# SUPPLEMENTARY MATERIAL FOR "AN APPROXIMATE BAYESIAN APPROACH TO MODEL-ASSISTED SURVEY ESTIMATION WITH MANY AUXILIARY VARIABLES"

Shonosuke Sugasawa[1] and Jae Kwang Kim[2]

[1] *The University of Tokyo and* [2] *Iowa State University*

This Supplementary Material contains a proof of (5), details of posterior computation, proofs of theorems and results of additional simulation suites.

## S1 Proof of (2.5)

We assume the same conditions in the proof of Theorem 1, given in Section S3. From (2.4), we have

$$\begin{aligned} E(R_n) &= -E\left\{(\hat{\hat{\boldsymbol{X}}}_{\rm HT} - \bar{\boldsymbol{X}}_N)^t(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1*})\right\} = -{\rm tr}\left\{{\rm Cov}\left(\hat{\hat{\boldsymbol{X}}}_{\rm HT}, \hat{\boldsymbol{\beta}}_1\right)\right\} \\ &= -\sum_{j=1}^{p}{\rm Cov}\left(\hat{\hat{x}}_{\rm HT,j}, \hat{\beta}_j\right) = O(p/n), \end{aligned}$$

where the expectation is taken with respect to the sampling distribution. 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 (2.5) follows.

## S2 Posterior computation

We provide the algorithm for generating the approximate posterior distribution of $\boldsymbol{\beta}_1$ given in (4.20) with two shrinkage priors, Laplace and horseshoe (4.18) 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 $\tau_k^2 \sim \mathrm{Exp}(\lambda^2/2)$, independently, for $k = 1, \ldots, p$. For $\lambda^2$, we consider the conjugate prior $\mathrm{Ga}(a, b)$, where $\mathrm{Ga}(a, b)$ is a gamma distribution with shape parameter $a$ and rate parameter $b$. The full conditional distribution of $\boldsymbol{\beta}_1$ is multivariate normal with mean $\boldsymbol{A}^{-1}\hat{\boldsymbol{V}}_{\beta 11}^{-1}\hat{\boldsymbol{\beta}}_1$ and variance-covariance matrix $\boldsymbol{A}^{-1}$ where $\boldsymbol{A} = \hat{\boldsymbol{V}}_{\beta 11}^{-1} + \mathbf{D}^{-1}$ with $\mathbf{D} = \mathrm{diag}(\tau_1^2, \ldots, \tau_p^2)$. The full conditional distribution of $\lambda^2$ is $\mathrm{Ga}(a + p, b + \sum_{k=1}^{p} \tau_k^2/2)$, and $\tau_1^2, \ldots, \tau_p^2$ are conditionally independent, with $1/\tau_j^2$ conditionally inverse-Gaussian with parameters $\mu = \sqrt{\lambda/\beta_j^2}$ in the parametrization 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

The prior for $\boldsymbol{\beta}_1$ can be expressed as a hierarchy: $\beta_k|u_k \sim N(0, \lambda^2 u_k^2)$ and $u_k \sim \mathrm{HC}(0, 1)$ independently for $k = 1, \ldots, p$, where $\mathrm{HC}(0, 1)$ is the standard half-Cauchy distribution. Using the hierarchical expression of the half-Cauchy distribution, we obtain the following Gibbs sampling steps. Let $\boldsymbol{A} = \hat{\boldsymbol{V}}_{\beta 11}^{-1} + \mathbf{B}^{-1}$, where $\mathbf{B} = \lambda^2 \mathrm{diag}(u_1^2, \ldots, u_p^2)$. The full conditional distribution of $\boldsymbol{\beta}_1$ is multivariate normal with mean $\boldsymbol{A}^{-1}\hat{\boldsymbol{V}}_{\beta 11}^{-1}\hat{\boldsymbol{\beta}}_1$ and variance-covariance matrix $\boldsymbol{A}^{-1}$. The full conditional distribution of $u_k^2$ and $\lambda^2$ are, respectively, give by

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

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

# S3 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 $\boldsymbol{\beta}_{1*}$ be the probability limit of $\hat{\bar{Y}}_{HT}$ and $\hat{\boldsymbol{\beta}}_1$, respectively. Let $B_n = (\bar{Y}_* - r_n, \bar{Y}_* + r_n)$ and $C_n$ be a ball with centre $\boldsymbol{\beta}_{1*}$ and radius $r_n \sim n^{\tau - 1/2}$ for $0 < \tau < 1/2$. Furthermore, let $\Theta_\beta$ and $\Theta_Y$ be the supports of the prior distributions of $\boldsymbol{\beta}_1$ and $\bar{Y}|\boldsymbol{\beta}_1$. We make the following regularity assumptions:

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

(C2) For fixed $\boldsymbol{\beta}_1 \in \Theta_\beta$, assume that the conditional prior distribution $\pi(\bar{Y}|\boldsymbol{\beta}_1)$ is positive and satisfies a Lipschitz condition over its support $\Theta_Y$, that is, there exists $C_1(\boldsymbol{\beta}_1) < \infty$ such that $|\pi(\theta_1|\boldsymbol{\beta}_1) - \pi(\theta_2|\boldsymbol{\beta}_1)| \le C_1(\boldsymbol{\beta}_1)|\theta_1 - \theta_2|$ for $\theta_1, \theta_2 \in \Theta_Y$ and $\boldsymbol{\beta}_1 \in \Theta_\beta$.

(C3) Assume that $\hat{\boldsymbol{V}}_{\beta 11} = \boldsymbol{V}_{\beta 11}\{1 + o_P(1)\}$ and $(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)^t \hat{\boldsymbol{V}}_{\beta 11}^{-1} (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) = (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)^t \boldsymbol{V}_{\beta 11}^{-1} (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)\{1 + o_P(1)\}$ for any $\boldsymbol{\beta}_1 \in C_n$ and $n \to \infty$.

(C4) Assume that $\pi(\boldsymbol{\beta}_1)$ is positive and finite over its support $\Theta_\beta$.

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 $g(\bar{Y},\boldsymbol{\beta}) = \phi(\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}_1);\bar{Y},\hat{V}_e(\boldsymbol{\beta}))\phi_p(\hat{\boldsymbol{\beta}}_1;\boldsymbol{\beta}_1,\hat{\boldsymbol{V}}_{\beta 11})\pi(\boldsymbol{\beta}_1)\pi(\bar{Y}|\boldsymbol{\beta}_1)$. Then, the approximated posterior distribution is given by

$$p(\bar{Y}|\hat{\bar{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1),\hat{\boldsymbol{\beta}}_1) = \frac{\int g(\bar{Y},\boldsymbol{\beta}_1)\mathrm{d}\boldsymbol{\beta}_1}{\iint g(\bar{Y},\boldsymbol{\beta}_1)\mathrm{d}\boldsymbol{\beta}_1\mathrm{d}\bar{Y}}.$$

Note that

$$\int g(\bar{Y},\boldsymbol{\beta}_1)\mathrm{d}\boldsymbol{\beta}_1 = \int_{C_n} g(\bar{Y},\boldsymbol{\beta}_1)\mathrm{d}\boldsymbol{\beta}_1 + \int_{\mathbb{R}^p\setminus C_n} g(\bar{Y},\boldsymbol{\beta}_1)\mathrm{d}\boldsymbol{\beta}_1 \tag{S3.1}$$

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

$$\operatorname*{plim}_{n\to\infty}\int_{C_n}\phi_p(\hat{\boldsymbol{\beta}}_1;\boldsymbol{\beta}_1,\hat{\boldsymbol{V}}_{\beta 11})\mathrm{d}\boldsymbol{\beta}_1 = 1,$$

so the second term in (S3.1) is $o_P(1)$. On the other hand, under condition (C3), $\phi_p(\hat{\boldsymbol{\beta}}_1;\boldsymbol{\beta}_1,\hat{\boldsymbol{V}}_{\beta 11}) = \phi_p(\hat{\boldsymbol{\beta}}_1;\boldsymbol{\beta}_1,\boldsymbol{V}_{\beta 11})\{1+o_P(1)\}$ as $n\to\infty$, for any $\boldsymbol{\beta}_1\in C_n$, thereby under condition (C4),

$$\begin{aligned}\int_{C_n} g(\bar{Y},\boldsymbol{\beta}_1)\mathrm{d}\boldsymbol{\beta}_1 &= \int_{C_n}\phi(\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}_1);\bar{Y},\hat{V}_e(\boldsymbol{\beta}_1))\phi_p(\hat{\boldsymbol{\beta}}_1;\boldsymbol{\beta}_1,\boldsymbol{V}_{\beta 11})\pi(\boldsymbol{\beta}_1)\mathrm{d}\pi(\bar{Y}|\boldsymbol{\beta}_1)\boldsymbol{\beta}_1\\ &= \phi(\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}_{1*});\bar{Y},\hat{V}_e(\boldsymbol{\beta}_{1*}))\pi(\boldsymbol{\beta}_{1*})\pi(\bar{Y}|\boldsymbol{\beta}_{1*})\{1+o_P(1)\}\end{aligned}$$

as $n \to \infty$ since $V \to 0$ and $\hat{\boldsymbol{\beta}}_1 \to \boldsymbol{\beta}_{1*}$ as $n \to \infty$. Hence, we have

$$\begin{aligned}
p(\bar{Y}|\hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1), \hat{\boldsymbol{\beta}}_1) &= \frac{\pi(\boldsymbol{\beta}_{1*})\phi(\hat{\hat{Y}}_{\text{reg}}(\boldsymbol{\beta}_{1*}); \bar{Y}, \hat{V}_e(\boldsymbol{\beta}_{1*}))\pi(\bar{Y}|\boldsymbol{\beta}_{1*})\{1 + o_P(1)\}}{\pi(\boldsymbol{\beta}_{1*}) \int \phi(\hat{\hat{Y}}_{\text{reg}}(\boldsymbol{\beta}_{1*}); \bar{Y}, \hat{V}_e(\boldsymbol{\beta}_{1*}))\pi(\bar{Y}|\boldsymbol{\beta}_{1*}) d\bar{Y}\{1 + o_P(1)\}} \\
&= \frac{\pi(\bar{Y}|\boldsymbol{\beta}_{1*})}{\pi(\bar{Y}_*|\boldsymbol{\beta}_{1*})}\phi(\hat{\hat{Y}}_{\text{reg}}(\boldsymbol{\beta}_{1*}); \bar{Y}, \hat{V}_e(\boldsymbol{\beta}_{1*}))\{1 + o_P(1)\} \\
&= \phi(\hat{\hat{Y}}_{\text{reg}}(\boldsymbol{\beta}_{1*}); \bar{Y}, \hat{V}_e(\boldsymbol{\beta}_{1*}))\{1 + o_P(1)\} \qquad \text{(S3.2)} \\
&= \phi(\hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1); \bar{Y}, \hat{V}_e(\hat{\boldsymbol{\beta}}_1))\{1 + o_P(1)\}, \qquad \text{(S3.3)}
\end{aligned}$$

for any $\bar{Y} \in B_n$ as $n \to \infty$, where (S3.2) follows from (C2), and (S3.3) follows from the properties $\hat{V}_e(\hat{\boldsymbol{\beta}}_1) = \hat{V}_e(\boldsymbol{\beta}_{1*})\{1 + o_P(1)\}$ and $\hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1) = \hat{\hat{Y}}_{\text{reg}}(\boldsymbol{\beta}_{1*})\{1 + o_P(1)\}$ under (C1). Let $R_n = \{\bar{Y} \in \Theta_Y : \hat{V}_e(\hat{\boldsymbol{\beta}}_1)^{-1}(\hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1) - \bar{Y})^2 \le \chi_1^2(q)\}$, where $\chi_k^2(q)$ is the upper $100q\%$-quantile of the chi-squared distribution with $k$ degree of freedom. Then, $\text{plim}_{n\to\infty} P(R_n) = q$. Since $\hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1) - \bar{Y}_* = O_p(n^{-1/2})$ 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,

$$\lim_{n\to\infty} P\left(\int_{B_n} \phi(\hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1); \bar{Y}, \hat{V}_e(\hat{\boldsymbol{\beta}}_1))\mathrm{d}\bar{Y} \ge \int_{R_n} \phi(\hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1); \bar{Y}, \hat{V}_e(\hat{\boldsymbol{\beta}}_1))\mathrm{d}\bar{Y}\right) = 1,$$

which means that

$$\lim_{n\to\infty} P\left(\int_{B_n} \phi(\hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1); \bar{Y}, \hat{V}_e(\hat{\boldsymbol{\beta}}_1))\mathrm{d}\bar{Y} \ge q\right) = 1$$

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

$$\underset{n\to\infty}{\text{plim}} \int_{B_n} \phi(\hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1); \bar{Y}, \hat{V}_e(\hat{\boldsymbol{\beta}}_1))\mathrm{d}\bar{Y} = 1. \qquad \text{(S3.4)}$$

Then,

$$\begin{aligned}
&\sup_{\bar{Y}\in\Theta_Y} \left|p(\bar{Y}|\hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1), \hat{\boldsymbol{\beta}}_1) - \phi(\bar{Y}; \hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1), \hat{V}_e(\hat{\boldsymbol{\beta}}_1))\right| \\
\le &\sup_{\bar{Y}\in B_n} \left|p(\bar{Y}|\hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1), \hat{\boldsymbol{\beta}}_1) - \phi(\bar{Y}; \hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1), \hat{V}_e(\hat{\boldsymbol{\beta}}_1))\right| \\
&+ \sup_{\bar{Y}\in\Theta_Y\setminus B_n} \left|p(\bar{Y}|\hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1), \hat{\boldsymbol{\beta}}_1) - \phi(\bar{Y}; \hat{\hat{Y}}_{\text{reg}}(\hat{\boldsymbol{\beta}}_1), \hat{V}_e(\hat{\boldsymbol{\beta}}_1))\right|,
\end{aligned}$$

which are both $o_P(1)$ from (S3.3) and (S3.4). This completes the proof. □

# S4 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 diverge at the origin $\beta_k = 0$. In what follows, let $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^t)$ and define $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_1^{(R)}$ in the same way. We use the following alternative condition for the shrinkage prior $\pi_\lambda(\boldsymbol{\beta}_1)$:

(C5) The regularized estimator $\hat{\boldsymbol{\beta}}_1^{(R)}$ under penalty $-\log \pi_\lambda(\boldsymbol{\beta}_1)$ is asymptotically normal, that is, $\sqrt{n}(\hat{\boldsymbol{\beta}}_1^{(R)} - \boldsymbol{\beta}_{1*}) \to N(0, \mathbf{C})$, where $\mathbf{C}$ is a positive definite matrix and $\lambda$ is appropriately chosen.

Under the Laplace prior, $\hat{\boldsymbol{\beta}}_1^{(R)}$ 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 $\pi_\lambda(\boldsymbol{\beta}_1)$, this condition holds if the assumption regarding the penalty term $P_\lambda(\boldsymbol{\beta}_1)$ given in Fan and Li (2001) is satisfied.

*Proof.* It is noted that

$$
\begin{aligned}
&\phi_p(\hat{\boldsymbol{\beta}}_1; \boldsymbol{\beta}_1, \hat{\boldsymbol{V}}_{\beta 11})\pi_\lambda(\boldsymbol{\beta}_1) \\
&\quad \propto \exp\left\{ -\frac{1}{2}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)^t \hat{\boldsymbol{V}}_{\beta 11}^{-1}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + \log \pi_\lambda(\boldsymbol{\beta}_1) \right\} \\
&\quad = \exp\left\{ -\frac{1}{2} \max_{\beta_0} \sum_{i \in A} \frac{1}{\pi_i}(y_i - \beta_0 - x_i^t \boldsymbol{\beta}_1)^2 + \log \pi_\lambda(\boldsymbol{\beta}_1) \right\} \{1 + o_P(1)\} \\
&\quad = \exp\left\{ -\frac{n}{2}(\hat{\boldsymbol{\beta}}_1^{(R)} - \boldsymbol{\beta}_1)^t \mathbf{C}^{-1}(\hat{\boldsymbol{\beta}}_1^{(R)} - \boldsymbol{\beta}_1) \right\} \{1 + o_P(1)\}.
\end{aligned}
$$

Define

$$
g(\bar{Y}, \boldsymbol{\beta}_1) = \phi(\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}_1); \bar{Y}, \hat{V}_e(\boldsymbol{\beta}_1))\phi(\hat{\boldsymbol{\beta}}_1; \boldsymbol{\beta}_1, \hat{\boldsymbol{V}}_{\beta 11})\pi_\lambda(\boldsymbol{\beta}_1)\pi(\bar{Y}|\boldsymbol{\beta}_1).
$$

Then, it holds that

$$
\int_{S_n} g(\bar{Y}, \boldsymbol{\beta}_1) \mathrm{d}\boldsymbol{\beta}_1 = \phi(\hat{\bar{Y}}_{\text{reg}}(\boldsymbol{\beta}_{1*}); \bar{Y}, \hat{V}_e(\boldsymbol{\beta}_{1*}))\pi_\lambda(\boldsymbol{\beta}_{1*})\pi(\bar{Y}|\boldsymbol{\beta}_{1*})\{1 + o_P(1)\}
$$

as $n \to \infty$, where $S_n$ is a ball with center $\boldsymbol{\beta}_{1*}$ 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 $\phi(\hat{\bar{Y}}_{\rm reg}(\boldsymbol{\beta}_{1*}); \bar{Y}, \hat{V}_e(\boldsymbol{\beta}_{1*})) = \phi(\hat{\bar{Y}}_{\rm reg}(\hat{\boldsymbol{\beta}}_1^{(R)}); \bar{Y}, \hat{V}_e(\hat{\boldsymbol{\beta}}_1^{(R)}))\{1 + o_P(1)\}$. □

# S5 Additional simulation results

We here provide additional simulation results. We first considered the same scenarios in the main document with larger sample size, namely, $n = 400$. The results are reported in Tables S1-S4. We can see the almost the same tendency confirmed in Tables 2-5 in the main document.

We next adopted two additional scenarios for the data generating processes. For simplicity, we here consider only linear regression with $n = 300$. In the first scenario, we generate $x_i = (x_{i1}, \ldots, x_{ip^*})$ from a $p^*$-dimensional multivariate normal distribution with vector $(1, \ldots, 1)^t$ and variance-covariance matrix $2R(0.5)$, so the correlations among the covariates are stronger than those in the main document. We set $\beta_0 = 1$ (non-zero intercept) and set the other settings to the same ones in the main document. In the second scenario, we investigate potential effects of unobserved covariates. To this end, we first generated $(w_i, x_i^t)$ from a $(p^* + 1)$-dimensional multivariate normal distribution with vector $(1, \ldots, 1)^t$ and variance-covariance matrix $2R(0.5)$. Then, the response variables $Y_i$ are generated from the following linear regression model:

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

where $\varepsilon_i \sim N(0, 2)$. We set $\delta = 1$, $(\beta_0, \beta_1, \beta_4, \beta_7, \beta_{10}) = (1, 1, -0.5, 1, -0.5)$ and the other $\beta_k$'s are set to zero. Although the covariate $w_i$ is included in the data generating process, we estimate the population mean by using only $x_i$. Therefore, $w_i$ is the unobserved covariate and the working model used in the model-assisted methods is misspecified. Under the two scenarios, we applied the same methods adopted in the main document and evaluated their performance in terms of MSE, Bias, CP and AL, where the results are shown in Tables S5-S8. We can again see the almost the same tendency confirmed in Tables 2-5 in the main document.

# Bibliography


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

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.

Isaki, C. T. and W. A. Fuller (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association 77*, 89–96.

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

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.

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

Wang, Z., J. K. Kim, and S. Yang (2018). Approximate Bayesian inference under informative sampling. *Biometrika 105*, 91–102.

Table S1: Square root of Mean squared errors (RMSE) and bias of point estimators under $p \in \{20, 30, 40, 50\}$ and $n = 400$ in scenarios (A) and (B) with linear regression. All values are multiplied by 100.

| | | (A) | | | | (B) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Method | 20 | 30 | 40 | 50 | 20 | 30 | 40 | 50 |
| | GREG | 10.3 | 10.4 | 10.6 | 10.9 | 9.9 | 10.1 | 10.3 | 10.5 |
| | GREG-L | 10.2 | 10.2 | 10.2 | 10.3 | 9.6 | 9.6 | 9.6 | 9.6 |
| | GREG-R | 10.3 | 10.5 | 10.7 | 10.9 | 9.8 | 10.0 | 10.2 | 10.3 |
| | GREG-V | 10.2 | 10.4 | 10.5 | 10.7 | 9.8 | 9.9 | 10.1 | 10.2 |
| RMSE | GREG-M | 10.3 | 10.4 | 10.6 | 10.9 | 9.9 | 10.1 | 10.3 | 10.5 |
| | AB | 10.3 | 10.5 | 10.7 | 11.0 | 10.0 | 10.2 | 10.5 | 10.6 |
| | ABL | 10.3 | 10.4 | 10.6 | 10.8 | 9.9 | 10.0 | 10.2 | 10.3 |
| | ABH | 10.2 | 10.2 | 10.3 | 10.3 | 9.7 | 9.7 | 9.8 | 9.8 |
| | HT | 14.8 | 14.8 | 14.8 | 14.8 | 12.6 | 12.6 | 12.6 | 12.6 |
| | GREG | -0.15 | -0.13 | -0.18 | -0.22 | 0.43 | 0.93 | 1.39 | 1.86 |
| | GREG-L | -0.22 | -0.22 | -0.25 | -0.24 | 0.12 | 0.18 | 0.25 | 0.34 |
| | GREG-R | -0.18 | -0.18 | -0.22 | -0.26 | 0.44 | 0.94 | 1.39 | 1.87 |
| | GREG-V | -0.21 | -0.22 | -0.24 | -0.23 | 0.27 | 0.62 | 0.95 | 1.30 |
| Bias | GREG-M | -0.15 | -0.13 | -0.18 | -0.22 | 0.43 | 0.93 | 1.39 | 1.86 |
| | AB | -0.17 | -0.16 | -0.20 | -0.24 | 0.43 | 0.93 | 1.37 | 1.85 |
| | ABL | -0.19 | -0.18 | -0.22 | -0.25 | 0.38 | 0.80 | 1.17 | 1.56 |
| | ABH | -0.20 | -0.20 | -0.21 | -0.25 | 0.14 | 0.30 | 0.46 | 0.63 |
| | HT | -0.29 | -0.29 | -0.29 | -0.29 | -0.39 | -0.39 | -0.39 | -0.39 |

Table S2: Coverage probabilities (CP) and average lengths (AL) of 95% confidence/credible intervals under $p \in \{20, 30, 40, 50\}$ and $n = 400$ in scenarios (A) and (B) with linear regression. All values are multiplied by 100.

| | | (A) | | | | (B) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Method | 20 | 30 | 40 | 50 | 20 | 30 | 40 | 50 |
| | GREG | 93.8 | 92.4 | 91.1 | 89.4 | 94.0 | 92.6 | 91.5 | 91.4 |
| | GREG-L | 94.2 | 94.0 | 94.1 | 93.9 | 94.9 | 94.4 | 94.8 | 94.5 |
| | GREG-R | 93.3 | 92.2 | 91.8 | 90.1 | 94.4 | 93.0 | 91.7 | 91.7 |
| | GREG-V | 94.1 | 93.3 | 92.8 | 90.9 | 94.5 | 94.0 | 92.8 | 92.4 |
| CP | GREG-M | 94.0 | 93.0 | 92.8 | 91.6 | 94.9 | 93.7 | 93.5 | 93.5 |
| | AB | 94.2 | 94.2 | 94.5 | 94.4 | 95.6 | 95.5 | 94.3 | 94.7 |
| | ABL | 94.2 | 94.3 | 94.9 | 94.9 | 95.6 | 95.0 | 94.5 | 94.6 |
| | ABH | 94.6 | 94.8 | 94.6 | 94.4 | 95.2 | 95.3 | 95.5 | 95.5 |
| | HT | 94.2 | 94.2 | 94.2 | 94.2 | 94.8 | 94.8 | 94.8 | 94.8 |
| | GREG | 37.4 | 36.9 | 36.4 | 35.9 | 37.6 | 37.1 | 36.6 | 36.1 |
| | GREG-L | 37.9 | 37.8 | 37.7 | 37.7 | 37.7 | 37.6 | 37.5 | 37.4 |
| | GREG-R | 37.5 | 37.0 | 36.6 | 36.2 | 37.4 | 36.9 | 36.4 | 35.9 |
| | GREG-V | 37.6 | 37.2 | 36.8 | 36.4 | 37.8 | 37.5 | 37.2 | 36.8 |
| AL | GREG-M | 38.1 | 38.1 | 38.2 | 38.2 | 38.4 | 38.5 | 38.6 | 38.7 |
| | AB | 39.2 | 39.5 | 39.9 | 40.2 | 39.6 | 40.2 | 40.8 | 41.3 |
| | ABL | 39.0 | 39.2 | 39.5 | 39.6 | 39.4 | 39.8 | 40.2 | 40.5 |
| | ABH | 38.8 | 38.8 | 38.8 | 38.9 | 39.0 | 39.2 | 39.3 | 39.3 |
| | HT | 57.5 | 57.5 | 57.5 | 57.5 | 51.1 | 51.1 | 51.1 | 51.1 |

Table S3: Square root of Mean squared errors (RMSE) and bias of point estimators under $p \in \{20, 30, 40, 50\}$ and $n = 400$ in scenarios (A) and (B) with logistic regression. All values are multiplied by 100.

| | | (A) | | | | (B) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Method | 20 | 30 | 40 | 50 | 20 | 30 | 40 | 50 |
| RMSE | GREG | 1.90 | 1.91 | 1.93 | 1.97 | 1.94 | 1.98 | 2.00 | 2.06 |
| | GREG-L | 1.86 | 1.87 | 1.87 | 1.87 | 1.90 | 1.91 | 1.91 | 1.92 |
| | GREG-R | 1.88 | 1.89 | 1.91 | 1.93 | 1.93 | 1.97 | 1.98 | 2.02 |
| | AB | 1.89 | 1.89 | 1.91 | 1.93 | 1.93 | 1.96 | 1.98 | 2.01 |
| | ABL | 1.88 | 1.88 | 1.89 | 1.90 | 1.91 | 1.91 | 1.89 | 1.88 |
| | ABH | 1.87 | 1.87 | 1.88 | 1.89 | 1.88 | 1.87 | 1.86 | 1.87 |
| | HT | 2.36 | 2.36 | 2.36 | 2.36 | 2.39 | 2.39 | 2.39 | 2.39 |
| Bias | GREG | -0.03 | -0.03 | -0.03 | -0.05 | -0.05 | 0.01 | 0.08 | 0.19 |
| | GREG-L | -0.02 | -0.01 | -0.01 | -0.02 | -0.12 | -0.11 | -0.10 | -0.09 |
| | GREG-R | -0.03 | -0.02 | -0.02 | -0.04 | -0.07 | -0.02 | 0.02 | 0.09 |
| | AB | -0.04 | -0.03 | -0.03 | -0.06 | -0.06 | 0.00 | 0.06 | 0.15 |
| | ABL | -0.03 | -0.02 | -0.01 | 0.00 | -0.07 | -0.02 | 0.02 | 0.08 |
| | ABH | -0.02 | -0.02 | -0.02 | -0.02 | -0.11 | -0.11 | -0.11 | -0.11 |
| | HT | 0.01 | 0.01 | 0.01 | 0.01 | -0.12 | -0.12 | -0.12 | -0.12 |

Table S4: Coverage probabilities (CP) and average lengths (AL) of 95% credible/confidence intervals under $p \in \{20, 30, 40, 50\}$ and $n = 400$ in scenarios (A) and (B) with logistic regression. All values are multiplied by 100.

| | | (A) | | | | (B) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Method | 20 | 30 | 40 | 50 | 20 | 30 | 40 | 50 |
| CP | GREG | 92.8 | 91.2 | 90.6 | 89.5 | 92.9 | 91.8 | 91.5 | 89.6 |
| | GREG-L | 93.0 | 92.9 | 92.9 | 93.0 | 94.3 | 94.3 | 94.5 | 94.0 |
| | GREG-R | 93.2 | 91.9 | 91.0 | 91.0 | 93.3 | 92.6 | 91.7 | 91.9 |
| | AB | 94.4 | 94.6 | 94.6 | 95.0 | 95.2 | 95.5 | 95.8 | 96.0 |
| | ABL | 94.5 | 94.4 | 95.3 | 95.3 | 95.2 | 95.8 | 96.7 | 97.4 |
| | ABH | 94.9 | 94.6 | 94.7 | 95.9 | 96.0 | 96.4 | 96.7 | 97.3 |
| | HT | 95.9 | 95.9 | 95.9 | 95.9 | 95.5 | 95.5 | 95.5 | 95.5 |
| AL | GREG | 7.01 | 6.87 | 6.73 | 6.58 | 7.27 | 7.12 | 6.96 | 6.79 |
| | GREG-L | 7.12 | 7.09 | 7.07 | 7.05 | 7.40 | 7.37 | 7.35 | 7.32 |
| | GREG-R | 7.09 | 6.98 | 6.88 | 6.81 | 7.36 | 7.24 | 7.14 | 7.05 |
| | AB | 7.47 | 7.57 | 7.67 | 7.78 | 7.80 | 7.92 | 8.06 | 8.24 |
| | ABL | 7.49 | 7.60 | 7.72 | 7.86 | 7.80 | 7.93 | 8.06 | 8.22 |
| | ABH | 7.45 | 7.54 | 7.65 | 7.78 | 7.75 | 7.83 | 7.94 | 8.06 |
| | HT | 9.60 | 9.60 | 9.60 | 9.60 | 9.53 | 9.53 | 9.53 | 9.53 |

Table S5: RMSE and bias of point estimators under $p \in \{20, 30, 40, 50\}$ and $n = 300$ in scenarios (A) and (B), where the simulated data is generated from linear regression with non-zero intercept and higher correlations among covariates. All values are multiplied by 100.

| | | (A) | | | | (B) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Method | 20 | 30 | 40 | 50 | 20 | 30 | 40 | 50 |
| RMSE | GREG | 11.9 | 12.0 | 12.1 | 12.4 | 11.7 | 11.7 | 11.9 | 12.2 |
| | GREG-L | 11.5 | 11.5 | 11.5 | 11.5 | 11.5 | 11.5 | 11.4 | 11.5 |
| | GREG-R | 11.7 | 11.7 | 11.8 | 12.1 | 11.7 | 11.9 | 12.1 | 12.3 |
| | GREG-V | 11.9 | 11.9 | 12.0 | 12.2 | 11.6 | 11.6 | 11.7 | 11.8 |
| | GREG-M | 11.9 | 12.0 | 12.1 | 12.4 | 11.7 | 11.7 | 11.9 | 12.2 |
| | AB | 11.7 | 11.8 | 12.0 | 12.4 | 11.2 | 11.4 | 11.8 | 12.3 |
| | ABL | 11.6 | 11.6 | 11.8 | 12.1 | 11.0 | 11.2 | 11.3 | 11.6 |
| | ABH | 11.5 | 11.5 | 11.6 | 11.7 | 10.8 | 10.9 | 10.9 | 10.9 |
| | HT | 16.2 | 16.2 | 16.2 | 16.2 | 13.6 | 13.6 | 13.6 | 13.6 |
| Bias | GREG | -1.08 | -1.23 | -1.26 | -1.35 | -0.53 | 0.33 | 1.07 | 1.96 |
| | GREG-L | 0.02 | 0.02 | 0.03 | 0.06 | -0.02 | 0.18 | 0.33 | 0.46 |
| | GREG-R | 0.11 | 0.05 | 0.10 | 0.13 | 0.76 | 1.75 | 2.54 | 3.31 |
| | GREG-V | -0.56 | -0.71 | -0.73 | -0.66 | -0.27 | 0.46 | 0.97 | 1.60 |
| | GREG-M | -1.08 | -1.23 | -1.26 | -1.35 | -0.53 | 0.33 | 1.07 | 1.96 |
| | AB | 0.16 | 0.12 | 0.22 | 0.21 | 1.07 | 2.17 | 3.11 | 4.08 |
| | ABL | 0.13 | 0.09 | 0.15 | 0.15 | 0.86 | 1.74 | 2.45 | 3.09 |
| | ABH | 0.10 | 0.04 | 0.05 | 0.07 | 0.28 | 0.62 | 0.90 | 1.17 |
| | HT | -0.08 | -0.08 | -0.08 | -0.08 | -1.61 | -1.61 | -1.61 | -1.61 |

Table S6: CP and AL of 95% credible/confidence intervals under $p \in \{20, 30, 40, 50\}$ and $n = 300$ in scenarios (A) and (B), where the simulated data is generated from linear regression with non-zero intercept and higher correlations among covariates. All values are multiplied by 100.

| | | (A) | | | | (B) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Method | 20 | 30 | 40 | 50 | 20 | 30 | 40 | 50 |
| CP | GREG | 94.3 | 93.2 | 92.2 | 90.9 | 94.2 | 93.3 | 92.6 | 90.7 |
| | GREG-L | 94.3 | 94.5 | 93.9 | 93.8 | 95.6 | 95.5 | 95.3 | 95.3 |
| | GREG-R | 94.4 | 93.9 | 93.5 | 92.3 | 94.2 | 93.6 | 92.0 | 90.9 |
| | GREG-V | 94.2 | 94.3 | 93.8 | 93.4 | 94.7 | 94.2 | 93.9 | 93.4 |
| | GREG-M | 94.7 | 94.7 | 94.2 | 93.6 | 95.3 | 95.3 | 94.4 | 92.9 |
| | AB | 95.1 | 95.1 | 95.1 | 94.3 | 95.4 | 95.1 | 95.1 | 94.4 |
| | ABL | 95.1 | 95.1 | 95.0 | 94.7 | 95.6 | 95.2 | 95.7 | 95.4 |
| | ABH | 95.4 | 94.9 | 94.8 | 94.7 | 95.9 | 95.7 | 95.6 | 95.9 |
| | HT | 95.5 | 95.5 | 95.5 | 95.5 | 95.4 | 95.4 | 95.4 | 95.4 |
| AL | GREG | 44.0 | 43.0 | 42.1 | 41.2 | 44.5 | 43.6 | 42.6 | 41.7 |
| | GREG-L | 43.8 | 43.6 | 43.6 | 43.5 | 44.4 | 44.2 | 44.1 | 43.9 |
| | GREG-R | 43.2 | 42.6 | 42.1 | 41.5 | 43.8 | 43.1 | 42.5 | 41.9 |
| | GREG-V | 44.5 | 43.8 | 43.1 | 42.4 | 45.0 | 44.4 | 43.8 | 43.2 |
| | GREG-M | 45.0 | 44.8 | 44.7 | 44.7 | 45.4 | 45.4 | 45.4 | 45.5 |
| | AB | 45.8 | 46.3 | 46.8 | 47.3 | 44.6 | 45.6 | 46.6 | 47.5 |
| | ABL | 45.5 | 45.8 | 46.0 | 46.2 | 44.1 | 44.6 | 45.2 | 45.6 |
| | ABH | 45.1 | 45.1 | 45.1 | 45.1 | 43.5 | 43.6 | 43.7 | 43.8 |
| | HT | 64.4 | 64.4 | 64.4 | 64.4 | 54.0 | 54.0 | 54.0 | 54.0 |

Table S7: RMSE and bias of point estimators under $p \in \{20, 30, 40, 50\}$ and $n = 300$ in scenarios (A) and (B), where the simulated data is generated from linear regression with non-zero intercept and an unobserved covariate. All values are multiplied by 100.

| | | (A) | | | | (B) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Method | 20 | 30 | 40 | 50 | 20 | 30 | 40 | 50 |
| RMSE | GREG | 12.4 | 12.5 | 12.8 | 13.0 | 11.7 | 11.9 | 12.2 | 12.6 |
| | GREG-L | 12.1 | 12.2 | 12.2 | 12.3 | 11.5 | 11.6 | 11.7 | 11.7 |
| | GREG-R | 12.2 | 12.5 | 12.7 | 12.9 | 11.8 | 12.2 | 12.7 | 13.1 |
| | GREG-V | 12.4 | 12.5 | 12.6 | 12.7 | 11.6 | 11.8 | 12.0 | 12.2 |
| | GREG-M | 12.4 | 12.5 | 12.8 | 13.0 | 11.7 | 11.9 | 12.2 | 12.6 |
| | AB | 12.2 | 12.5 | 12.7 | 13.0 | 11.0 | 11.5 | 12.2 | 12.7 |
| | ABL | 12.1 | 12.4 | 12.5 | 12.7 | 10.9 | 11.2 | 11.6 | 11.9 |
| | ABH | 12.1 | 12.1 | 12.2 | 12.3 | 10.7 | 10.8 | 10.8 | 10.9 |
| | HT | 17.7 | 17.7 | 17.7 | 17.7 | 14.1 | 14.1 | 14.1 | 14.1 |
| Bias | GREG | -1.67 | -1.79 | -1.88 | -2.01 | -0.19 | 0.72 | 1.49 | 2.43 |
| | GREG-L | -0.13 | -0.12 | -0.14 | -0.09 | 0.74 | 1.06 | 1.17 | 1.35 |
| | GREG-R | -0.16 | -0.04 | 0.00 | -0.03 | 1.50 | 2.54 | 3.42 | 4.33 |
| | GREG-V | -0.97 | -1.05 | -1.11 | -1.20 | 0.22 | 0.87 | 1.59 | 2.21 |
| | GREG-M | -1.67 | -1.79 | -1.88 | -2.01 | -0.19 | 0.72 | 1.49 | 2.43 |
| | AB | -0.21 | -0.10 | -0.07 | -0.12 | 1.61 | 2.72 | 3.65 | 4.67 |
| | ABL | -0.22 | -0.09 | -0.08 | -0.13 | 1.43 | 2.32 | 3.03 | 3.76 |
| | ABH | -0.21 | -0.16 | -0.15 | -0.14 | 0.85 | 1.21 | 1.49 | 1.81 |
| | HT | 0.66 | 0.66 | 0.66 | 0.66 | -0.75 | -0.75 | -0.75 | -0.75 |

Table S8: CP and AL of 95% credible/confidence intervals under $p \in \{20, 30, 40, 50\}$ and $n = 300$ in scenarios (A) and (B), where the simulated data is generated from linear regression with non-zero intercept and an unobserved covariate. All values are multiplied by 100.

| | | (A) | | | | (B) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Method | 20 | 30 | 40 | 50 | 20 | 30 | 40 | 50 |
| CP | GREG | 92.9 | 91.9 | 90.9 | 88.5 | 95.0 | 93.9 | 91.9 | 91.0 |
| | GREG-L | 93.0 | 92.4 | 92.4 | 92.5 | 95.8 | 94.7 | 94.6 | 94.6 |
| | GREG-R | 92.2 | 91.8 | 91.0 | 89.0 | 94.3 | 92.5 | 90.5 | 89.4 |
| | GREG-V | 93.7 | 92.7 | 91.8 | 91.4 | 95.2 | 94.8 | 93.5 | 92.8 |
| | GREG-M | 93.6 | 93.1 | 92.9 | 91.4 | 95.0 | 95.2 | 93.8 | 93.7 |
| | AB | 93.8 | 93.6 | 93.2 | 93.5 | 95.6 | 95.2 | 94.5 | 92.8 |
| | ABL | 93.5 | 93.1 | 93.6 | 92.9 | 96.1 | 95.4 | 94.5 | 94.1 |
| | ABH | 94.0 | 93.7 | 93.2 | 93.0 | 96.2 | 96.4 | 95.5 | 95.8 |
| | HT | 94.3 | 94.3 | 94.3 | 94.3 | 95.2 | 95.2 | 95.2 | 95.2 |
| AL | GREG | 45.0 | 44.0 | 43.1 | 42.1 | 45.7 | 44.6 | 43.7 | 42.6 |
| | GREG-L | 44.6 | 44.5 | 44.3 | 44.3 | 45.4 | 45.2 | 45.1 | 44.9 |
| | GREG-R | 44.0 | 43.4 | 42.8 | 42.3 | 44.8 | 44.1 | 43.4 | 42.7 |
| | GREG-V | 45.6 | 44.8 | 44.1 | 43.4 | 46.3 | 45.6 | 44.9 | 44.2 |
| | GREG-M | 46.0 | 45.8 | 45.7 | 45.6 | 46.6 | 46.5 | 46.5 | 46.5 |
| | AB | 46.6 | 47.2 | 47.7 | 48.2 | 45.1 | 46.0 | 47.1 | 48.0 |
| | ABL | 46.4 | 46.7 | 46.9 | 47.0 | 44.6 | 45.2 | 45.7 | 46.1 |
| | ABH | 46.0 | 46.0 | 46.0 | 46.0 | 43.9 | 44.1 | 44.3 | 44.3 |
| | HT | 67.1 | 67.1 | 67.1 | 67.1 | 55.3 | 55.3 | 55.3 | 55.3 |