An Approximate Bayesian Approach to Regression Estimation with Many Auxiliary Variables

Shonosuke Sugasawa[1] Jae Kwang Kim[2]

June 12, 2019

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 shinkage prior for model selection. The proposed Bayesian approach enables us to get not only efficient point estimates but also reasonable credible intervals for population means. Results from two limited simulation studies are presented to facilitate comparison with existing frequentist methods.

Keywords: 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, the Hotvitz-Thompson estimator obtained from a probability sample is often used, which satisfies consistency and the resulting inference is justified from the randomization perspective (Horvitz and Thompson, 1952). However, the Horvitz-Thompson estimator uses the first-order inclusion probability only and

[1] Center for Spatial Information Science, The University of Tokyo

[2] Department of Statistics, Iowa State University

does not fully incorporate all available information from the finite population. To improve efficiency, regression estimation is often used to incorporate auxiliary information in survey sampling. Deville and Sa¨rndal (1992), Fuller (2002), Kim and Park (2010), and Breidt and Opsomer (2017) present comprehensive overviews of such variants of regression estimation in survey sampling.

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

$$y_i = x^t_i \beta + e_i, \tag{1}$$

where $E(e_i) = 0$ and $Var(e_i) = \sigma^2$. The superpopulation model does not necessarily hold in the sample as the sampling design can be informative in the sense of Pfeffermann and Sverchkov (1999). Under the regression superpopulation model in (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 variance under the joint distribution of the sampling design and the superpopulation model in (1).

However, the above optimality of the regression estimator is untenable if the dimension of the auxiliary variables x is large. 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 the uncertainty can fail to be fully accounted for by the standard linearization variance estimation, resulting in misleading inference.

To overcome the problem, S¨arndal and Lundstro¨m (2005) select a subset of the auxiliary variables for regression estimation. The classical model selection

approach is based on a step-wise method. However, the step-wise methods will not necessarily produce the best model if there are redundant predictors. Another approach is to employ regularized estimation of regression coefficients. For example, McConville et al. (2017) propose a regularized regression estimation approach based on the LASSO penalty of Tibshirani (1996). However, there are two main problems in the regularization approach. First, the choice of the regularization parameter is somewhat unclear. Second, after model selection, the frequentist inference is notoriously difficult to make.

In this paper, we propose a unified Bayesian framework to handle regularized regression estimation. We first present a Bayesian approach for regression estimation when $p = \dim(x)$ 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 better coverage properties. Second, the proposed Bayesian method solves the problem of large p in regularized regression estimation.

The penalty function for regularization is incorporated into the prior distribution and the uncertainty associated with model selection and parameter estimation is fully captured in the Bayesian machinery. Furthermore, the penalty parameter λ can be optimized by having its own prior distribution. The proposed method provides a unified approach to Bayesian inference with sparse regression estimation. It is a calibrated Bayesian (Little, 2012) in the sense that it is asymptotically equivalent to the frequentist design-based approach.

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. In Section 4, the proposed method is extended to handle 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.

2 Basic setup

Consider a finite population of a known size N. Associated with unit i in the finite population, we consider measurement (x_i^t, y_i) where 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 $\{x_i, y_i\}_{i\in A}$ from the sample. The Horvitz-Thompson estimator $\hat{\bar{Y}}_{HT} = N^{-1}\sum_{i\in A}\pi_i^{-1} y_i$ is design unbiased but it is not necessarily efficient.

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

$$\hat{\bar{Y}}_{\mathrm{reg}} = \bar{\boldsymbol{X}}^t\hat{\boldsymbol{\beta}} + \frac{1}{N}\sum_{i\in A}\frac{1}{\pi_i}\left(y_i - \boldsymbol{x}_i^t\hat{\boldsymbol{\beta}}\right) \tag{2}$$

where π_i is the first-order inclusion probability of unit i, and $\hat{\beta}$ is an estimator of β in (1). Typically, we use $\hat{\beta}$ obtained by minimizing the weighted quadratic loss

$$Q(\beta) = \sum_{i\in A}\pi^{-1}(y_i - x_i^t\beta)^2, \tag{3}$$

motivated from model (1).

To derive the asymptotic properties of $\hat{\bar{Y}}_{\mathrm{reg}}$, we may use

$$\begin{aligned}\hat{\bar{Y}}_{\mathrm{reg}} - \bar{Y} &= \hat{\bar{Y}}_{\mathrm{HT}} - \bar{Y} + \left(\bar{\boldsymbol{X}} - \hat{\bar{\boldsymbol{X}}}_{\mathrm{HT}}\right)^t\hat{\boldsymbol{\beta}} \\ &= \hat{\bar{Y}}_{\mathrm{HT}} - \bar{Y} + \left(\bar{\boldsymbol{X}} - \hat{\bar{\boldsymbol{X}}}_{\mathrm{HT}}\right)^t\boldsymbol{\beta}_* + R_n\end{aligned} \tag{4}$$

where $\hat{\bar{X}}_{\mathrm{HT}} = N^{-1} \sum_{i \in A} \pi_i^{-1} x_i$ and

$$R_n = \left(\bar{\boldsymbol{X}} - \hat{\bar{\boldsymbol{X}}}_{\mathrm{HT}}\right)^t \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*\right)$$

for any β_*. If we choose $\beta_* = \mathrm{plim}_{n\to\infty} \hat{\beta}$ and the dimension p is fixed in the asymptotic setup, then we can obtain $R_n = O_p(n^{-1})$ and safely use the main terms of (4) to describe the asymptotic behavior of $\hat{\bar{Y}}_{\mathrm{reg}}$. To emphasize its dependence on $\hat{\beta}$ in the regression estimator, we can write $\hat{\bar{Y}}_{\mathrm{reg}} = \hat{\bar{Y}}_{\mathrm{reg}}(\hat{\beta})$. Roughly speaking, we can obtain

$$\sqrt{n}\left(\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\beta}) - \hat{\bar{Y}}_{\mathrm{reg}}(\beta_*)\right) = O_p(n^{-1/2}p) \tag{5}$$

and, if $p = o(n^{1/2})$ then we can safely ignore the effect of estimating β_* in the regression estimator. See Appendix A for a sketch proof of (5).

If, on the other hand, the dimension p is large, then we cannot ignore the effect of estimating β_*. 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

$$\hat{\beta}^{(R)} = \mathrm{argmin}_{\beta}\{Q(\beta) + p_\lambda(\beta)\}, \tag{6}$$

where $Q(\beta)$ is defined in (3) and $p_\lambda(\beta)$ is a penalty function with parameter λ. Some popular penalty functions are presented in Table 1. Once the solution to (6) is obtained, then the regularized regression estimator is given by

$$\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}}^{(R)}) = \bar{\boldsymbol{X}}^t \hat{\boldsymbol{\beta}}^{(R)} + \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \left(y_i - \boldsymbol{x}_i^t \hat{\boldsymbol{\beta}}^{(R)}\right) \tag{7}$$

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)	$p_\lambda(\boldsymbol{\beta}) = \lambda \sum_{j=1}^{p} \lvert\beta_j\rvert$
Adaptive LASSO	Zou (2006)	$p_\lambda(\boldsymbol{\beta}) = \lambda \sum_{j=1}^{p} \left(\lvert\beta_j\rvert \,/\, \left\lvert\hat{\beta}_j\right\rvert \right)$
Elastic Net	Zou and Hastie (2005)	$p_\lambda(\boldsymbol{\beta}) = \lambda_1 \sum_{j=1}^{p} \lvert\beta_j\rvert + \lambda_2 \sum_{j=1}^{p} \beta_j^2$

Statistical inference with the regularized regression estimator in (7) 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 the sampling design is non-informative and the uncertainty in model selection is not fully incorporated in their inference. Generally speaking, making frequentist inference after model selection is difficult. The approximated Bayesian method we propose in this paper will capture the full uncertainty in the Bayesian framework.

3 Approximate Bayesian survey regression estimation

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

$$p(\theta \mid \hat{\theta}) = \frac{f(\hat{\theta} \mid \theta)\pi(\theta)}{\int f(\hat{\theta} \mid \theta)\pi(\theta)\mathrm{d}\theta}, \tag{8}$$

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

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

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

$$\hat{V}_\beta = \left(\sum_{i \in A} \pi_i^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^t \right)^{-1} \left(\sum_{i \in A} \sum_{j \in A} \frac{\Delta_{ij}}{\pi_{ij}} \frac{\hat{e}_i \boldsymbol{x}_i}{\pi_i} \frac{\hat{e}_j \boldsymbol{x}_j^t}{\pi_j} \right) \left(\sum_{i \in A} \pi_i^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^t \right)^{-1}, \quad (9)$$

where $\hat{e}_i = y_i - \boldsymbol{x}_i^t \hat{\beta}$, $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$ 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

$$\hat{V}_\beta^{-1/2} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right) \mid \boldsymbol{\beta} \xrightarrow{\mathcal{L}} N(0, I) \quad (10)$$

as $n \to \infty$, where

$$\boldsymbol{\beta} = \left(\sum_{i=1}^{N} \boldsymbol{x}_i \boldsymbol{x}_i^t \right)^{-1} \sum_{i=1}^{N} \boldsymbol{x}_i y_i .$$

Thus, using (8) and (10), we can obtain the approximate posterior distribution of β as

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

where φ_p denotes a p-dimensional multivariate normal density and $\pi(\beta)$ is a prior distribution for β.

Now, we wish to find the posterior distribution of $\bar{Y}$ for a given β. First, define

$$\hat{\bar{Y}}_{\mathrm{reg}}(\boldsymbol{\beta}) = \bar{\boldsymbol{X}}^t \boldsymbol{\beta} + \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \left(y_i - \boldsymbol{x}_i^t \boldsymbol{\beta} \right) .$$

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

$$\frac{\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}) - \bar{Y}}{\sqrt{\hat{V}_e(\boldsymbol{\beta})}} \mid \bar{Y}, \boldsymbol{\beta} \xrightarrow{\mathcal{L}} N(0,1), \tag{12}$$

where

$$\hat{V}_e(\boldsymbol{\beta}) = \frac{1}{N^2} \sum_{i \in A} \sum_{j \in A} \frac{\Delta_{ij}}{\pi_{ij}} \frac{1}{\pi_i} \frac{1}{\pi_j} (y_i - \boldsymbol{x}_i^t \boldsymbol{\beta})(y_j - \boldsymbol{x}_j^t \boldsymbol{\beta}), \tag{13}$$

is a design consistent variance estimator of $\hat{\bar{Y}}_{reg}(\boldsymbol{\beta})$ for given $\boldsymbol{\beta}$. We then use $\varphi(\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}); \bar{Y}, \hat{V}_e(\boldsymbol{\beta}))$ as the density for the approximate sampling distribution of $\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta})$ in (12), where $\varphi(\cdot;\mu,\sigma^2)$ is the normal density function with mean μ and variance σ^2. Thus, the approximate posterior distribution of $\bar{Y}$ given $\boldsymbol{\beta}$ can be defined as

$$p(\bar{Y} \mid \hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}), \boldsymbol{\beta}) \propto \varphi(\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}); \bar{Y}, \hat{V}_e(\boldsymbol{\beta})) \pi(\bar{Y} \mid \boldsymbol{\beta}), \tag{14}$$

where $\pi(\bar{Y})$ is a conditional prior distribution of $\bar{Y}$ given $\boldsymbol{\beta}$. Without extra assumptions, we can use a flat prior distribution for $\pi(\bar{Y} \mid \boldsymbol{\beta})$. See Remark 1 below.

Therefore, combining (11) and (14), the approximate posterior distribution of $\bar{Y}$ can be obtained as

$$p(\bar{Y} \mid \hat{\bar{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}) = \frac{\int \phi(\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}); \bar{Y}, \hat{V}_e(\boldsymbol{\beta})) \phi_p(\hat{\boldsymbol{\beta}}; \boldsymbol{\beta}, \hat{\boldsymbol{V}}_\beta) \pi(\boldsymbol{\beta}) \pi(\bar{Y} \mid \boldsymbol{\beta}) \mathrm{d}\boldsymbol{\beta}}{\iint \phi(\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}); \bar{Y}, \hat{V}_e(\boldsymbol{\beta})) \phi_p(\hat{\boldsymbol{\beta}}; \boldsymbol{\beta}, \hat{\boldsymbol{V}}_\beta) \pi(\boldsymbol{\beta}) \pi(\bar{Y} \mid \boldsymbol{\beta}) \mathrm{d}\boldsymbol{\beta} \mathrm{d}\bar{Y}}. \tag{15}$$

Generating posterior samples from (15) can be easily carried out via the following two steps:

1. Generate posterior sample $\boldsymbol{\beta}^*$ of $\boldsymbol{\beta}$ from (11).

2. Generate posterior sample of $\bar{Y}$ from the conditional posterior (14) given θ^*.

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

Remark 1. *If an intercept term is included in* $\mathbf{x}_i$*, that is,* $\mathbf{a}^t\mathbf{x}_i = 1, \forall i \in \{1,\cdots,N\}$*, for some* $\mathbf{a}$*, then we have* $\bar{Y} = \bar{\mathbf{X}}^t\theta$ *and the parameter* $\bar{Y}$ *is completely determined from* θ*. In this case, the posterior distribution in (15) reduces to*

$$p(\bar{Y}|\hat{\bar{Y}}_{\text{reg}}(\hat{\theta}),\hat{\theta}) = \int p(\theta \mid \hat{\theta})\pi(\bar{Y} \mid \theta)\mathrm{d}\theta,$$

where $p(\theta \mid \hat{\theta})$ *is defined in (11) and* $\pi(\bar{Y} \mid \theta)$ *is a degenerating distribution at* $\bar{Y} = \bar{\mathbf{X}}^t\theta$*.*

The following theorem presents an asymptotic property of the proposed approximate Bayesian method.

Theorem 1. *Under the regularity conditions described in the Appendix, conditional on the full sample data,*

$$\sup_{\bar{Y}\in\Theta_Y} \left| p(\bar{Y}|\hat{\bar{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}),\hat{\boldsymbol{\beta}}) - \phi(\bar{Y};\hat{\bar{Y}}_{\text{reg}},\hat{V}_e) \right| \to 0, \tag{16}$$

as $n \to \infty$ *in probability, where* Θ_Y *is the feasible set for* $\bar{Y}$ *and* $p(\bar{Y}|\hat{\bar{Y}}_{\text{reg}}(\hat{\theta}),\hat{\theta})$ *is given in (15).*

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 proof is given in the Appendix. According to Theorem 1, the credible interval for $\bar{Y}$ constructed from the approximated posterior distribution (15) is asymptotically equivalent to the frequentist confidence interval based on the asymptotic normality of the common

survey regression estimator. Therefore, the frequentist survey regression estimator can be formally interpreted by the Bayesian inference. The consistency of the Bayesian point estimator (e.g. posterior mean) follows directly from (16) since $V_e(\hat{\theta}) \to 0$ in probability as $n \to \infty$.

4 Approximate Bayesian method with shrinkage priors

We now consider the case when there are many auxiliary variables in applying regression estimation. When p is large, it is important to select suitable auxiliary variables that are associated with the response variable to present irrelevant covariates from rendering the resulting estimator inefficient. To this end, we assume that the regression model in (1) contains an intercept term. That is,

$$\mathrm{E}(y_i \mid x_i) = \beta_0 + x_i^t \theta_1, \tag{17}$$

where β_0 is an intercept term.

To deal with the problem in the Bayesian way, we may define the approximate posterior distribution of $\bar{Y}$ given both β_0 and θ_1 as similar to (15). That is, we use the asymptotic distribution of the estimators $\hat{\beta}_0$ and $\hat{\theta}_1$ of β_0 and θ_1, respectively, and assign a shrinkage prior for θ_1 and flat prior for β_0. Let $\pi_\lambda(\theta_1)$ be the shrinkage prior for θ_1 with a structural parameter λ which might be multivariate.

Among the several choices of shrinkage priors, we specifically consider two priors for θ_1: Laplace (Park and Casella, 2008) and horseshoe (Carvalho et al., 2009, 2010). The Laplace prior is given by $\pi_\lambda(\theta_1) \propto \exp(-\lambda \sum_{k=1}^{p} |\beta_k|)$, 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 with the form:

$$\pi_\lambda(\boldsymbol{\beta}_1) = \prod_{k=1}^{p} \int_0^\infty \phi(\beta_k; 0, \lambda^2 u_k^2) \frac{2}{\pi(1+u_k^2)} \mathrm{d}u_k \tag{18}$$

where $\varphi(\cdot;a,b)$ 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 $\hat{\boldsymbol{\beta}}_1$ than the Laplace prior, thus allowing strong signals to remain large (Carvalho et al., 2009).

Let $\hat{V}_\beta$ be the asymptotic variance-covariance matrix of $(\hat{\beta}_0, \hat{\boldsymbol{\beta}}_1^t)$. Then, under the flat prior for β_0, the approximate posterior distribution of $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^t)$ can be defined as

$$p_\lambda(\beta_0, \boldsymbol{\beta}_1 | \hat{\beta}_0, \hat{\boldsymbol{\beta}}_1) = \frac{\phi((\hat{\beta}_0, \hat{\boldsymbol{\beta}}_1^t); (\beta_0, \boldsymbol{\beta}_1), \hat{\boldsymbol{V}}_\beta) \pi_\lambda(\boldsymbol{\beta}_1)}{\iint \phi((\hat{\beta}_0, \hat{\boldsymbol{\beta}}_1^t); (\beta_0, \boldsymbol{\beta}_1), \hat{\boldsymbol{V}}_\beta) \pi_\lambda(\boldsymbol{\beta}_1) \mathrm{d}\beta_0 \mathrm{d}\boldsymbol{\beta}_1}. \tag{19}$$

The marginal posterior of $\hat{\boldsymbol{\beta}}_1$ is given by

$$p_\lambda(\boldsymbol{\beta}_1 | \hat{\boldsymbol{\beta}}_1) = \frac{\phi(\hat{\boldsymbol{\beta}}_1; \boldsymbol{\beta}_1, \hat{V}_{\beta 11}) \pi_\lambda(\boldsymbol{\beta}_1)}{\int \phi(\hat{\boldsymbol{\beta}}_1; \boldsymbol{\beta}_1, \hat{V}_{\beta 11}) \pi_\lambda(\boldsymbol{\beta}_1) \mathrm{d}\boldsymbol{\beta}_1}, \tag{20}$$

where $\hat{V}_{\beta 11}$ is the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\beta}}_1$, which is a submatrix of $\hat{V}_\beta$. Under both shrinkage priors, we can derive efficient algorithms for doing posterior computations of $\hat{\boldsymbol{\beta}}_1$ as well as $\bar{Y}$. The details are provided in the Appendix. On the other hand, the conditional posterior of β_0 given $\hat{\boldsymbol{\beta}}_1$ is the normal distribution with mean $\hat{\beta}_0 + \hat{V}_{\beta 01} \hat{V}_{\beta 11}^{-1} (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)$ and variance $\hat{V}_{\beta 00} - \hat{V}_{\beta 01} \hat{V}_{\beta 11}^{-1} \hat{V}_{\beta 10}$, where

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

Thus, we can generate posterior samples of β_0 and $\hat{\boldsymbol{\beta}}_1$ from (19) via Markov Chain Monte Carlo in which we iteratively sample from the marginal posterior distribution of $\hat{\boldsymbol{\beta}}_1$ and conditional posterior distribution of β_0 given $\hat{\boldsymbol{\beta}}_1$. Once $\hat{\boldsymbol{\beta}}$ are

sampled from (19), we can use (14) to obtain the posterior distribution of $\bar{Y}$ for a given $\boldsymbol{\beta}$. Therefore, the approximate posterior distribution of $\bar{Y}$ can be obtained as

$$p_\lambda(\bar{Y}|\hat{\bar{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}) = \frac{\int \phi(\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}); \bar{Y}, \hat{V}_e(\boldsymbol{\beta}))\phi((\hat{\beta}_0, \hat{\boldsymbol{\beta}}_1^t); (\beta_0, \boldsymbol{\beta}_1), \hat{\boldsymbol{V}}_\beta)\pi_\lambda(\boldsymbol{\beta}_1)\pi(\bar{Y} \mid \boldsymbol{\beta})\mathrm{d}\boldsymbol{\beta}_1}{\iint \phi(\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}); \bar{Y}, \hat{V}_e(\boldsymbol{\beta}))\phi((\hat{\beta}_0, \hat{\boldsymbol{\beta}}_1^t); (\beta_0, \boldsymbol{\beta}_1), \hat{\boldsymbol{V}}_\beta)\pi_\lambda(\boldsymbol{\beta}_1)\pi(\bar{Y} \mid \boldsymbol{\beta})\mathrm{d}\boldsymbol{\beta}_1\mathrm{d}\bar{Y}}, \quad (21)$$

where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\boldsymbol{\beta}}_1^t)^t$. Generating posterior samples from (21) can be easily carried out via the following two steps:

1. For a given λ, generate posterior sample $\boldsymbol{\beta}^*$ of $\boldsymbol{\beta}$ from (19).

2. Generate posterior sample of $\bar{Y}$ from the conditional posterior (14) given $\boldsymbol{\beta}^*$.

Remark 2. *Let* $\hat{\beta}_0^{(R)}$ *and* $\hat{\boldsymbol{\beta}}_1^{(R)}$ *be the estimator of* β_0 *and* $\boldsymbol{\beta}_1$ *defined as*

$$(\hat{\beta}_0^{(R)}, \hat{\boldsymbol{\beta}}_1^{(R)}) = \text{argmin}^{\beta_0,\beta_1} \left\{ \sum_{i\in A} \frac{1}{\pi_i}(y_i - \beta_0 - \boldsymbol{x}_i^t\boldsymbol{\beta}_1)^2 + \mathrm{P}_\lambda(\boldsymbol{\beta}_1) \right\}, \quad (22)$$

where $\mathrm{P}(\boldsymbol{\beta}_1) = -2\log\pi_\lambda(\boldsymbol{\beta}_1)$ *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* $\mathrm{P}(\boldsymbol{\beta}_1) = 2\lambda\sum_{k=1}^p |\beta_k|$*, 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* $-\mathrm{P}_{i\in A}\pi_i^{-1}(y_i - \beta_0 - x^t{}_i\boldsymbol{\beta}_1)^2$ *is close to the approximated likelihood* $\phi_p((\hat{\beta}_0, \hat{\boldsymbol{\beta}}_1^t); (\beta_0, \boldsymbol{\beta}_1^t), \hat{\boldsymbol{V}}_\beta)$ *used in the approximated Bayesian method when n is large, the mode of the approximated posterior of* $(\beta_0, \boldsymbol{\beta}_1^t)$ *would be close to the frequentist estimator (22) as well.*

Remark 3. *By the frequent 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 parameter 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 as the Laplace prior and a half-Cauchy prior for λ as the horseshoe prior (18). They both lead to familiar forms of full conditional posterior distributions of λ or λ^2. The details are given in the Appendix.

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 Appendix, conditional on the full sample data,*

$$\sup_{\bar{Y}\in\Theta_Y} \left| p_\lambda(\bar{Y}|\hat{\bar{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}), \hat{\beta}_0, \hat{\boldsymbol{\beta}}_1) - \phi(\bar{Y}; \hat{\bar{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}^{(R)}), \hat{V}_e(\hat{\beta}_0^{(R)}, \hat{\boldsymbol{\beta}}_1^{(R)})) \right| \to 0, \tag{23}$$

as $n \to \infty$ in probability, where Θ_Y is the feasible set for $\bar{Y}$ and $p_\lambda(\bar{Y}|\hat{\bar{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}), \hat{\beta}_0, \hat{\boldsymbol{\beta}}_1)$ is given in (21).

Theorem 2 ensures that the proposed approximate Bayesian method is asymptotically equivalent to the frequentist version in which $\boldsymbol{\beta}_1$ 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 non-linear 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 $E(y_i \mid x_i) = m(x_i;\beta) = m_i$ and $Var(y_i \mid x_i) = \sigma^2 a(m_i)$ for some known functions $m(\cdot;\cdot)$ and $a(\cdot)$. The modelassisted regression estimator for $\bar{Y}$ with β known is then

$$\hat{\bar{Y}}_{\mathrm{reg,m}}(\boldsymbol{\beta}) = \frac{1}{N}\left\{\sum_{i=1}^{N} m(\boldsymbol{x}_i;\boldsymbol{\beta}) + \sum_{i\in A}\frac{1}{\pi_i}\Big(y_i - m(\boldsymbol{x}_i;\boldsymbol{\beta})\Big)\right\},$$

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 $\bar{Y}$ given β. That is, similarly to (14), we can obtain

$$p(\bar{Y}\mid \hat{\bar{Y}}_{\mathrm{reg,m}}(\beta),\beta) \propto \varphi(\hat{\bar{Y}}_{\mathrm{reg,m}}(\beta);\bar{Y}, \hat{V}_{e,m}(\beta))\pi(\bar{Y}\mid\beta). \qquad (24)$$

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

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

where $\mathbf{h}(x_i;\beta) = (\partial m_i/\partial\beta)/a(m_i)$. For example, for binary y_i, we may use a logistic model with $m(x_i;\beta) = \exp(x_i^t\beta)/\{1+\exp(x_i^t\beta)\}$ and $Var(y_i) = m_i(1-m_i)$, which leads to $h(x_i;\beta) = x_i$.

Under some regularity conditions, we can establish the asymptotic normality of $\hat{\beta}$. That is,

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

where

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

with $\hat{e}_i = y_i - m(x_i;\hat{\theta})$, $\hat{\mathbf{h}}_i = \mathbf{h}(x_i;\hat{\theta})$, and $\dot{m}(x;\theta) = \partial m(x;\theta)/\partial\theta$. Note that $\dot{m}(x;\theta) = m_i(1 - m_i)x_i$ under a logistic model.

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

$$p(\theta \mid \hat{\theta}) \propto \varphi(\hat{\theta} \mid \theta, \hat{V}_\beta)\pi(\theta). \tag{25}$$

We can use a shrinkage prior $\pi(\theta)$ for θ in (25) if necessary. Once θ^* is generated from (25), the posterior values of $\bar{Y}$ are generated from (24) for a given θ^*.

This formula enables us to define the approximate posterior distribution of θ of the form (11), 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 Appendix 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.

6.1 Simulation study: linear regression

In the first simulation, we generate $x_i = (x_{i1},...,x_{ip^*})^t$, $i = 1,...,N$, from a multivariate normal distribution with mean vector $(1,...,1)^t$ 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:

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

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 consider two scenarios for the sampling probability: (A) $\pi_i = 0.04$ and (B) $\text{logit}(1 - \pi_i) = 3.1 + 0.1 y_i$. The sampled units are selected via Poisson sampling, which leads to an average sample size of around 400 in both scenarios.

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 (18) for β_1, which are denoted by AB, ABL and ABH, respectively. For all the Bayesian methods, we use $\pi(\bar{Y}) \propto 1$. 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) and ridge estimation of β_1 (GREG-R; Rao and Singh, 1997). We also apply the Horwitz-Thompson (HT) estimator 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 Ga($\lambda_*^2, 1)$ for λ^2 in ABL, where λ_* is the selected value for λ in GREG-L. In ABH, we assign a prior for the tuning parameter and generate posterior samples. Based on 1,000 Monte Carlo samples, we calculate the mean squared errors (MSE), the coverage probabilities (CP) and the average length of the 95% confidence (credible) intervals, which are reported in Table 2.

Based on the results, MSEs of AB and GREG are almost the same in all cases since AB is an approximate Bayesian version of GREG. Since AB can take account of the variability in estimating β, the coverage probabilities of AB are closer to the nominal level (95%) than GREG, which is an important advantage of the proposed method. The GREG shows shorter confidence intervals with large values of p, as the variance estimator is negatively biased, and the coverage rate is lower than the nominal levels. As p gets larger, direct use of the auxiliary information makes the point estimates more inefficient as shown in Table 2, and the methods with shrinkage estimation of β such as ABH, ABL and GREG-L provide better point estimates than AB and GREG, in terms of MSEs. We note that GREG-R does not obtain much gain compared with other shrinkage methods. Comparing ABH, ABL and GREG-L, GREG-L tends to produce short confidence intervals whose coverage probabilities are smaller than the nominal level when p is large, but the proposed ABH and ABL methods produce wider credible intervals than GREG and have coverage probabilities closer to the nominal level.

6.2 Simulation study: logistic regression

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 variable Y_i are generated from the following logistic regression model:

$$Y_i \sim \text{Ber}(\delta_i), \quad \log\left(\frac{\delta_i}{1-\delta_i}\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p^*} x_{ip^*}, \qquad i = 1, \ldots, N,$$

where $\beta_0 = -1$ and the other settings are the same as the linear regression case. We again apply the same six methods 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 MSE, CP and AL based on 1,000 Monte Carlo samples are reported in Table 3, which shows again the superiority of the proposed Bayesian approach to the frequentist approach in terms of uncertainty quantification.

7 Example

We applied the proposed methods to the synthetic income data available from the sae package (Molina and Marhuenda, 2015) in R language. In the dataset, the equivalized annual net income is observed for a certain number of individuals in each province of Spain. As auxiliary variables, we used 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). Moreover, we considered 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. Here 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. In order to perform joint estimation and inference in the three provinces, we employed the following working model:

$$y_i = \sum_{h=1}^{3} x_{0i}^{(h)} \beta_0^{(h)} + \boldsymbol{x}_i^t \boldsymbol{\beta}_1 + e_i \quad (26)$$

where $x_{0i}^{(h)} = 1$ if i belong to province h, where h = 1 for Palencia, h = 2 for Segovia, and h = 3 for Soria, and 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 (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

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

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 logtransformed 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. 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 (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 design-consistent 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 to 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.

8 Concluding Remarks

We have proposed an approximate Bayesian method for survey regression estimation using a parametric regression model as the working model. The proposed Bayesian method captures the uncertainty in estimating regression parameters even when the number of the auxiliary variables is large. A main advantage of the proposed method is that it uses a 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 Bayesian inference. Although we only consider two popular prior distributions here, Laplace prior and the horseshoe prior, other priors, such as the

spike-and-slab prior (Ishwaran and Rao, 2005), can be considered. Further investigation regarding the choice of the shrinkage prior distributions will be an important research topic in the future.

Although our working model is parametric, the proposed approximate 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. Such extensions are beyond the scope of this paper and will be topics for future research.

Acknowledgement

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).

Appendix

A. Proof of (5)

From (4), we have

$$E(R_n) = -E\left\{(\hat{\bar{\boldsymbol{X}}}_{\mathrm{HT}} - \bar{\boldsymbol{X}}_N)^t(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*)\right\} = -\mathrm{tr}\left\{ \mathrm{Cov}\left(\hat{\bar{\boldsymbol{X}}}_{\mathrm{HT}}, \hat{\boldsymbol{\beta}}\right)\right\}$$
$$= -\sum_{j=1}^{p} V\left(\hat{\bar{x}}_{\mathrm{HT,j}}, \hat{\beta}_j\right) = O(p/n).$$

Also, we can show that $V(R_n) = O(p/n^2)$. Therefore, using Chebychev inequality, we have $R_n = O_p(p/n)$ and result (5) follows.

B. Posterior computation

We provide the algorithm for generating the approximate posterior distribution of β_1 given in (20) with two shrinkage priors, Laplace and horseshoe (18) priors. Using the mixture representation of both priors, we get the following Gibbs sampling algorithm.

Laplace prior

We consider the mixture representation of Laplace distribution: $\beta_k|\tau_k \sim N(0,\tau_k^2)$ and $\tau_k^2 \sim \text{Exp}(\lambda^2/2)$, independently, for $k = 1,...,p$. For λ^2, we consider the conjugate prior Ga(a,b), where Ga(a,b) is a gamma distribution with shape parameter a and rate parameter b. The full conditional distribution of β is multivariate normal with mean $A^{-1}\hat{\boldsymbol{V}}_\beta^{-1}\hat{\boldsymbol{\beta}}$ and variance-covariance matrix A^{-1} where $A = \hat{\boldsymbol{V}}_\beta^{-1} + \mathbf{D}^{-1}$ with $\mathbf{D}$ = diag($\tau_1^2,\ldots,\tau_p^2$). The full conditional distribution of λ^2

is Ga($a+p$,$b+\sum_{k=1}^{p}\tau_k^2/2$), and $\tau_1^2,\ldots,\tau_p^2$ are conditionally independent, with $1/\tau_j^2$

τ_j^2 in the parametrization conditionally inverse-Gaussian with parameters $\mu = \lambda/\beta_j$ of the inverse-Gaussian density given by

$$f(x) = \sqrt{\frac{\lambda}{2\pi}}x^{-3/2}\exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \quad x > 0$$

.

Horseshoe prior

From (18), the prior for β_1 can be expressed as a hierarchy: $\beta_k|u_k \sim N(0,\lambda^2 u^2_k)$ and $u_k \sim \text{HC}(0,1)$ independently for $k = 1,...,p$, where HC(0,1) is the standard half-Cauchy distribution. Using the hierarchical expression of the half-Cauchy dis-

tribution, we obtain the following Gibbs sampling steps. Let $A = \hat{\boldsymbol{V}}_\beta^{-1} + \mathbf{B}^{-1}$, where $\mathbf{B} = \lambda^2\text{diag}(u_1^2,\ldots,u_p^2)$. The full conditional distribution of β is multivariate normal with mean $A^{-1}\hat{\boldsymbol{V}}_\beta^{-1}\hat{\boldsymbol{\beta}}$ and variance-covariance matrix A^{-1}. The full conditional distribution of u^2_k and λ^2 are, respectively, give by

$$\text{IG}\left(1, \frac{1}{\xi_k} + \frac{\beta_k^2}{2\lambda^2}\right) \quad \text{and} \quad \text{IG}\left(\frac{p+1}{2}, \frac{1}{\gamma} + \frac{1}{2}\sum_{k=1}^{p}\frac{\beta_k^2}{u_k^2}\right),$$

where IG(a,b) denotes an inverse-Gamma distribution with shape parameter a and rate parameter b. Here ξ_k and γ are additional latent variables, and their full conditional distributions are given by IG(1,1 + $1/\delta_k^2$) and IG(1,1 + $1/\lambda^2$), respectively.

C. A sketched proof of Theorem 1

To discuss the asymptotic properties of the approximate Bayesian method, we first assume a sequence of finite populations and samples with finite fourth moments as in Isaki and Fuller (1982). The finite population is a random sample from an unknown superpopulation model. Let $\bar{Y}_*$ and θ_* be the true values of $\bar{Y}$ and θ.

Let $B_n = (\bar{Y}_* - r_n, \bar{Y}_* + r_n)$ and C_n be a ball with centre θ_* and radius $r_n \sim n^{\tau-1/2}$ for $0 < \tau < 1/2$. We make the following regularity assumptions

(C1) Assume that the sufficient conditions for the asymptotic normality of $\hat{\bar{Y}}_{\text{reg}}$ for $\bar{Y} \in B_n$ hold for the sequence of finite populations and samples.

(C2) Assume that the prior distribution $\pi(\bar{Y})$ is positive and satisfies a Lipschitz condition over its support Θ_Y; that is, there exists $C_1 < \infty$ such that $|\pi(\theta_1) - \pi(\theta_2)| \le C_1|\theta_1 - \theta_2|$ for $\theta_1, \theta_2 \in \Theta_Y$.

(C3) Assume that $\hat{V}_\beta = V_\beta\{1+o_P(1)\}$ and $(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^t\hat{\boldsymbol{V}}_\beta^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^t\boldsymbol{V}_\beta^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\{1 + o_P(1)\}$ for any $\theta \in C_n$ and $n \to \infty$.

(C4) Assume that $\pi(\theta)$ is positive and finite over its support Θ_β.

Sufficient conditions for (C1) are discussed within various asymptotic structures (e.g. Binder, 1983; Pfeffermann and Sverchkov, 2009). Conditions (C2) and (C4) are satisfied for common priors such as (multivariate) normal distribution . Condition (C3) essentially requires that the design variance estimators be consistent and meet a certain continuity condition.

Proof. Let $g(\bar{Y}, \theta) = \varphi(\hat{\bar{Y}}_{\mathrm{reg}}(\theta); \bar{Y}, \hat{V}_e(\theta))\varphi_p(\theta; \hat{\theta}, \hat{V}_\beta)\pi(\theta)$. Then, the approximated posterior distribution is given by

$$p(\bar{Y}|\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}) = \frac{\int g(\bar{Y}, \boldsymbol{\beta})\mathrm{d}\boldsymbol{\beta}}{\int\int g(\bar{Y}, \boldsymbol{\beta})\mathrm{d}\boldsymbol{\beta}\mathrm{d}\bar{Y}},$$

Note that

$$\int g(\bar{Y}, \boldsymbol{\beta})\mathrm{d}\boldsymbol{\beta} = \int_{\beta\in C_n} g(\bar{Y}, \boldsymbol{\beta})\mathrm{d}\boldsymbol{\beta} + \int_{\beta\in\mathbb{R}^p\setminus C_n} g(\bar{Y}, \boldsymbol{\beta})\mathrm{d}\boldsymbol{\beta} \quad (27)$$

By the same argument in the proof of Theorem 1 in Wang et al. (2018), we have

$$\operatorname*{plim}_{n\to\infty} \int_{\beta\in C_n} \phi_p(\hat{\boldsymbol{\beta}}; \boldsymbol{\beta}, \hat{\boldsymbol{V}}_\beta)\mathrm{d}\boldsymbol{\beta} = 1,$$

so the second term in (27) is $o_P(1)$. On the other hand, under condition (C3), $\varphi_p(\theta; \hat{\theta}, \hat{V}_\beta) = \varphi_p(\theta; \hat{\theta}, V_\beta)\{1+o_P(1)\}$ as $n \to \infty$, for any $\theta \in C_n$, thereby under condition (C4),

$$\int_{\beta\in C_n} g(\bar{Y}, \theta)\mathrm{d}\theta = \int_{\beta\in C_n} \varphi(\hat{\bar{Y}}_{\mathrm{reg}}(\theta); \bar{Y}, \hat{V}_e(\theta))\varphi_p(\theta; \hat{\theta}, V_\beta)\pi(\theta)\mathrm{d}\theta$$

$$= \varphi(\hat{\bar{Y}}_{\mathrm{reg}}(\theta_*); \bar{Y}, \hat{V}_e(\theta_*))\pi(\theta_*)\{1 + o_P(1)\}$$

as $n \to \infty$ since $V \to 0$ and $\hat{\theta} \to \theta_*$ as $n \to \infty$. Hence, we have

$$
\begin{aligned}
p(\bar{Y}|\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}}),\hat{\boldsymbol{\beta}}) &= \frac{\pi(\boldsymbol{\beta}_*)\phi(\hat{\bar{Y}}_{\mathrm{reg}}(\boldsymbol{\beta}_*);\bar{Y},\hat{V}_e(\boldsymbol{\beta}_*))\pi(\bar{Y})\{1+o_P(1)\}}{\pi(\boldsymbol{\beta}_*)\int\phi(\hat{\bar{Y}}_{\mathrm{reg}}(\boldsymbol{\beta}_*);\bar{Y},\hat{V}_e(\boldsymbol{\beta}_*))\pi(\bar{Y})d\bar{Y}\{1+o_P(1)\}} \\
&= \frac{\pi(\bar{Y})}{\pi(\bar{Y}_*)}\phi(\hat{\bar{Y}}_{\mathrm{reg}}(\boldsymbol{\beta}_*);\bar{Y},\hat{V}_e(\boldsymbol{\beta}_*))\{1+o_P(1)\} \\
&= \phi(\hat{\bar{Y}}_{\mathrm{reg}}(\boldsymbol{\beta}_*);\bar{Y},\hat{V}_e(\boldsymbol{\beta}_*))\{1+o_P(1)\} \qquad (28)
\end{aligned}
$$

$$
= \varphi(\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}});\bar{Y},\hat{V}_e(\hat{\boldsymbol{\beta}}))\{1+o_P(1)\}, \qquad (29)
$$

for any $\bar{Y} \in B_n$ as $n \to \infty$, where (28) follows from (C2), and (29) follows from the properties $\hat{V}_e(\hat{\boldsymbol{\beta}}) = \hat{V}_e(\boldsymbol{\beta}_*)\{1+o_P(1)\}$ and $\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}}) = \hat{\bar{Y}}_{\mathrm{reg}}(\boldsymbol{\beta}_*)\{1+o_P(1)\}$ under (C1). Let $R_n = \{\bar{Y} \in \Theta_Y : \hat{V}_e(\hat{\boldsymbol{\beta}})^{-1}(\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}}) - \bar{Y})^2 \le \chi^2_{1,\beta_0}\}$, where χ^2_{1,β_0} is the upper β_0-quantile of the chi-squared distribution with 1 degree of freedom. Then, $\mathrm{plim}_{n\to\infty} P(R_n) = \beta_0$. Since $\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}}) - \bar{Y}_* = O_p(n^{-1/2})$ and $r_n = n^{\tau-1/2}$, which is slower than $n^{-1/2}$, it holds that $\lim_{n\to\infty} P(R_n \subset B_n) = 1$. Then,

$$
\lim_{n\to\infty} P\left(\int_{B_n}\phi(\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}});\bar{Y},\hat{V}_e(\hat{\boldsymbol{\beta}}))\mathrm{d}\bar{Y} \ge \int_{R_n}\phi(\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}});\bar{Y},\hat{V}_e(\hat{\boldsymbol{\beta}}))\mathrm{d}\bar{Y}\right) = 1,
$$

which means that

$$
\lim_{n\to\infty} P\left(\int_{B_n}\phi(\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}});\bar{Y},\hat{V}_e(\hat{\boldsymbol{\beta}}))\mathrm{d}\bar{Y} \ge \beta_0\right) = 1
$$

for any $\beta_0 \in (0,1)$, implying

$$
\operatorname*{plim}_{n\to\infty}\int_{B_n}\phi(\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}});\bar{Y},\hat{V}_e(\hat{\boldsymbol{\beta}}))\mathrm{d}\bar{Y} = 1. \qquad (30)
$$

Then,

$$
\begin{aligned}
&\sup_{\bar{Y}\in\Theta_Y}\left|p(\bar{Y}|\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}}),\hat{\boldsymbol{\beta}}) - \phi(\bar{Y};\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}}),\hat{V}_e(\hat{\boldsymbol{\beta}}))\right| \\
\le &\sup_{\bar{Y}\in B_n}\left|p(\bar{Y}|\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}}),\hat{\boldsymbol{\beta}}) - \phi(\bar{Y};\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}}),\hat{V}_e(\hat{\boldsymbol{\beta}}))\right| \\
&+ \sup_{\bar{Y}\in\Theta_Y\setminus B_n}\left|p(\bar{Y}|\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}}),\hat{\boldsymbol{\beta}}) - \phi(\bar{Y};\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\boldsymbol{\beta}}),\hat{V}_e(\hat{\boldsymbol{\beta}}))\right|,
\end{aligned}
$$

which are both $o_P(1)$ from (29) and (30). This completes the proof. □

D. A sketched proof of Theorem 2

The condition (C4) given in the proof of Theorem 1 may not be satisfied for shrinkage priors. For example, the horseshoe prior (18) diverge at the origin $\beta_k = 0$. In what follows, let $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^t)$ and define $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^{(R)}$ in the same way. We use the
following alternative condition for the shrinkage prior $\pi_\lambda(\boldsymbol{\theta})$:

(C5) The regularized estimator $\hat{\boldsymbol{\theta}}_R$ under penalty $-\log\pi_\lambda(\boldsymbol{\theta}_1)$ is asymptotically normal, that is, $\sqrt{n}(\hat{\boldsymbol{\theta}}^{(R)} - \boldsymbol{\theta}_*) \to N(0, \mathbf{C})$, where $\mathbf{C}$ is a positive definite matrix and λ is appropriately chosen.

Under the Laplace prior, $\hat{\boldsymbol{\theta}}^{(R)}$ is equivalent to the Lasso estimator, and the above property holds if $\lambda = o(\sqrt{n})$ (Knight and Fu, 2000; McConville et al., 2017). For general prior $\pi_\lambda(\boldsymbol{\theta}_1)$, this condition holds if the assumption regarding the penalty term $P_\lambda(\boldsymbol{\theta}_1)$ given in Fan and Li (2001) is satisfied.

Proof. It is noted that

$$\begin{aligned}
&\phi_p((\hat{\beta}_0, \hat{\boldsymbol{\beta}}^t); (\beta_0, \boldsymbol{\beta}^t), \hat{\boldsymbol{V}}_\beta)\pi_\lambda(\boldsymbol{\beta}_1) \\
&\quad \propto \exp\left\{-\frac{1}{2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^t \hat{\boldsymbol{V}}_\beta^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \log\pi_\lambda(\boldsymbol{\beta}_1)\right\} \\
&\quad = \exp\left\{-\frac{1}{2}\sum_{i\in A}\frac{1}{\pi_i}(y_i - \beta_0 - \boldsymbol{x}_i^t\boldsymbol{\beta}_1)^2 + \log\pi_\lambda(\boldsymbol{\beta}_1)\right\}\{1 + o_P(1)\} \\
&\quad = \exp\left\{-\frac{n}{2}(\hat{\boldsymbol{\beta}}^{(R)} - \boldsymbol{\beta})^t\mathbf{C}^{-1}(\hat{\boldsymbol{\beta}}^{(R)} - \boldsymbol{\beta})\right\}\{1 + o_P(1)\}.
\end{aligned}$$

Define

$$g(Y, \bar{\boldsymbol{\theta}}) = \varphi(\bar{\hat{Y}}_{\text{reg}}(\boldsymbol{\theta}); Y, \bar{\hat{V}}_e(\boldsymbol{\theta}))\varphi(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta}, \hat{V}_\beta)\pi_\lambda(\boldsymbol{\theta}_1).$$

Then, it holds that

$$\int g(\bar{Y}, \beta)\mathrm{d}\beta = \varphi(\hat{\bar{Y}}_{\mathrm{reg}}(\beta_*); \bar{Y}, \hat{V}_e(\beta_*))\{1 + o_P(1)\} \int_{\beta \in R_n}$$

as $n \to \infty$, where R_n is a ball with center β_* and radius $O(n^{\tau - 1/2})$ for $0 < \tau < 1/2$. Hence, the statement can be proved in the same way as the proof of Theorem 1 since $\varphi(\hat{\bar{Y}}_{\mathrm{reg}}(\beta_*); \bar{Y}, \hat{V}_e(\beta_*)) = \varphi(\hat{\bar{Y}}_{\mathrm{reg}}(\hat{\beta}_{(R)}); \bar{Y}, \hat{V}_e(\hat{\beta}_{(R)}))\{1 + o_P(1)\}$. □

References

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Reviews 51*, 279–292.

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.

Deville, J. C. and C. E. Sa¨rndal (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association 87*, 376–382.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its Oracle properties. *Journal of the American Statistical Association 96*, 1348–1360.

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.

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.

Knight, K. and W. Fu (2000). Asymptotics for Lasso-type estimators. *Journal of Official Statistics 28*, 1356–1378.

Little, R. J. A. (2012). Calibrated Bayes, an alternative inferential paradigm for official statistics. *Journal of Official Statistics*.

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, 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. *Sankhy¯a, Series B 61*, 166–186.

Pfeffermann, D. and M. Sverchkov (2009). Inference under informative sampling. *Handbook of Statistics*.

Rao, J. and A. Singh (1997). A ridge-shrinkage method for range-restricted weight calibration in survey sampling. In *Proceedings of the Section on Survey Research Methods*, pp. 57–65. American Statistical Association.

Sa¨rndal, C. E. and Lundstr¨om (2005). *Estimation in surveys with nonresponse*. Chichester: John Wiley & Sons.

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.

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.

Table 2: Summary of the simulation results in scenarios (A) and (B) with linear regression. All values are multiplied by 100.

Scenario (A): $\pi_i = 0.04$

p	GREG	GREG-L	GREG-R	AB	ABL			ABH	HT
30	1.06	1.00	1.07	1.06	1.04			1.00	2.32
40	1.01	0.98	1.02	1.01	1.01	0.99	2.32	1.01	2.32
	1.15	1.00	1.14	1.15	1.05	1.03	2.32		
	93.9	94.2	94.2	95.2	94.7	95.3	95.4		
	92.8	94.5	93.1	94.6	94.5	95.3	95.4		
	92.0	94.3	92.7	94.5	94.8	95.1	95.4		
	90.9	93.8	91.7	93.8	94.4	94.8	95.4		
	37.3	37.8	37.3	39.1	39.4	38.7	60.5		
	36.8	37.7	36.9	39.4	39.6	38.7	60.5		
	36.3	37.6	36.5	39.8	39.6	38.8	60.5		
	35.8	37.6	36.1	40.1	39.5	38.8	60.5		
		1.10	1.10	1.05					

	20	1.14	1.12	1.16	1.14	1.21	1.12	3.35
MSE								
	50	1.40	1.15	1.44	1.40	1.27	1.17	3.35
	20	92.8	93.0	92.5	94.4	94.0	94.2	94.4
	30	91.6	92.9	91.2	93.0	93.9	93.7	94.4
CP	40	89.9	93.4	89.8	92.5	94.3	93.6	94.4
	50	88.1	92.8	88.3	91.9	93.4	93.9	94.4
	20	39.1	39.9	39.3	41.1	42.9	41.0	70.3
	30	38.4	39.8	38.6	41.3	43.2	41.1	70.3
AL	40	37.7	39.8	38.1	41.5	43.2	41.1	70.3
	50	37.0	39.7	37.5	41.7	43.1	41.1	70.3

Scenario (B): $\text{logit}(1 - \pi_i) = 3.1 + 0.1y_i$

p	GREG	GREG-L	GREG-R	AB	ABL			ABH	HT
30	1.21	1.12	1.24	1.21	1.24	1.13	3.35		
40	1.30	1.13	1.34	1.30	1.25	1.15	3.35		

Table 3: Summary of the simulation results in scenarios (A) and (B) with logistic regression. MSE values are multiplied by 10,000 and CP and AL values are multiplied by 100.

Scenario (A): $\pi_i = 0.04$

p	GREG	GREG-L	GREG-R	AB	ABL			ABH	HT
303.76	3.56	3.74	3.71	3.66				3.53	12.4

40	3.66	3.56	3.67	3.64	3.60	3.51	12.4	3.5512.4
	3.97	3.58	3.94	3.86	3.81	3.60	12.4	
	93.0	94.2	94.1	95.4	95.7	95.7	94.3	
	92.4	93.9	92.8	94.7	96.2	95.8	94.3	
	91.1	94.0	92.4	94.9	96.4	96.0	94.3	
	89.6	94.1	91.7	95.2	96.3	96.2	94.3	
	7.01	7.12	7.10	7.45	7.58	7.44	13.6	
	6.88	7.09	6.99	7.52	7.69	7.52	13.6	
	6.73	7.08	6.89	7.57	7.81	7.59	13.6	
	6.58	7.06	6.82	7.61	7.94	7.69	13.6	
		3.83	3.78	3.72				

Scenario (B): logit$(1 - \pi_i) = 3.1 + 0.1y_i$

p	GREG	GREG-L	GREG-R	AB	ABL	ABH	HT

	20	3.69	3.58	3.69	3.67	3.64	3.56	12.4
MSE								
	50	4.01	3.62	3.94	3.91	3.80	3.63	12.4
CP	20	92.3	93.1	92.9	94.3	94.3	94.3	95.0
	30	91.5	92.8	92.1	94.3	94.3	94.2	95.0
	40	90.5	92.6	90.9	94.4	95.7	94.7	95.0
	50	89.0	92.6	90.1	93.4	95.8	95.1	95.0
AL	20	6.88	7.00	6.98	7.31	7.44	7.30	13.8
	30	6.75	6.97	6.87	7.37	7.54	7.36	13.8
	40	6.62	6.95	6.77	7.43	7.65	7.45	13.8
	50	6.47	6.93	6.69	7.47	7.78	7.54	13.8
	30	3.78	3.60	3.76	3.75	3.71	3.58	12.4
	40	3.88	3.60	3.85	3.82	3.75	3.61	12.4

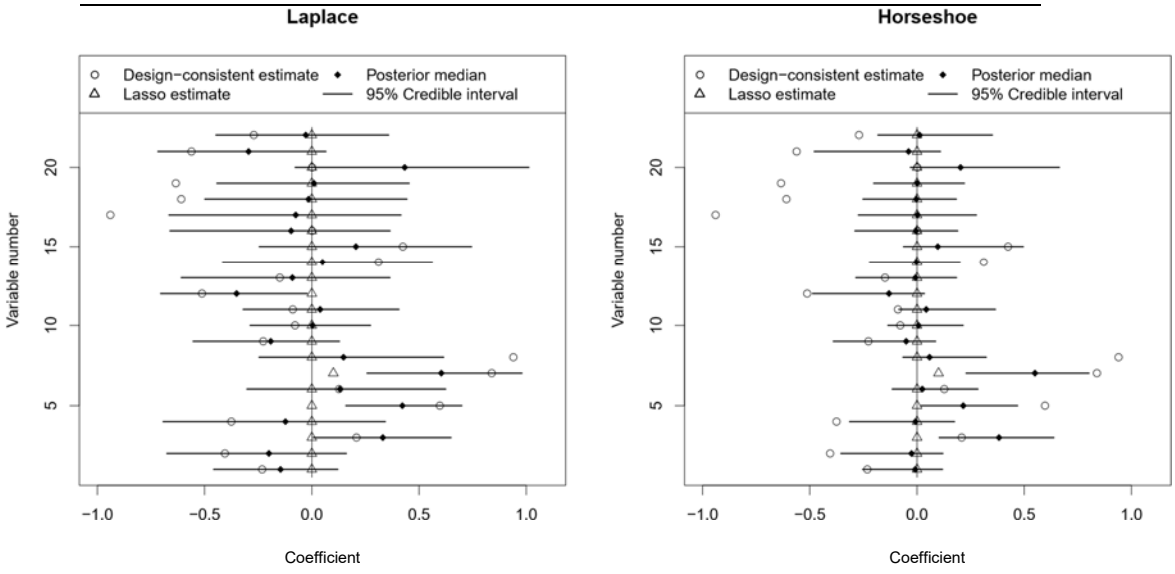

Figure 1: 95% credible intervals of regression coefficients under Laplace (left) and horseshoe (right) priors.

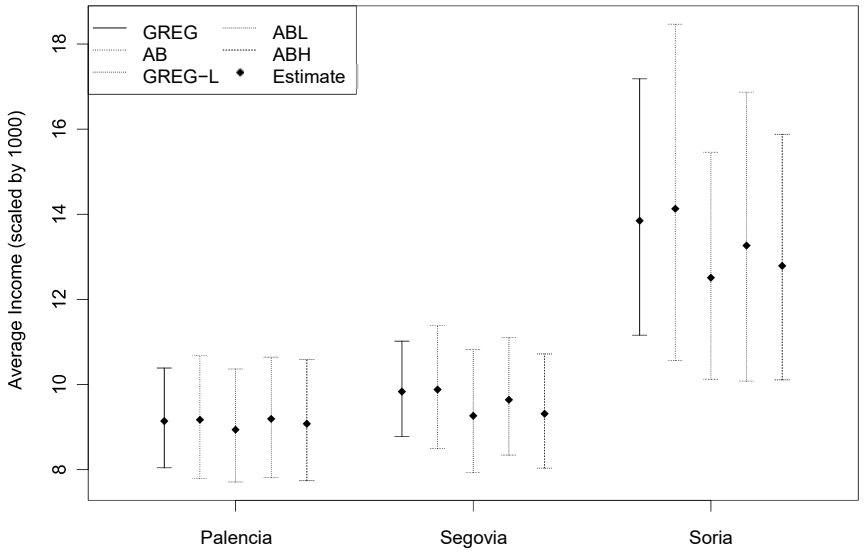

Figure 2: 95% confidence and credible intervals for average income based on five methods in three provinces in Spain.