.pdf.

You, Y. and Hidiroglou, M. (2023). Application of sampling variance smoothing methods for small area proportion estimation. *Journal of Official Statistics*, 39, 571-590.

You, Y. and Rao, J.N.K. (2000). Hierarchical Bayes estimation of small area means using multi-level models. *Survey Methodology*, 26(2), 173-181. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2000002/article/5537-eng.pdf.

You, Y. and Rao, J.N.K. (2002) Small area estimation using unmatched sampling and linking models. *The Canadian Journal of Statistics*, 30, 3-15.

You, Y., Rao, J.N.K. and Dick, P. (2004). Benchmarking hierarchical Bayes small area estimators in the Canadian census undercoverage estimation. *Statistics in Transition*, 6, 631-640.

You, Y., Rao, J.N.K. and Gambino, J. (2003). Model-based unemployment rate estimation for the Canadian Labor Force Survey: A hierarchical Bayes approach. *Survey Methodology*, 29(1), 25-32. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2003001/article/6602-eng.pdf.

You, Y. and Zhou, Q.M. (2011). Hierarchical Bayes small area estimation under a spatial model with application to health survey data. *Survey Methodology*, 37(1), 25-37. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2011001/article/11445-eng.pdf.

Zhang, J.L. and Bryant, J. (2020). Fully Bayesian benchmarking of small area estimation models. *Journal of Official Statistics*, 36, 197-223. DOI: https://doi.org/10.2478/jos-2020-0010.

Zondervan-Zwijnenburg, M., Peeters, M., Depaoli, S. and Van de Schoot, R. (2017). Where do priors come from? Applying guidelines to construct informative priors in small sample research. *Research in Human Development*, 14, 305-320. DOI: 10.1080/15427609.2017.1370966.