Statistica Sinica Preprint No: SS-2019-0239						
Title	An Approximate Bayesian Approach to Model-assisted					
	Survey Estimation with Many Auxiliary Variables					
Manuscript ID	SS-2019-0239					
URL	http://www.stat.sinica.edu.tw/statistica/					
DOI	10.5705/ss.202019.0239Shonosuke Sugasawa andJae Kwang Kim					
Complete List of Authors						
Corresponding Author	Shonosuke Sugasawa					
E-mail	shonosuke622@gmail.com					
Notice: Accepted version subject to English editing.						

Statistica Sinica: Newly accepted Paper (accepted author-version subject to English editing)

Statistica Sinica

AN APPROXIMATE BAYESIAN APPROACH TO MODEL-ASSISTED SURVEY ESTIMATION WITH MANY AUXILIARY VARIABLES

Shonosuke Sugasawa¹ and Jae Kwang Kim²

¹The University of Tokyo and ²Iowa State University

Abstract: Model-assisted estimation with complex survey data is an important practical problem in survey sampling. When there are many auxiliary variables, selecting significant variables associated with the study variable would be necessary to achieve efficient estimation of population parameters of interest. In this paper, we formulate a regularized regression estimator in the framework of Bayesian inference using the penalty function as the shrinkage prior for model selection. The proposed Bayesian approach enables us to get not only efficient point estimates but also valid credible intervals. Results from two limited simulation studies are presented to compare with existing frequentist methods.

Key words and phrases: Generalized regression estimation; Regularization; Shrinkage prior; Survey Sampling

1. Introduction

Probability sampling is a scientific tool for obtaining a representative sample from the target population. In order to estimate a finite population total from a target population, Horvitz-Thompson (HT) estimator obtained from a probability sample satisfies design-consistency and the resulting inference is justified from the randomization perspective (Horvitz and Thompson, 1952). However, the HT estimator uses the first-order inclusion probability only and does not fully incorporate all available information in the finite population. To improve its efficiency, regression estimation is often used by incorporating auxiliary information in the finite population. Deville and Särndal (1992), Fuller (2002), Kim and Park (2010), and Breidt and Opsomer (2017) present comprehensive overviews of variants of regression estimation in survey sampling. There are also other directions of improvement on the HT estimator based on prediction using augmented models (e.g. Zeng and Little, 2003, 2005; Zanganeh and Little, 2015).

The regression estimation approaches in survey sampling assume a model for the finite population, i.e., the superpopulation model, such as

$$y_i = \boldsymbol{x}_i^t \boldsymbol{\beta} + e_i, \tag{1.1}$$

where y_i is a response variable, x_i and β are vectors of auxiliary variables and regression coefficients, respectively, and e_i is an error term satisfying $E(e_i) = 0$ and $Var(e_i) = \sigma^2$. The superpopulation model does not necessarily hold in the sample as the sampling design can be informative (e.g. Pfeffermann and Sverchkov, 1999; Little, 2004). Under the regression superpopulation model in (1.1), Isaki and Fuller (1982) show that the asymptotic variance of the regression estimator achieves the lower bound of Godambe and Joshi (1965). Thus, the regression estimator is asymptotically efficient in the sense of achieving the minimum anticipated variance under the joint distribution of the sampling design and the superpopulation model in (1.1).

On the other hand, the dimension of the auxiliary variables x_i could be large in practice. Even when the number of observed covariates is not necessarily large, the dimension of x_i could be very large once we include polynomial or interaction terms to achieve flexible modeling, as considered in Section 7. However, in this case, the optimality of the regression estimator is untenable. When there are many auxiliary variables, the asymptotic bias of the regression estimator using all the auxiliary variables is no longer negligible and the resulting inference can be problematic. Simply put, including irrelevant auxiliary variables can introduce substantial variability in point estimation, but its uncertainty is not fully accounted for by the stan-

dard linearization variance estimation, resulting in misleading inference.

To overcome the problem, variable selection techniques for regression estimation have been considered in literatures (e.g. Silva and Skinner, 1997; Särndal and Lundström, 2005). The classical selection approach is based on a step-wise method. However, the step-wise methods will not necessarily produce the best model (e.g. Dempster et al., 1977) although the potential effect on prediction could be limited. Another approach is to employ regularized estimation of regression coefficients. Recently, McConville et al. (2017) propose a regularized regression estimation approach based on the LASSO penalty of Tibshirani (1996). However, there are two main problems with the regularization approach in regression estimation. First, the choice of the regularization parameter is not straightforward under survey sampling. Second, the frequentist inference accounting for model selection uncertainty is notoriously difficult to make.

In this paper, to overcome the above difficulties, we adopt a Bayesian framework in the regularized regression estimation. We first introduce an approximate Bayesian approach for regression estimation when $p + 1 = \dim(\boldsymbol{x})$ is fixed, using the approximate Bayesian approach considered in Wang et al. (2018). The proposed Bayesian method fully captures the uncertainty in parameter estimation for the regression estimator and has good

coverage properties. Second, the proposed Bayesian method is extended to the problem of large p in regularized regression estimation. By incorporating the penalty function for regularization into the prior distribution, the uncertainty associated with model selection and parameter estimation is fully captured in the Bayesian machinery. Furthermore, the choice of the penalty parameter can be handled by using its posterior distribution. Hence, the proposed method provides a unified approach to Bayesian inference with sparse model-assisted survey estimation. The proposed method is a calibrated Bayesian (Little, 2012) and it is asymptotically equivalent to the frequentist model-assisted approach for a fixed p.

The paper is organized as follows. In Section 2, the basic setup is introduced. In Section 3, the approximate Bayesian inference using regression estimation is proposed under a fixed p setup. In Section 4, the proposed method is extended to high dimensional setup by developing sparse regression estimation using shrinkage prior distributions. In Section 5, the proposed method is extended to non-linear regression models. In Section 6, results from two limited simulation studies are presented. The proposed method is applied to the real data example in Section 7. Some concluding remarks are made in Section 8. R code is available at GitHub repository (https://github.com/sshonosuke/ABMASE).

2. Basic setup

Consider a finite population of a known size N. Associated with unit i in the finite population, we consider measurement $\{\boldsymbol{x}_i, y_i\}$ where \boldsymbol{x}_i is the vector of auxiliary variables with dimension p and y_i is the study variable of interest. We are interested in estimating the finite population mean $\bar{Y} = N^{-1} \sum_{i=1}^{N} y_i$ from a sample selected by a probability sampling design. Let A be the index set of the sample and we observe $\{\boldsymbol{x}_i, y_i\}_{i \in A}$ from the sample. The HT estimator $\hat{Y}_{HT} = N^{-1} \sum_{i \in A} \pi_i^{-1} y_i$, where π_i is the first-order inclusion probability of unit i, is design unbiased but it is not necessarily efficient.

If the finite population mean $\bar{\boldsymbol{X}} = N^{-1} \sum_{i=1}^{N} \boldsymbol{x}_i$ is known, then we can improve the efficiency of \hat{Y}_{HT} by using the following regression estimator:

$$\hat{ar{Y}}_{ ext{reg}} = rac{1}{N}\sum_{i=1}^{N}oldsymbol{x}_{i}^{t}\hat{oldsymbol{eta}}$$

where $\hat{\boldsymbol{\beta}}$ is an estimator of $\boldsymbol{\beta}$ in (1.1). Typically, we use $\hat{\boldsymbol{\beta}}$ obtained by minimizing the weighted quadratic loss

$$Q(\boldsymbol{\beta}) = \sum_{i \in A} \pi_i^{-1} (y_i - \boldsymbol{x}_i^t \boldsymbol{\beta})^2, \qquad (2.2)$$

motivated from the model (1.1). If an intercept term is included in \boldsymbol{x}_i such that $\boldsymbol{x}_i^t = (1, \boldsymbol{x}_{1i}^t)$, we can express

$$\hat{\bar{Y}}_{\text{reg}} = \hat{\beta}_0 + \bar{\boldsymbol{X}}_1^t \hat{\boldsymbol{\beta}}_1 = \hat{N}^{-1} \sum_{i \in A} \pi_i^{-1} \left(y_i - \boldsymbol{x}_{1i}^t \hat{\boldsymbol{\beta}}_1 \right) + \bar{\boldsymbol{X}}_1^t \hat{\boldsymbol{\beta}}_1 \qquad (2.3)$$

where $\hat{N} = \sum_{i \in A} \pi_i^{-1}$ and $\hat{\beta}_1$ is given by

$$\hat{\boldsymbol{\beta}}_{1} = \left\{ \sum_{i \in A} \pi_{i}^{-1} (\boldsymbol{x}_{1i} - \hat{\bar{\boldsymbol{X}}}_{1,\pi})^{\otimes 2} \right\}^{-1} \sum_{i \in A} \pi_{i}^{-1} (\boldsymbol{x}_{1i} - \hat{\bar{\boldsymbol{X}}}_{1,\pi}) y_{i} \qquad (2.4)$$

where $\hat{\bar{X}}_{1,\pi} = \hat{N}^{-1} \sum_{i \in A} \pi_i^{-1} x_{1i}$ and $B^{\otimes 2} = BB'$ for some matrix B.

To discuss some asymptotic properties of \hat{Y}_{reg} in (2.3), we consider a sequence of finite populations and samples as discussed in Isaki and Fuller (1982), where N increases with n. Note that

$$\hat{\bar{Y}}_{\text{reg}} - \bar{Y} = \hat{\bar{Y}}_{\pi} - \bar{Y} + \left(\bar{\boldsymbol{X}}_{1} - \hat{\bar{\boldsymbol{X}}}_{1,\pi}\right)^{t} \hat{\boldsymbol{\beta}}_{1}$$
$$= \hat{\bar{Y}}_{\pi} - \bar{Y} + \left(\bar{\boldsymbol{X}}_{1} - \hat{\bar{\boldsymbol{X}}}_{1,\pi}\right)^{t} \boldsymbol{\beta}_{1} + R_{n}$$
(2.5)

where $\hat{Y}_{\pi} = \hat{N}^{-1} \sum_{i \in A} \pi_i^{-1} y_i$ and

$$R_n = \left(\bar{\boldsymbol{X}}_1 - \hat{\bar{\boldsymbol{X}}}_1\right)^t \left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\right)$$

for any β_1 . If we choose $\beta_1 = p \lim_{n \to \infty} \hat{\beta}_1$ with respect to the sampling mechanism and $p = \dim(\mathbf{x}_1)$ is fixed in the asymptotic setup, then we can obtain $R_n = O_p(n^{-1})$ and safely use the main terms of (2.5) to describe the asymptotic behavior of \hat{Y}_{reg} . To emphasize its dependence on $\hat{\beta}_1$ in the regression estimator, we can write $\hat{Y}_{reg} = \hat{Y}_{reg}(\hat{\beta}_1)$. Roughly speaking, we can obtain

$$\sqrt{n}\left\{\hat{Y}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1) - \hat{Y}_{\text{reg}}(\boldsymbol{\beta}_1)\right\} = O_p(n^{-1/2}p).$$
(2.6)

and, if $p = o(n^{1/2})$ then we can safely ignore the effect of estimating β_1 in the regression estimator. See Supplementary Material for a sketched proof of (2.6).

If, on the other hand, the dimension p is larger than $O(n^{1/2})$, then we cannot ignore the effect of estimating β_1 . In this case, we can consider using some variable selection idea to reduce the dimension of X. For variable selection, we may employ techniques of regularized estimation of regression coefficients. The regularization method can be described as finding

$$(\hat{\beta}_0^{(R)}, \hat{\boldsymbol{\beta}}_1^{(R)}) = \operatorname{argmin}_{\beta_0, \beta_1} \{ Q(\boldsymbol{\beta}) + p_{\lambda}(\boldsymbol{\beta}_1) \}, \qquad (2.7)$$

where $Q(\boldsymbol{\beta})$ is defined in (2.2) and $p_{\lambda}(\boldsymbol{\beta}_1)$ is a penalty function with parameter λ . Some popular penalty functions are presented in Table 1. Once the solution to (2.7) is obtained, then the regularized regression estimator is given by

$$\hat{\bar{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_{1}^{(R)}) = \bar{\boldsymbol{X}}_{1}^{t} \hat{\boldsymbol{\beta}}_{1}^{(R)} + \frac{1}{\hat{N}} \sum_{i \in A} \frac{1}{\pi_{i}} \left(y_{i} - \boldsymbol{x}_{1i}^{t} \hat{\boldsymbol{\beta}}_{1}^{(R)} \right).$$
(2.8)

Statistical inference with the regularized regression estimator in (2.8) is not fully investigated in the literature. For example, Chen et al. (2018) consider the regularized regression estimator using adaptive LASSO of Zou (2006), but they assume that the sampling design is non-informative and the uncertainty in model selection is not fully incorporated in their inference.

Table 1: Popular penalized regression methods									
Method	Reference	Penalty function							
Ridge	Hoerl and Kennard (1970)	$p_{\lambda}(\boldsymbol{\beta}) = \lambda \sum_{j=1}^{p} \beta_{j}^{2}$							
LASSO	Tibshirani (1996)	$p_{\lambda}(\boldsymbol{eta}) = \lambda \sum_{j=1}^{p} \beta_j $							
Adaptive LASSO	Zou (2006)	$p_{\lambda}(\boldsymbol{eta}) = \lambda \sum_{j=1}^{p} \left(\left \beta_{j} \right / \left \hat{\beta}_{j} \right \right)$							
Elastic Net	Zou and Hastie (2005)	$p_{\lambda}(\boldsymbol{\beta}) = \lambda_1 \sum_{j=1}^p \beta_j + \lambda_2 \sum_{j=1}^p \beta_j^2$							

Roughly speaking, making inference after model selection under superpopulation frequentist framework is difficult. The approximated Bayesian method introduced in the next section will capture the full uncertainty in the Bayesian framework.

3. Approximate Bayesian survey regression estimation

Developing Bayesian model-assisted inference under complex sampling is a challenging problem in statistics. Wang et al. (2018) recently propose the so-called approximate Bayesian method for design-based inference using asymptotic normality of a design-consistent estimator. Specifically, for a given parameter θ with a prior distribution $\pi(\theta)$, if one can find a designconsistent estimator $\hat{\theta}$ of θ , then the approximate posterior distribution of θ is given by

$$p(\theta \mid \hat{\theta}) = \frac{f(\hat{\theta} \mid \theta)\pi(\theta)}{\int f(\hat{\theta} \mid \theta)\pi(\theta)d\theta},$$
(3.9)

where $f(\hat{\theta} \mid \theta)$ is the sampling distribution of $\hat{\theta}$, which is often approximated by a normal distribution.

Drawing on this idea, one can develop an approximate Bayesian approach to capture the full uncertainty in the regression estimator. Let

$$\hat{\boldsymbol{eta}} = \left(\sum_{i\in A} \pi_i^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^t\right)^{-1} \sum_{i\in A} \pi_i^{-1} \boldsymbol{x}_i y_i$$

be the design-consistent estimator of $\boldsymbol{\beta}$ and $\hat{\boldsymbol{V}}_{\beta}$ be the corresponding asymptotic variance-covariance matrix of $\hat{\boldsymbol{\beta}}$, given by

$$\hat{\boldsymbol{V}}_{\beta} = \left(\sum_{i \in A} \pi_i^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^t\right)^{-1} \left(\sum_{i \in A} \sum_{j \in A} \frac{\Delta_{ij}}{\pi_{ij}} \frac{\hat{e}_i \boldsymbol{x}_i}{\pi_i} \frac{\hat{e}_j \boldsymbol{x}_j^t}{\pi_j}\right) \left(\sum_{i \in A} \pi_i^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^t\right)^{-1},$$
(3.10)

where $\hat{e}_i = y_i - \boldsymbol{x}_i^t \hat{\boldsymbol{\beta}}$, $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$ and π_{ij} is the joint inclusion probability of unit *i* and *j*. Under some regularity conditions, as discussed in Chapter 2 of Fuller (2009), we can establish

$$\hat{\boldsymbol{V}}_{\beta 11}^{-1/2} \left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \right) \mid \boldsymbol{\beta} \stackrel{\mathcal{L}}{\longrightarrow} N(0, I)$$
(3.11)

as $n \to \infty$, where $\hat{\boldsymbol{V}}_{\beta 11}$ is the submatrix of $\hat{\boldsymbol{V}}_{\beta}$ with

$$\hat{\boldsymbol{V}}_{\beta} = \begin{pmatrix} \hat{V}_{\beta 00} & \hat{V}_{\beta 01} \\ \hat{V}_{\beta 10} & \hat{V}_{\beta 11} \end{pmatrix}.$$
(3.12)

Thus, using (3.9) and (3.11), we can obtain the approximate posterior distribution of $\boldsymbol{\beta}$ as

$$p(\boldsymbol{\beta}_1|\hat{\boldsymbol{\beta}}_1) = \frac{\phi_p(\hat{\boldsymbol{\beta}}_1; \boldsymbol{\beta}_1, \hat{\boldsymbol{V}}_{\beta 11}) \pi(\boldsymbol{\beta}_1)}{\int \phi_p(\hat{\boldsymbol{\beta}}_1; \boldsymbol{\beta}_1, \hat{\boldsymbol{V}}_{\beta 11}) \pi(\boldsymbol{\beta}_1) \mathrm{d}\boldsymbol{\beta}_1},$$
(3.13)

where ϕ_p denotes a *p*-dimensional multivariate normal density and $\pi(\boldsymbol{\beta}_1)$ is a prior distribution for $\boldsymbol{\beta}_1$. We use a flat prior here but will use a shrinkage prior in Section 4.

Now, we consider the conditional posterior distribution of \bar{Y} for a given β_1 . First, define

$$\hat{Y}_{ ext{reg}}(oldsymbol{eta}_1) = oldsymbol{ar{X}}_1^t oldsymbol{eta}_1 + rac{1}{\hat{N}} \sum_{i \in A} rac{1}{\pi_i} \left(y_i - oldsymbol{x}_{1i}^t oldsymbol{eta}_1
ight).$$

Note that $\hat{Y}_{reg}(\boldsymbol{\beta}_1)$ is an approximately design-unbiased estimator of \bar{Y} , regardless of $\boldsymbol{\beta}_1$. Under some regularity conditions, we can show that $\hat{Y}_{reg}(\boldsymbol{\beta}_1)$ follows a normal distribution asymptotically. Thus, we obtain

$$\frac{\bar{Y}_{\text{reg}}(\boldsymbol{\beta}_1) - \bar{Y}}{\sqrt{\hat{V}_e(\boldsymbol{\beta}_1)}} \mid \bar{Y}, \boldsymbol{\beta}_1 \xrightarrow{\mathcal{L}} N(0, 1), \qquad (3.14)$$

where

$$\hat{V}_{e}(\boldsymbol{\beta}_{1}) = \frac{1}{N^{2}} \sum_{i \in A} \sum_{j \in A} \frac{\Delta_{ij}}{\pi_{ij}} \frac{1}{\pi_{i}} \frac{1}{\pi_{j}} (y_{i} - \boldsymbol{x}_{1i}^{t} \boldsymbol{\beta}_{1}) (y_{j} - \boldsymbol{x}_{1j}^{t} \boldsymbol{\beta}_{1}), \qquad (3.15)$$

is a design consistent variance estimator of $\hat{Y}_{reg}(\boldsymbol{\beta}_1)$ for given $\boldsymbol{\beta}_1$. We then use $\phi(\hat{Y}_{reg}(\boldsymbol{\beta}_1); \bar{Y}, \hat{V}_e(\boldsymbol{\beta}_1))$ as the density for the approximate sampling distribution of $\hat{Y}_{reg}(\boldsymbol{\beta}_1)$ in (3.14), where $\phi(\cdot; \mu, \sigma^2)$ is the normal density function with mean μ and variance σ^2 . Thus, the approximate conditional posterior distribution of \bar{Y} given $\boldsymbol{\beta}$ can be defined as

$$p(\bar{Y}|\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}_1),\boldsymbol{\beta}_1) \propto \phi(\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}_1);\bar{Y},\hat{V}_e(\boldsymbol{\beta}_1))\pi(\bar{Y}\mid\boldsymbol{\beta}_1), \qquad (3.16)$$

where $\pi(\bar{Y} \mid \boldsymbol{\beta}_1)$ is a conditional prior distribution of \bar{Y} given $\boldsymbol{\beta}_1$. Without extra assumptions or any prior information, we can use a flat prior distribution, namely, $\pi(\bar{Y} \mid \boldsymbol{\beta}_1) \propto 1$.

Therefore, combining (3.13) and (3.16), the approximate posterior distribution of \bar{Y} can be obtained as

$$p(\bar{Y}|\hat{Y}_{reg}(\hat{\boldsymbol{\beta}}_{1}), \hat{\boldsymbol{\beta}}_{1}) = \frac{\int p(\boldsymbol{\beta}_{1}|\hat{\boldsymbol{\beta}}_{1})\phi(\hat{Y}_{reg}(\boldsymbol{\beta}_{1}); \bar{Y}, \hat{V}_{e}(\boldsymbol{\beta}_{1}))\pi(\bar{Y} \mid \boldsymbol{\beta}_{1})d\boldsymbol{\beta}_{1}}{\int \int p(\boldsymbol{\beta}_{1}|\hat{\boldsymbol{\beta}}_{1})\phi(\hat{Y}_{reg}(\boldsymbol{\beta}_{1}); \bar{Y}, \hat{V}_{e}(\boldsymbol{\beta}_{1}))\pi(\bar{Y} \mid \boldsymbol{\beta}_{1})d\boldsymbol{\beta}_{1}d\bar{Y}} = \frac{\int \phi(\hat{Y}_{reg}(\boldsymbol{\beta}_{1}); \bar{Y}, \hat{V}_{e}(\boldsymbol{\beta}_{1}))\phi_{p}(\hat{\boldsymbol{\beta}}_{1}; \boldsymbol{\beta}_{1}, \hat{\boldsymbol{V}}_{\beta 11})\pi(\boldsymbol{\beta}_{1})\pi(\bar{Y} \mid \boldsymbol{\beta}_{1})d\boldsymbol{\beta}_{1}}{\int \int \phi(\hat{Y}_{reg}(\boldsymbol{\beta}_{1}); \bar{Y}, \hat{V}_{e}(\boldsymbol{\beta}_{1}))\phi_{p}(\hat{\boldsymbol{\beta}}_{1}; \boldsymbol{\beta}_{1}, \hat{\boldsymbol{V}}_{\beta 11})\pi(\boldsymbol{\beta}_{1})\pi(\bar{Y} \mid \boldsymbol{\beta}_{1})d\boldsymbol{\beta}_{1}d\bar{Y}}.$$

$$(3.17)$$

Generating posterior samples from (3.17) can be easily carried out via the following two steps:

- 1. Generate posterior sample β_1^* of β_1 from (3.13).
- 2. Generate posterior sample of \bar{Y} from (3.16) for given β_1^* .

Based on the approximate posterior samples of \bar{Y} , we can compute the posterior mean as a point estimator as well as credible intervals for uncertainty quantification for \bar{Y} including the variability in estimating β_1 .

The following theorem presents an asymptotic property of the proposed approximate Bayesian method. **Theorem 1.** Under the regularity conditions described in the Supplementary Material, conditional on the full sample data,

$$\sup_{\bar{Y}\in\Theta_{Y}} \left| p(\bar{Y}|\hat{\bar{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_{1}), \hat{\boldsymbol{\beta}}_{1}) - \phi(\bar{Y}; \hat{\bar{Y}}_{\text{reg}}, \hat{V}_{e}) \right| \to 0,$$
(3.18)

in probability as $n \to \infty$ while p is fixed, and $n/N \to f \in [0, 1)$, where Θ_Y is some Borel set for \bar{Y} and $p(\bar{Y}|\hat{Y}_{reg}(\hat{\beta}_1), \hat{\beta}_1)$ is given in (3.17).

Theorem 1 is a special case of the Bernstein-von Mises theorem (van der Vaart, 2000, Section 10.2) in survey regression estimation, and its sketched proof is given in the Supplementary Material. The proof is not quite rigorous but contains enough details to deliver the main ideas. According to Theorem 1, the credible interval for \bar{Y} constructed from the approximated posterior distribution (3.17) is asymptotically equivalent to the frequentist confidence interval based on the asymptotic normality of the common survey regression estimator. Therefore, the proposed Bayesian method implements the frequentist inference of the survey regression estimator at least asymptotically.

4. Approximate Bayesian method with shrinkage priors

We consider the case when there are many auxiliary variables in applying regression estimation. When p is large, it is desirable to select a suitable

subset of auxiliary variables that are associated with the response variable to avoid inefficient regression estimation due to including irrelevant covariates.

To deal with the problem in a Bayesian way, we may define the approximate posterior distribution of \bar{Y} given β_1 as similar to (3.17) with a different prior for β_1 to implement variable selection. That is, we use the same asymptotic distribution of the estimators $\hat{\beta}_1$ of β_1 and assign a shrinkage prior for β_1 . Let $\pi_{\lambda}(\beta_1)$ be the shrinkage prior for β_1 with a structural parameter λ which might be multivariate.

Among the several choices of shrinkage priors, we specifically consider two priors for β_1 : Laplace (Park and Casella, 2008) and horseshoe (Carvalho et al., 2009, 2010). The Laplace prior is given by $\pi_{\lambda}(\beta_1) \propto \exp(-\lambda \sum_{k=1}^p |\beta_k|)$, which is related to Lasso regression (Tibshirani, 1996), so that the proposed approximated Bayesian method can be seen as the Bayesian version of a survey regression estimator with Lasso (McConville et al., 2017). The horseshoe prior is a more advanced shrinkage prior of the form:

$$\pi_{\lambda}(\boldsymbol{\beta}_{1}) = \prod_{k=1}^{p} \int_{0}^{\infty} \phi(\beta_{k}; 0, \lambda^{2} u_{k}^{2}) \frac{2}{\pi(1+u_{k}^{2})} \mathrm{d}u_{k}, \qquad (4.19)$$

where $\phi(\cdot; a, b)$ denotes the normal density function with mean a and variance b. It is known that the horseshoe prior enjoys more severe shrinkage for the zero elements of β_1 than the Laplace prior, thus allowing strong signals to remain large (Carvalho et al., 2009).

Similarly to (3.13), we can develop a posterior distribution of β_1 using the shrinkage prior

$$p_{\lambda}(\boldsymbol{\beta}_{1}|\hat{\boldsymbol{\beta}}_{1}) = \frac{\phi(\hat{\boldsymbol{\beta}}_{1};\boldsymbol{\beta}_{1},\hat{V}_{\beta11})\pi_{\lambda}(\boldsymbol{\beta}_{1})}{\int \phi(\hat{\boldsymbol{\beta}}_{1};\boldsymbol{\beta}_{1},\hat{V}_{\beta11})\pi_{\lambda}(\boldsymbol{\beta}_{1})\mathrm{d}\boldsymbol{\beta}_{1}}, \qquad (4.20)$$

where $\hat{V}_{\beta 11}$ is the asymptotic variance-covariance matrix of $\hat{\beta}_1$, defined in (3.12). Once β_1 are sampled from (4.20), we can use the same posterior distribution of \bar{Y} in (3.16) for a given β_1 . Under the Laplace and horseshoe priors, generating posterior samples of β_1 can be easily carried out by simple Gibbs sampling algorithms. The details are given in the Supplementary Material.

Therefore, the approximate posterior distribution of \overline{Y} is obtained as

$$p_{\lambda}(\bar{Y}|\hat{Y}_{reg}(\hat{\boldsymbol{\beta}}_{1}),\hat{\boldsymbol{\beta}}_{1}) = \frac{\int \phi(\hat{Y}_{reg}(\boldsymbol{\beta}_{1});\bar{Y},\hat{V}_{e}(\boldsymbol{\beta}_{1}))\phi_{p}(\hat{\boldsymbol{\beta}}_{1};\boldsymbol{\beta}_{1},\hat{\boldsymbol{V}}_{\beta 11})\pi_{\lambda}(\boldsymbol{\beta}_{1})\pi(\bar{Y}|\boldsymbol{\beta}_{1})d\boldsymbol{\beta}_{1}}{\int \int \phi(\hat{Y}_{reg}(\boldsymbol{\beta}_{1});\bar{Y},\hat{V}_{e}(\boldsymbol{\beta}_{1}))\phi_{p}(\hat{\boldsymbol{\beta}}_{1};\boldsymbol{\beta}_{1},\hat{\boldsymbol{V}}_{\beta 11})\pi_{\lambda}(\boldsymbol{\beta}_{1})\pi(\bar{Y}|\boldsymbol{\beta}_{1})d\boldsymbol{\beta}_{1}d\bar{Y}}.$$

$$(4.21)$$

Generating posterior samples from (4.21) can be easily carried out via the following two steps:

- 1. For a given λ , generate posterior sample β_1^* of β_1 from (4.20).
- 2. Generate posterior sample of \overline{Y} from (3.16) for given β_1^* .

Remark 1. Let $\hat{\beta}_0^{(R)}$ and $\hat{\beta}_1^{(R)}$ be the estimator of β_0 and β_1 defined as

$$(\hat{\beta}_0^{(R)}, \hat{\boldsymbol{\beta}}_1^{(R)}) = \operatorname{argmin}_{\beta_0, \beta_1} \left\{ \sum_{i \in A} \frac{1}{\pi_i} (y_i - \beta_0 - \boldsymbol{x}_{1i}^t \boldsymbol{\beta}_1)^2 + \mathcal{P}_{\lambda}(\boldsymbol{\beta}_1) \right\}, \quad (4.22)$$

where $P(\boldsymbol{\beta}_1) = -2\log \pi_{\lambda}(\boldsymbol{\beta}_1)$ is the penalty (regularization) term for $\boldsymbol{\beta}_1$ induced from prior $\pi_{\lambda}(\boldsymbol{\beta}_1)$. For example, the Laplace prior for $\pi_{\lambda}(\boldsymbol{\beta}_1)$ leads to the penalty term $P(\boldsymbol{\beta}_1) = 2\lambda \sum_{k=1}^p |\boldsymbol{\beta}_k|$, in which $\hat{\boldsymbol{\beta}}_1^{(R)}$ corresponds to the regularized estimator of $\boldsymbol{\beta}_1$ used in McConville et al. (2017). Since the exponential of $-\sum_{i\in A} \pi_i^{-1}(y_i - \beta_0 - \boldsymbol{x}_i^t \boldsymbol{\beta}_1)^2$ is close to the approximated likelihood $\phi_p((\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_1^t); (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1^t), \hat{\boldsymbol{V}}_{\boldsymbol{\beta}})$ used in the approximated Bayesian method when *n* is large, the mode of the approximated posterior of $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1^t)$ would be close to the frequentist estimator (4.22) as well.

Remark 2. In the frequentist approach, λ is often called the tuning parameter and can be selected via a data-dependent procedure such as cross validation as used in McConville et al. (2017). On the other hand, in the Bayesian approach, we assign a prior distribution on the hyperparameter λ and consider integration with respect to the posterior distribution of λ , which means that uncertainty of the hyperparameter estimation can be taken into account. Specifically, we assign a gamma prior for λ^2 in the Laplace prior and a half-Cauchy prior for λ in the horseshoe prior (4.19). They both lead to familiar forms of full conditional posterior distributions

of λ or λ^2 . The details are given in the Supplementary Material.

As in Section 3, we obtain the following asymptotic properties of the proposed approximate Bayesian method.

Theorem 2. Under the regularity conditions described in the Supplementary Material, conditional on the full sample data,

$$\sup_{\bar{Y}\in\Theta_{Y}}\left|p_{\lambda}(\bar{Y}|\hat{Y}_{\text{reg}}(\hat{\boldsymbol{\beta}}_{1}),\hat{\boldsymbol{\beta}}_{1})-\phi(\bar{Y};\hat{Y}_{\text{reg}}(\hat{\boldsymbol{\beta}}_{1}^{(R)}),\hat{V}_{e}(\hat{\boldsymbol{\beta}}_{1}^{(R)}))\right|\to0,\qquad(4.23)$$

in probability as $n \to \infty$ while p is fixed, and $n/N \to f \in [0, 1)$, where Θ_Y is some Borel set for \bar{Y} , $p_{\lambda}(\bar{Y}|\hat{Y}_{reg}(\hat{\beta}_1), \hat{\beta}_1)$ is given in (4.21).

The sketched proof is given in the Supplementary Material. Theorem 2 ensures that the proposed approximate Bayesian method is asymptotically equivalent to the frequentist version in which β_1 is estimated by the regularized method with penalty corresponding to the shrinkage prior used in the Bayesian method. Moreover, the proposed Bayesian method can be extended to cases using general non-linear regression, as demonstrated in the next section.

5. An Extension to non-linear models

The proposed Bayesian methods can be readily extended to work with nonlinear regression. Some extensions of the regression estimator to nonlinear models are also considered in Wu and Sitter (2001), Breidt et al. (2005), and Montanari and Ranalli (2005).

We consider a general working model for y_i as $E(y_i | \boldsymbol{x}_i) = m(\boldsymbol{x}_i; \boldsymbol{\beta}) = m_i$ and $Var(y_i | \boldsymbol{x}_i) = \sigma^2 a(m_i)$ for some known functions $m(\cdot; \cdot)$ and $a(\cdot)$. The model-assisted regression estimator for \bar{Y} with $\boldsymbol{\beta}$ known is then

$$\hat{Y}_{\text{reg,m}}(\boldsymbol{\beta}) = \frac{1}{N} \left\{ \sum_{i=1}^{N} m(\boldsymbol{x}_{i}; \boldsymbol{\beta}) + \sum_{i \in A} \frac{1}{\pi_{i}} (y_{i} - m(\boldsymbol{x}_{i}; \boldsymbol{\beta})) \right\},\$$

and its design-consistent variance estimator is obtained by

$$\hat{V}_{e,m}(\boldsymbol{\beta}) = \frac{1}{N^2} \sum_{i \in A} \sum_{j \in A} \frac{\Delta_{ij}}{\pi_{ij}} \frac{1}{\pi_i} \frac{1}{\pi_j} \{y_i - m(\boldsymbol{x}_i; \boldsymbol{\beta})\} \{y_j - m(\boldsymbol{x}_j; \boldsymbol{\beta})\}$$

which gives the approximate conditional posterior distribution of \overline{Y} given β . That is, similarly to (3.16), we can obtain

$$p(\bar{Y}|\hat{Y}_{\text{reg,m}}(\boldsymbol{\beta}),\boldsymbol{\beta}) \propto \phi(\hat{Y}_{\text{reg,m}}(\boldsymbol{\beta});\bar{Y},\hat{V}_{e,m}(\boldsymbol{\beta}))\pi(\bar{Y}\mid\boldsymbol{\beta}).$$
(5.24)

To generate the posterior values of $\boldsymbol{\beta}$, we first find a design-consistent estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$. Note that a consistent estimator $\hat{\boldsymbol{\beta}}$ can be obtained by solving

$$\hat{U}(\boldsymbol{\beta}) \equiv \sum_{i \in A} \pi_i^{-1} \{ y_i - m(\boldsymbol{x}_i; \boldsymbol{\beta}) \} h(\boldsymbol{x}_i; \boldsymbol{\beta}) = 0,$$

where $h(\boldsymbol{x}_i; \boldsymbol{\beta}) = (\partial m_i / \partial \boldsymbol{\beta}) / a(m_i)$. For example, for binary y_i , we may use a logistic regression model with $m(\boldsymbol{x}_i; \boldsymbol{\beta}) = \exp(\boldsymbol{x}_i^t \boldsymbol{\beta}) / \{1 + \exp(\boldsymbol{x}_i^t \boldsymbol{\beta})\}$ and $\operatorname{Var}(y_i) = m_i(1 - m_i)$, which leads to $h(\boldsymbol{x}_i; \boldsymbol{\beta}) = \boldsymbol{x}_i$. Under some regularity conditions, we can establish the asymptotic normality of $\hat{\beta}$. That is,

$$\hat{\boldsymbol{V}}_{\beta}^{-1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \mid \boldsymbol{\beta} \stackrel{\mathcal{L}}{\longrightarrow} N(0,I),$$

where

$$\hat{\boldsymbol{V}}_{\beta} = \left\{ \sum_{i \in A} \frac{1}{\pi_i} \hat{\boldsymbol{h}}_i \dot{\boldsymbol{m}}(\boldsymbol{x}_i; \hat{\boldsymbol{\beta}})^t \right\}^{-1} \left(\sum_{i \in A} \sum_{j \in A} \frac{\Delta_{ij}}{\pi_{ij}} \frac{\hat{e}_i \hat{\boldsymbol{h}}_i}{\pi_i} \frac{\hat{e}_j \hat{\boldsymbol{h}}_j^t}{\pi_j} \right) \left\{ \sum_{i \in A} \frac{1}{\pi_i} \hat{\boldsymbol{h}}_i \dot{\boldsymbol{m}}(\boldsymbol{x}_i; \hat{\boldsymbol{\beta}})^t \right\}^{-1}$$
with $\hat{e}_i = y_i - \boldsymbol{m}(\boldsymbol{x}_i; \hat{\boldsymbol{\beta}}), \ \hat{\boldsymbol{h}}_i = h(\boldsymbol{x}_i; \hat{\boldsymbol{\beta}}), \text{ and } \dot{\boldsymbol{m}}(\boldsymbol{x}; \boldsymbol{\beta}) = \partial \boldsymbol{m}(\boldsymbol{x}; \boldsymbol{\beta}) / \partial \boldsymbol{\beta}.$ Note that $\dot{\boldsymbol{m}}(\boldsymbol{x}; \boldsymbol{\beta}) = m_i (1 - m_i) \boldsymbol{x}_i$ under a logistic regression model.

Thus, the posterior distribution of β given $\hat{\beta}$ can be obtained by

$$p(\boldsymbol{\beta} \mid \hat{\boldsymbol{\beta}}) \propto \phi(\hat{\boldsymbol{\beta}} \mid \boldsymbol{\beta}, \hat{\boldsymbol{V}}_{\beta}) \pi(\boldsymbol{\beta}).$$
 (5.25)

We can use a shrinkage prior $\pi(\beta)$ for β in (5.25) if necessary. Once β^* is generated from (5.25), the posterior values of \bar{Y} are generated from (5.24) for a given β^* .

This formula enables us to define the approximate posterior distribution of β of the form (3.13), so that the approximate Bayesian inference for \bar{Y} can be carried out in the same way as in the linear regression case. Note that Theorem 1 still holds under the general setup as long as the regularity conditions given in the Supplementary Material are satisfied.

6. Simulation

We investigate the performance of the proposed approximate Bayesian methods against standard frequentist methods using two limited simulation studies. In the first simulation, we consider a linear regression model for a continuous y variable. In the second simulation, we consider a binary y and apply the logistic regression model for the non-linear regression estimation.

In the first simulation, we generate $x_i = (x_{i1}, \ldots, x_{ip^*})^t$, $i = 1, \ldots, N$, from a multivariate normal distribution with mean vector $(1, \ldots, 1)^t$ and variance-covariance matrix 2R(0.2), where $p^* = 50$ and the (i, j)-th element of $R(\rho)$ is $\rho^{|i-j|}$. The response variables Y_i are generated from the following linear regression model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p^*} x_{ip^*} + \varepsilon_i, \quad i = 1, \dots, N,$$

where N = 10,000, $\varepsilon_i \sim N(0,2)$, $\beta_1 = 1$, $\beta_4 = -0.5$, $\beta_7 = 1$, $\beta_{10} = -0.5$ and the other β_k 's are set to zero. For the dimension of the auxiliary information, we consider four scenarios for p of 20, 30, 40 and 50. For each p, we assume that we can access only $(x_{i1}, \ldots, x_{ip})^t$ a subset of the full information $(x_{i1}, \ldots, x_{ip^*})^t$. Note that for all scenarios the auxiliary variables significantly related with Y_i are included, and so only the amount of irrelevant information gets larger as p gets larger. We selected a sample size of n = 300 from the finite population, using two sampling mechanism: (A) simple random sampling (SRS) and (B) probability-proportional-tosize sampling (PPS) with size measure $z_i = \max\{\log(1 + |Y_i + e_i|), 1\}$ with $e_i \sim \text{Exp}(2)$. The parameter of interest is $\bar{Y} = N^{-1} \sum_{i=1}^{N} Y_i$. We assume that $\bar{X}_k = N^{-1} \sum_{i=1}^{N} x_{ik}$ is known for all $k = 1, \ldots, p$.

For the simulated dataset, we apply the proposed approximate Bayesian methods with the uniform prior $\pi(\beta_1) \propto 1$, Laplace prior and horseshoe prior (4.19) for β_1 , which are denoted by AB, ABL and ABH, respectively. For all the Bayesian methods, we use $\pi(\bar{Y}|\boldsymbol{\beta}_1) \propto 1$. We generate 5,000 posterior samples of \bar{Y} after discarding the first 500 samples and compute the posterior mean of \bar{Y} as the point estimate. As for the frequentist methods, we apply the original generalized regression estimator without variable selection (GREG) as well as the GREG method with Lasso regularization (GREG-L; McConville et al., 2017), ridge estimation of β_1 (GREG-R; Rao and Singh, 1997) and forward variable selection (GREG-V) using adjusted coefficient of determination. We also adopted the mixed modeling approach to the GREG estimation (GREG-M; Park and Fuller, 2009) which is similar to GREG-R. Moreover, the HT estimator is employed as a benchmark for efficiency comparison. In GREG-L, the tuning parameter is selected via 10-fold cross validation, and we use the gamma prior $\operatorname{Ga}(\lambda_*^2, 1)$ for λ^2 in ABL, where λ_* is the selected value for λ in GREG-L. In ABH, we assign the half-Cauchy prior HC(0, 1) for the tuning parameter λ^2 . Based on 1,000 replications, we calculate the square root of mean squared errors (RMSE) and bias of point estimators which are reported in Table 2. We also evaluated the performance of 95% confidence (credible) intervals using coverage probabilities (CP) and the average length (AL), which are shown in Table 3.

Table 2 shows that RMSE and bias of AB and GREG are almost identical, which is consistent with the fact that AB is a Bayesian version of GREG. Moreover, the results show that the existing shrinkage methods such as GREG-L and the proposed Bayesian methods ABL and ABH tend to produce smaller RMSEs and smaller absolute biases than GREG or AB as p increases, which indicates the importance of suitable selection of auxiliary variables when p is large. From Table 3, it is observed that the CPs of GREG decreases as p increases and are significantly smaller than the nominal level since GREG ignores the variability in estimating β and the variability increases as p increases. On the other hand, the Bayesian version AB can take account of the variability in estimating β and the CPs are around the nominal level and ALs of AB are larger than those of GREG. Although the performance of GREG-L is much better than GREG due to the shrinkage techniques, the CPs are not necessarily close to the nominal level. Note that GREG-M takes account of the variability in estimating β , but not in other parameters, thereby the coverage performance is limited. It is also confirmed that the proposed ABH and ABL methods produce narrower intervals than AB.

In the second simulation study, we consider the binary case for y_i and apply the non-linear regression method discussed in Section 5. The binary response variables Y_i are generated from the following logistic regression model:

$$Y_i \sim \text{Ber}(\delta_i), \quad \log\left(\frac{\delta_i}{1-\delta_i}\right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \quad i = 1, \dots, N,$$

where $\beta_0 = -1$ and the other settings are the same as the linear regression case. We selected a sample size of n = 300 from the finite population, using two sampling mechanism: (A) simple random sampling and (B) probabilityproportional-to-size sampling with size measure $z_i = \max\{\log(1 + 0.5Y_i + e_i), 0.5\}$ with $e_i \sim \exp(3)$. We again apply the three Bayesian methods and three frequents methods, GREG, GREG-L and GREG-R, based on a logistic regression model to obtain point estimates and confidence/credible intervals of the population mean $\bar{Y} = N^{-1} \sum_{i=1}^{N} Y_i$. The obtained RMSE and bias of point estimates and CP and AL of intervals based on 1,000 replications are reported in Tables 4 and 5, respectively, which also shows again the superiority of the proposed Bayesian approach to the frequentist approach in terms of uncertainty quantification.

In the Supplementary Material, we report additional simulation results under larger sample sizes or different data generation scenarios.

7. Example

We applied the proposed methods to the synthetic income data available from the sae package (Molina and Marhuenda, 2015) in R. In the dataset, the normalized annual net income is observed for a certain number of individuals in each province of Spain. The dataset contains 9 covariates; four indicators of the four groupings of ages (16 - 24, 25 - 49, 50 - 64 and ≥ 65 denoted by ag1,...,ag4, respectively), the indicator of having Spanish nationality na, the indicators of education levels (primary education ed1 and post-secondary education ed2), and the indicators of two employment categories (employed em1 and unemployed em2). We also adopted 13 interaction variables: ag1*na, ag2*na, ag3*na, ag4*na, ag2*ed1, ag3*ed1, ag4*ed1, ag1*em1, ag2*em1, ag3*em1, ag4*em1, ed1*em1 and ed2*em1, as auxiliary variables, thereby p = 22 in this example. The dataset also contains information of survey weights, so that we used its inverse value as the sampling probability. Since there is no information regarding the details of sampling mechanism, we approximate the joint inclusion probability as the product of two sampling probabilities. In this example, we focus on estimating average income in three provinces, Palencia, Segovia and Soria, where the number of sampled units are 72, 58 and 20, respectively. The number of non-sampled units were around 10^6 . It should be noted that the number of sample sizes are not so large compared with the number of auxiliary variables, especially in Soria. Hence, the estimation error of regression coefficients would not be negligible and the proposed Bayesian methods would be appealing in this case.

In order to perform joint estimation and inference in the three provinces, we employed the following working model:

$$y_i = \alpha + \sum_{h \in \{1,2,3\}} x_{0i}^{(h)} \beta_0^{(h)} + \boldsymbol{x}_i^t \boldsymbol{\beta}_1 + e_i, \qquad (7.26)$$

where α is an intercept term, $x_{0i}^{(h)} = 1$ if *i* belong to province *h*, where h = 1 for Palencia, h = 2 for Segovia, and h = 3 for Soria, and \boldsymbol{x}_i is the vector of auxiliary variables with dimension p = 22 (9 auxiliary variables and 13 interaction variables). Here y_i is the log-transformed net income and e_i is the error term.

Under the working model (7.26), the posterior distribution of \bar{Y}_h is

$$p\{\bar{Y}_h \mid \hat{\bar{Y}}_{h,\mathrm{reg}}(\beta_0^{(h)},\boldsymbol{\beta}_1), \beta_0^{(h)},\boldsymbol{\beta}_1\} \propto \phi(\hat{\bar{Y}}_{h,\mathrm{reg}}(\beta_0^{(h)},\boldsymbol{\beta}_1) \mid \bar{Y}_h, \hat{V}_{e,h}(\boldsymbol{\beta}))\pi(\bar{Y}_h)$$

where

$$\hat{\bar{Y}}_{h,\text{reg}} = \hat{\beta}_0^{(h)} + \bar{\boldsymbol{X}}_h^t \hat{\boldsymbol{\beta}}_1 + \frac{1}{N_h} \sum_{i \in A_h} \frac{1}{\pi_i} \left(y_i - \hat{\beta}_0^{(h)} - \boldsymbol{x}_i^t \hat{\boldsymbol{\beta}}_1 \right),$$

and

$$\hat{V}_{e,h}(\boldsymbol{\beta}) = \frac{1}{N_h^2} \sum_{i \in A_h} \sum_{j \in A_h} \frac{\Delta_{ij}}{\pi_{ij}} \frac{1}{\pi_i} \frac{1}{\pi_j} \left(y_i - \beta_0^{(h)} - \boldsymbol{x}_i^t \boldsymbol{\beta}_1 \right) \left(y_j - \beta_0^{(h)} - \boldsymbol{x}_j^t \boldsymbol{\beta}_1 \right)$$

Based on the above formulas, we performed the proposed approximate Bayesian methods for \bar{Y}_h for each h, and computed 95% credible intervals for the log-transformed average income with 5000 posterior samples after discarding the first 500 samples as burn-in period. We considered three types of priors for β_1 , flat, Laplace and horseshoe priors as considered in Section 6, where we adopted the same priors for the hyperparameters in the Laplace and horseshoe priors. In the Laplace priors, we applied two different priors for the hyperparameter λ^2 , given by Ga(1, 1) and Ga(1/p, 1), but the results were almost the same. We also calculated 95% confidence intervals of the log-transformed average income based on the two frequentist methods, GREG and GREG-L, using the working model (7.26). In applying GREG-L, the tuning parameter in the Lasso estimator was selected via 10 fold cross validation.

The 95% credible intervals of β_1 based on the approximate posterior distributions under Laplace and horseshoe priors are shown in Figure 1, in which the design-consistent and Lasso estimates of β_1 are also given. It is observed that the approximate posterior mean of β_1 shrinks the designconsistent estimates of β_1 toward 0 although exactly zero estimates are not produced as the frequentist Lasso estimator does. The Lasso estimate selects only one variable among 22 candidates, and the variable is also significant in terms of the credible interval in both two priors. Moreover, the two Bayesian methods detect one or two more variables to be significant judging from the credible intervals. Comparing the results from two priors, the horseshoe prior provides narrower credible intervals than the Laplace prior.

In Figure 2, we show the resulting credible and confidence intervals of the average income in the three provinces. It is observed that the proposed Bayesian methods, AB and ABL, tend to produce wider credible intervals than the confidence intervals of the corresponding frequencies methods, GREG and GREG-L, respectively, which is consistent with the simulation results in Section 6. We can also confirm that the credible intervals of ABH are slightly narrower than those of ABL, which would reflect the differences of interval lengths of β_1 as shown in Figure 1.



Figure 1: 95% credible intervals of regression coefficients under Laplace (left) and horseshoe (right) priors.

8. Concluding Remarks

We proposed an approximate Bayesian method for model-assisted survey estimation using parametric regression models as working models. The proposed method is justified under the frequentist framework. A main advantage of the proposed method is that it can naturally implement a



Figure 2: 95% confidence and credible intervals for average income based on five methods in three provinces in Spain.

shrinkage prior for regularized regression estimation, which not only provides an efficient point estimator, but also fully captures the uncertainty associated with model selection and parameter estimation via a Bayesian framework. Although we only consider two popular prior distributions, the Laplace prior and the horseshoe prior, other priors, such as the spikeand-slab prior (Ishwaran and Rao, 2005), can be adopted in the same way.

Further investigation will be an important research topic in the future.

Although our working model is parametric, the proposed Bayesian method can be applied to other semiparametric models such as local polynomial model (Breidt and Opsomer, 2000), P-spline regression model (Breidt et al., 2005), or a neural network model (Montanari and Ranalli, 2005). By finding suitable prior distributions for the semiparametric models, the model complexity parameters will be determined automatically and the uncertainty will be captured in the approximate Bayesian framework.

Finally, under more complicated sampling design such as multi-stage stratified cluster sampling, the main idea can be applied similarly since the proposed Bayesian method relies on the sampling distribution of the GREG estimator, which is asymptotically normal as shown by Krewski and Rao (1981). If the asymptotic normality is questionable, one can use a weighted likelihood bootstrap to approximate Bayesian posterior, as in Lyddon et al. (2019). Such extensions are beyond the scope of this paper and will be considered in the future.

Supplementary Materials

Supplementary Material includes technical details for posterior computation, proofs of theorems and additional results of simulation studies.

Acknowledgement

We thank the AE and three anonymous referees for very constructive comments. The first author was supported by Japan Society for the Promotion of Science KAKENHI grant number JP18K12757. The second author was supported by US National Science Foundation (MMS-1733572).

References

- Breidt, F. J., G. Claeskens, and J. D. Opsomer (2005). Model-assisted estimation for complex surveys using penalised splines. *Biometrika 92*, 831–846.
- Breidt, F. J. and J. D. Opsomer (2000). Local polynomial regression estimators in survey sampling. *Annals of Statistics 28*, 403–427.
- Breidt, F. J. and J. D. Opsomer (2017). Model-assisted survey estimation with modern prediction techniques. *Statistical Science* 32, 190–205.
- Carvalho, C. M., N. G. Polson, and J. G. Scott (2009). Handling sparsity via the horseshoe. Proceedings of the 12th International Confe- rence on Artificial Intelligence and Statistics (AISTATS 2009).

- Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika* 97, 465–480.
- Chen, J. K. T., R. L. Valliant, and M. R. Elliott (2018). Model-assisted calibration of non-probability sample survey data using adaptive LASSO. *Survey Methodology* 44, 117–144.
- Dempster, A. P., M. Schatzoff, and N. Wermuth (1977). A simulation study of alternatives to ordinary least squares. Journal of the American Statistical Association 72, 77–91.
- Deville, J. C. and C. E. Särndal (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association* 87, 376–382.
- Fuller, W. A. (2002). Regression estimation for sample surveys. Survey Methodology 28, 5–23.

Fuller, W. A. (2009). Sampling Statistics. Wiley.

- Godambe, V. P. and V. M. Joshi (1965). Admissibility and Bayes estimation in sampling finite populations, 1. Annals of Mathematical Statistics 36, 1707–1722.
- Hoerl, E. and R. W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12, 55–67.

REFERENCES33

- Horvitz, D. G. and D. J. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47, 663–685.
- Isaki, C. T. and W. A. Fuller (1982). Survey design under the regression superpopulation model. Journal of the American Statistical Association 77, 89–96.
- Ishwaran, H. and J. S. Rao (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *Annals of Statistics 33*, 730–773+.
- Kim, J. K. and M. Park (2010). Calibration estimation in survey sampling. International Statistical Review 78, 21–39.
- Krewski, D. and J. N. K. Rao (1981). Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods. *Annals of Statistics 9*, 1010–1019.
- Little, R. J. (2004). To model or not to model? competing modes of inference for finite population sampling. *Journal of the American Statistical Association 99*, 546–556.
- Little, R. J. A. (2012). Calibrated Bayes, an alternative inferential paradigm for official statistics. *Journal of Official Statistics* 28, 309–334.

- Lyddon, S. P., C. C. Holmes, and S. G. Walker (2019). General bayesian updating and the loss-likelihood bootstrap. *Biometrika* 106, 465–478.
- McConville, K., F. Breidt, T. Lee, and G. Moisen (2017). Model-assisted survey regression estimation with the LASSO. *Journal of Survey Statistics and Methodology 5*, 131–158.
- Molina, I. and Y. Marhuenda (2015). sae: An R package for small area estimation. *The R Journal* 7, 81–98.
- Montanari, G. E. and M. G. Ranalli (2005). Nonparametric model calibration estimation in survey sampling. Journal of the American Statistical Association 100, 1429–1442.
- Park, M. and W. A. Fuller (2009). The mixed model for survey regression estimation. *Journal of Statistical Planning and Inference 139*, 1320–1331.
- Park, T. and G. Casella (2008). The Bayesian Lasso. Journal of the American Statistical Association 103, 681–686.
- Pfeffermann, D. and M. Sverchkov (1999). Parametric and semiparametric estimation of regression models fitted to survey data. Sankhyā, Series B 61, 166–186.
- Rao, J. and A. Singh (1997). A ridge-shrinkage method for range-restricted

REFERENCES35

weight calibration in survey sampling. In *Proceedings of the Section on* Survey Research Methods, pp. 57–65. American Statistical Association.

- Särndal, C. E. and Lundström (2005). *Estimation in surveys with nonresponse*. Chichester: John Wiley & Sons.
- Silva, P. L. D. N. and C. J. Skinner (1997). Variable selection for regression estimation in finite populations. Survey Methodology 23, 23–32.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B 58, 267–288.
- van der Vaart, A. W. (2000). Asymptotic Statistics. New York: Cambridge University Press.
- Wang, Z., J. K. Kim, and S. Yang (2018). Approximate Bayesian inference under informative sampling. *Biometrika* 105, 91–102.
- Wu, C. and R. R. Sitter (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association 96*, 185–193.
- Zanganeh, S. Z. and R. J. A. Little (2015). Bayesian inference for the finite population total from a heteroscedastic probability proportional to size sample. *Journal of Survey Statistics and Methodology 3*, 162–192.

- Zeng, H. and R. J. A. Little (2003). Penalized spline model-based estimation of the finite populations total from probability-proportional-to-size samples. *Journal of Official Statistics* 19, 99–117.
- Zeng, H. and R. J. A. Little (2005). Inference for the population total from probability-proportional-to-size samples based on predictions from a penalized spline nonparametric model. *Journal of Official Statistics 21*, 1–20.
- Zou, H. (2006). The adaptive Lasso and its oracle properties. Journal of the American Statistical Association 101, 1418–1429.
- Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B* 67, 301–320.

Center for Spatial Information Science, The University of Tokyo E-mail: sugasawa@csis.u-tokyo.ac.jp Department of Statistics, Iowa State University

E-mail: jkim@iastate.edu

Table 2: Square root of Mean squared errors (RMSE) and bias of point estimators under $p \in \{20, 30, 40, 50\}$ in scenarios (A) and (B) with linear regression. All values are multiplied by 100.

_				(1	A)		(B)				
_	Method		20	30	40	50	20	30	40	50	
		GREG	11.7	11.8	12.0	12.3	11.4	11.8	12.1	12.3	
		GREG-L	11.7	11.7	11.7	11.8	11.1	11.1	11.1	11.1	
		GREG-R	11.8	11.9	12.1	12.4	11.4	11.6	11.8	12.0	
		GREG-V	11.6	11.7	11.8	12.0	11.3	11.5	11.8	12.0	
	MSE	GREG-M	11.7	11.8	12.0	12.3	11.4	11.8	12.1	12.3	
		AB	11.7	11.9	12.1	12.4	11.6	11.9	12.2	12.5	
		ABL	11.7	11.8	11.9	12.2	11.4	11.7	11.8	12.0	
		ABH	11.6	11.6	11.6	11.8	11.2	11.3	11.3	11.4	
_		HT	17.5	17.5	17.5	17.5	14.8	14.8	14.8	14.8	
		GREG	0.21	0.12	0.13	0.23	0.54	1.24	1.87	2.41	
		GREG-L	0.19	0.16	0.18	0.19	0.00	0.11	0.20	0.26	
		GREG-R	0.22	0.16	0.18	0.31	0.56	1.21	1.79	2.32	
		GREG-V	0.16	0.05	0.08	0.17	0.29	0.80	1.26	1.64	
	Bias	GREG-M	0.21	0.12	0.13	0.23	0.54	1.24	1.87	2.41	
		AB	0.19	0.10	0.11	0.22	0.60	1.28	1.92	2.44	
		ABL	0.19	0.11	0.11	0.21	0.49	1.06	1.55	1.95	
		ABH	0.16	0.12	0.11	0.17	0.06	0.29	0.51	0.71	
		HT	0.78	0.78	0.78	0.78	-1.08	-1.08	-1.08	-1.08	

Table 3: Coverage probabilities (CP) and average lengths (AL) of 95% confidence/credible intervals under $p \in \{20, 30, 40, 50\}$ in scenarios (A) and (B) with linear regression. All values are multiplied by 100.

			(1	A)		(B)					
	Method	20	30	40	50	20	30	40	50		
	GREG	92.8	92.8	92.7	89.9	94.2	92.1	92.1	90.1		
	GREG-L	93.5	93.4	93.2	93.3	94.5	94.8	94.4	94.8		
	GREG-R	93.0	92.4	91.8	90.0	93.3	92.4	91.9	90.4		
	GREG-V	93.6	93.7	93.3	91.4	94.1	93.8	92.5	91.2		
CP	GREG-M	93.9	93.9	93.9	92.9	94.5	93.7	93.8	92.9		
	AB	95.3	94.8	94.9	94.2	95.1	94.8	94.9	95.2		
	ABL	95.2	94.6	94.8	94.5	95.3	95.3	95.1	94.9		
	ABH	94.8	95.0	95.0	94.7	95.4	95.9	95.1	95.5		
	HT	94.5	94.5	94.5	94.5	95.2	95.2	95.2	95.2		
	GREG	43.1	42.3	41.5	40.7	43.1	42.3	41.5	40.7		
	GREG-L	43.8	43.7	43.6	43.5	43.3	43.1	42.9	42.8		
	GREG-R	43.2	42.5	41.9	41.4	42.8	42.0	41.3	40.7		
	GREG-V	43.4	42.8	42.2	41.6	43.4	42.9	42.3	41.8		
AL	GRREG-M	44.2	44.2	44.3	44.4	44.3	44.4	44.6	44.8		
	AB	45.8	46.3	46.8	47.3	46.2	47.0	47.8	48.7		
	ABL	45.6	45.9	46.1	46.3	45.8	46.4	46.8	47.3		
	ABH	45.1	45.2	45.2	45.1	45.2	45.4	45.4	45.6		
	HT	66.4	66.4	66.4	66.4	59.1	59.1	59.1	59.1		

Table 4: Square root of Mean squared errors (RMSE) and bias of point estimators under $p \in \{20, 30, 40, 50\}$ in scenarios (A) and (B) with logistic regression. All values are multiplied by 100.

		(A)					(B)			
	Method	20	30	40	50	2	20	30	40	50
	GR	2.24	2.29	2.32	2.36	2	.32	2.39	2.50	2.57
	GRL	2.17	2.18	2.19	2.20	2	.27	2.29	2.31	2.30
	GRR	2.22	2.26	2.29	2.31	2	.32	2.38	2.44	2.49
RMSE	AB	2.23	2.26	2.28	2.30	2	.31	2.37	2.45	2.50
	ABL	2.21	2.23	2.24	2.25	2	.27	2.28	2.26	2.23
	ABH	2.18	2.20	2.23	2.26	2	.26	2.27	2.28	2.32
	HT	2.80	2.80	2.80	2.80	2	.83	2.83	2.83	2.83
	GR	-0.10	-0.12	-0.12	-0.11	0	.10	0.18	0.31	0.43
	GRL	-0.11	-0.11	-0.10	-0.11	0.	.03	0.05	0.07	0.08
	GRR	-0.11	-0.12	-0.12	-0.12	0.	.07	0.13	0.20	0.27
Bias	AB	-0.11	-0.13	-0.13	-0.13	0.	.09	0.17	0.27	0.38
	ABL	-0.10	-0.10	-0.07	-0.02	0	.07	0.13	0.19	0.22
	ABH	-0.10	-0.11	-0.10	-0.11	0	.01	0.03	0.04	0.03
	HT	-0.15	-0.15	-0.15	-0.15	0	.07	0.07	0.07	0.07

Table 5: Coverage probabilities (CP) and average lengths (AL) of 95% credible/confidence intervals under $p \in \{20, 30, 40, 50\}$ in scenarios (A) and (B) with logistic regression. All values are multiplied by 100.

			(1		(I	3)			
	Method	20	30	40	50	20	30	40	50
	GR	92.3	90.8	88.8	86.4	91.9	90.3	87.3	84.6
	GRL	94.1	94.1	93.9	93.2	93.2	93.0	92.6	92.9
	GRR	92.8	92.1	91.0	90.6	92.0	90.8	89.6	89.0
CP	AB	94.8	95.5	95.4	96.1	94.6	94.1	94.5	95.1
	ABL	95.1	95.7	95.9	96.5	94.6	95.2	96.6	97.2
	ABH	95.1	96.0	96.0	96.2	95.1	95.2	95.9	96.2
	HT	95.3	95.3	95.3	95.3	94.5	94.5	94.5	94.5
	GR	8.02	7.80	7.56	7.30	8.20	7.95	7.69	7.39
	GRL	8.21	8.17	8.14	8.11	8.42	8.37	8.33	8.30
	GRR	8.15	7.99	7.88	7.79	8.34	8.17	8.04	7.94
AL	AB	8.74	8.90	9.10	9.42	9.05	9.27	9.59	10.10
	ABL	8.79	8.99	9.24	9.55	9.07	9.31	9.61	9.99
	ABH	8.76	8.96	9.18	9.45	9.02	9.22	9.46	9.75
	HT	11.14	11.14	11.14	11.14	11.00	11.00	11.00	11.00