On high-dimensional variance estimation in survey sampling

Esther Eustache[(a)] , Mehdi Dagdoug[(b)] and David Haziza[(c)]

(a) University of Lausanne, Institute of Sport Sciences, Lausanne, Switzerland
(b) McGill University, Department of Mathematics and Statistics, Montreal, Canada
(c) University of Ottawa, Department of Mathematics and Statistics, Ottawa, Canada

Abstract

Utilizing predictive modeling at different survey stages can improve the accuracy of a point estimator or help tackle issues such as missing values. So far, the existing literature on predictive models for survey data has predominantly concentrated on scenarios with low-dimensional data, wherein the number of variables is small compared to the sample size. In this paper, assuming a linear regression model, we show that customary variance estimators based on a first Taylor expansion or jackknife may suffer from substantial bias in a high-dimensional setting. We explain why this is so through a mix of theoretical and empirical investigations. We propose some bias-adjusted variance estimators and show, theoretically and empirically, that the proposed variance estimators perform well in terms of bias, even in a high-dimensional setting.

Key words: Bias-adjusted variance estimator; Generalized regression estimator; Jackknife variance estimation; Linear regression imputation; Taylor-based variance estimator.

1 Introduction

Predictive modeling can be applied at various stages of a survey to enhance the precision of a point estimator and to address the problem of missing values, among others. Using predictive models enables us to exploit a relationship between a survey variable Y and a set of predictors $X_1, X_2, \ldots, X_p$. For instance, model-assisted estimation procedures use a set of predicted values to improve the efficiency of point estimators; e.g., see Särndal (1992) and Breidt and Opsomer (2017). To mitigate the potential nonresponse bias caused by item nonresponse, it is common practice to employ some form of imputation, which involves generating a set of predictions to substitute for the missing values; e.g., see Haziza (2009) and Chen and Haziza (2019).

The literature on predictive models for survey data has primarily focused on low-dimensional data settings, where the number of variables p is small relative to the sample size n. Formalized mathematically, it means that $p/n \to 0$. Some notable exceptions include Cardot et al. (2017), Ta et al. (2020), Chauvet and Goga (2022) and Dagdoug et al. (2022). With the advent of big data sets, moderate to high-dimensional settings are becoming more prevalent. In this article, a high-dimensional setting refers to a situation where the number of predictors p is of the same magnitude as the sample size n so that $p/n \to \kappa \in (0, 1)$.

High-dimensional linear regression models pose some challenges compared to traditional linear regression models with fewer predictors. In this paper, we focus on variance estimation, an important aspect for national statistical offices, which routinely produce measures such as the coefficient of variation for point estimates or confidence intervals. This work constitutes a first step in understanding the behavior of common variance estimators in high-dimensional settings. We show that the common variance estimation procedures tend to break down when $p/n \to \kappa \in (0, 1) > 0$. Specifically, variance estimators based on first-order Taylor expansions tend to underestimate the variance of point estimators, while resampling procedures, such as the jackknife and bootstrap, may substantially overestimate the true variance. In a low dimensional setting, resampling methods in surveys are discussed in Wolter (2007), Mashreghi et al. (2016), Wang et al. (2022), and Stefan and Hidiroglou (2024).

The contributions of this paper are as follows: (i) We explain why variance estimators based on first-order Taylor expansions and jackknife variance estimators tend to break down, using a mix of empirical and theoretical investigations. We consider two distinct setups involving the customary linear regression model: (a) the model-assisted estimation framework, utilizing the generalized regression estimator (see, e.g., Särndal, 1980; Särndal, 1992; Särndal, 2007); and (b) the deterministic linear regression imputation framework (e.g., Chen and Haziza, 2019). (ii) In the context of Bernoulli sampling and simple random sampling without replacement, we propose bias-adjusted variance estimators, that are shown to work well, at least in our experiments.

We adopt the following notations. Let $U := \{1, 2, ..., N\}$ be a finite population of size N. Our interest lies in estimating the finite population mean

$$\mu_y := \frac{1}{N} \sum_{k \in U} y_k,$$

of a survey variable Y, where y_k denotes the y-value attached to unit k. We select a sample, S, of (expected) size n, according to a probability sampling design $\mathcal{P}(S \mid \boldsymbol{Z})$, where $\boldsymbol{Z} \in \mathbb{R}^{N \times d}$ denotes the matrix of design information. We restrict our attention to non-informative sampling designs; see, e.g., Pfeffermann and Sverchkov (2009). The sample S is fully characterized by the vector of sample selection indicators, $\boldsymbol{I} := [I_1, I_2, \ldots, I_N]^\top$, where $I_k := 1$ if $k \in S$, and $I_k := 0$, otherwise. We denote by $\pi_k := \mathbb{P}(I_k = 1) > 0$ and $\pi_{k\ell} := \mathbb{P}(I_k = 1, I_\ell = 1) > 0$, for $k, \ell \in U$, the first-order and the second-order inclusion probabilities, respectively.

2 Linear prediction in survey sampling

We consider the customary linear regression model:

$$y_k = \mathbf{x}_k^\top \boldsymbol{\beta} + \epsilon_k, \qquad k \in U, \tag{1}$$

where $\boldsymbol{\beta}$ is a p-vector of unknown coefficients and the errors ϵ_k are independent and identically distributed variables such that $\mathbb{E}[\epsilon_k | \mathbf{x}_k] = 0$, and $\mathbb{E}\left[\epsilon_k^2 | \mathbf{x}_k\right] := \sigma^2 < \infty$. We assume that the intercept is included in the covariates; i.e., the first component of $\mathbf{x}_k$ is 1 for all $k \in U$.

Although we assume a homoscedastic variance structure, our results can be extended to the case of a heteroscedastic variance structure.

Below, we use the notation $\boldsymbol{y}_S \in \mathbb{R}^{n_s}$ and $\boldsymbol{X}_S \in \mathbb{R}^{n_s \times p}$ to denote the vector of y-values and the design matrix corresponding to the sample, respectively. Also, we use $\boldsymbol{\Pi}_S \in \mathbb{R}^{n_s \times n_s}$ to denote a diagonal matrix, whose k-th diagonal element is π_k.

2.1 Model-assisted estimation

In this section, we assume that the observed data are given by

$$\mathcal{D}_{ma} := \{(\mathbf{x}_k, y_k) \ ; \ k \in S\}.$$

In addition, we assume that the vector of population totals,

$$\mathbf{t}_{\mathbf{x}} := \left[\sum_{k \in U} x_{k1}, \sum_{k \in U} x_{k2}, ..., \sum_{k \in U} x_{kp}\right]^\top,$$

is available from an external source. The Generalized REGression (GREG) estimator of μ_y is defined by

$$\widehat{\mu}_{greg} := \frac{1}{N}\left(\sum_{k \in U} \mathbf{x}_k^\top \widehat{\boldsymbol{\beta}}_S + \sum_{k \in S} \frac{y_k - \mathbf{x}_k^\top \widehat{\boldsymbol{\beta}}_S}{\pi_k}\right), \tag{2}$$

where

$$\widehat{\boldsymbol{\beta}}_S = \left(\boldsymbol{X}_S^\top \boldsymbol{\Pi}_S^{-1} \boldsymbol{X}_S\right)^{-1} \boldsymbol{X}_S \boldsymbol{\Pi}_S^{-1} \boldsymbol{y}_S$$

is the weighted least squares estimator of $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}}_S := \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{k \in S} \frac{\left(y_k - \mathbf{x}_k^\top \boldsymbol{\beta}\right)^2}{\pi_k}. \tag{3}$$

Throughout the paper, we assume, for simplicity, that the $p \times p$ matrix, $\boldsymbol{A}_{\Pi S} := \boldsymbol{X}_S^\top \boldsymbol{\Pi}_S^{-1} \boldsymbol{X}_S$, is non-singular.

In our setting, the GREG estimator can be written in the so-called projection form:

$$\widehat{\mu}_{greg} = \frac{1}{N}\sum_{k\in U}\mathbf{x}_k^\top\widehat{\boldsymbol{\beta}}_S$$

since

$$\sum_{k\in S}\pi_k^{-1}\widehat{\epsilon}_{kS} = 0,$$

where $\widehat{\epsilon}_{kS} := y_k - \mathbf{x}_k^\top\widehat{\boldsymbol{\beta}}_S$ denotes the sample residual attached to unit $k \in S$, (Särndal et al., 1992, Chapter 6).

In this section, as the reference distribution for studying the properties of variance estimators, we use the joint distribution induced by the superpopulation model (1) and the sampling design. Consider the following decomposition:

$$\mathbb{V}_{mp}\left(\widehat{\mu}_{greg}\right) = \mathbb{E}_m\left[\mathbb{V}_p\left(\widehat{\mu}_{greg}\right)\right] + \mathbb{V}_m\left(\mathbb{E}_p\left[\widehat{\mu}_{greg}\right]\right), \tag{4}$$

where the subscripts p and m are used to denote the sampling design and the model, respectively.

Based on (4), an estimator of the variance of $\widehat{\mu}_{greg}$ based on a first-order Taylor expansion is given by

$$\widehat{V}_{\text{tay}} = \frac{1}{N^2}\sum_{k\in S}\sum_{\ell\in S}\frac{\Delta_{k\ell}}{\pi_{k\ell}}\frac{\widehat{\epsilon}_{kS}}{\pi_k}\frac{\widehat{\epsilon}_{\ell S}}{\pi_\ell} + \frac{\widehat{\sigma}^2}{N}, \tag{5}$$

where $\widehat{\sigma}^2$ is defined by

$$\widehat{\sigma}^2 = \frac{1}{n_s - p}\sum_{k\in S}(y_k - \mathbf{x}_k^\top\widehat{\boldsymbol{\beta}}_S)^2. \tag{6}$$

The first term on the right side of (5) is the focus of this article. Indeed, the noise variance estimator $\widehat{\sigma}^2$ in (6) is unbiased for any values of p and n. In view of that, in our proofs, we assume without loss of generality that σ^2 is known.

As an alternative to the first term on the right side of (5), Särndal et al. (1989) advocated the use of a g-weighted version, leading to

$$\widehat{V}_g = \frac{1}{N^2}\sum_{k\in S}\sum_{\ell\in S}\frac{\Delta_{k\ell}}{\pi_{k\ell}}\frac{g_k\widehat{\epsilon}_{kS}}{\pi_k}\frac{g_\ell\widehat{\epsilon}_{\ell S}}{\pi_\ell} + \frac{\widehat{\sigma}^2}{N}, \tag{7}$$

where

$$g_k := 1 + \left(\mathbf{t_x} - \widehat{\mathbf{t}}_{\mathbf{x},\pi}\right)^\top \boldsymbol{A}_{\Pi S}^{-1}\mathbf{x}_k, \qquad k\in S, \tag{8}$$

is the so-called g-weight attached to unit $k \in S$, with $\widehat{\mathbf{t}}_{\mathbf{x},\pi}$ denoting the Horvitz-Thompson estimator of $\mathbf{t_x}$. Since the intercept is included in the model and the variance structure is assumed to be homoscedastic, the g-weight in (8) reduces to

$$g_k = \mathbf{t}_\mathbf{x}^\top \boldsymbol{A}_{\Pi S}^{-1}\mathbf{x}_k, \qquad k\in S. \tag{9}$$

Jackknife variance estimation for the GREG estimator has been discussed in Yung and Rao (1996), Duchesne (2000) and Valliant (2002), among others. Here, we consider the generalized jackknife variance estimator of Campbell (1980) and Berger and Skinner (2005). Let $\widetilde{h}^\pi_{kk} := \mathbf{x}_k^\top \boldsymbol{A}_{\Pi S}^{-1} d_k \mathbf{x}_k$ be the survey weighted leverage of element $k\in S$, with $d_k = \pi_k^{-1}$. The next proposition gives a closed-form expression of the generalized jackknife variance estimator in the case of the GREG estimator.

Proposition 2.1. *An estimator of (4) based on the generalized jackknife variance estimator of Berger and Skinner (2005) has a closed-formed formula given by*

$$\widehat{V}_{\text{jack}} = \frac{1}{N^2}\sum_{k\in S}\sum_{\ell\in S}\frac{\Delta_{k\ell}}{\pi_{k\ell}}\frac{(1-w_k)\,g_k\,\widehat{\epsilon}_{kS}}{\left(1-\widetilde{h}^\pi_{kk}\right)\pi_k}\frac{(1-w_\ell)\,g_\ell\,\widehat{\epsilon}_{\ell S}}{\left(1-\widetilde{h}^\pi_{\ell\ell}\right)\pi_\ell} + \frac{\widehat{\sigma}^2}{N}, \tag{10}$$

where $w_k := (N\pi_k)^{-1}$ *for* $k \in S$.

Proof. See Appendix B.1. ■

2.2 Deterministic linear regression imputation

Predictions based on a linear regression model are also used in the context of imputation for item nonresponse. In this context, the survey variable Y is observed only for a subset $S_r \subseteq S$, called the set of respondents to item Y. We denote by $S_m = S - S_r$ the set of nonrespondents to item Y. Let $\boldsymbol{R} := [R_1, R_2, \ldots, R_N]^\top$ be the N-vector of response indicators, where $R_k = 1$ if $k \in S_r$, and $R_k = 0$, otherwise. Here, the predictors $X_1, \ldots, X_p$, are assumed to be available for both the respondents and the nonrespondents. We assume that: (i) The data $\{(\mathbf{x}_k, y_k, r_k)\}_{k \in U}$ are identically and independently distributed; (ii) The data are Missing At Random (Rubin, 1976, MAR):

$$\mathbb{P}\left(R_k = 1 | \mathbf{x}_k, y_k\right) = \mathbb{P}\left(R_k = 1 | \mathbf{x}_k\right);$$

(iii) The positivity assumption is satisfied; i.e., $\mathbb{P}\left(R_k = 1 | \mathbf{x}_k\right) > 0$, almost surely. Available to the imputer are the data

$$\mathcal{D}_{imp} := \{(\mathbf{x}_k, y_k) \ ; \ k \in S_r\} \bigcup \{\mathbf{x}_k \ ; \ k \in S_m\}.$$

Imputation involves estimating the relationship between Y and $X_1, \ldots, X_p$ based on the respondents and applying this relationship to the nonrespondents.

An estimator of μ_y after deterministic linear imputation is given by

$$\widehat{\mu}_{lr} := \frac{1}{\widehat{N}} \sum_{k \in S} \frac{\widetilde{y}_k}{\pi_k}, \tag{11}$$

where $\widehat{N} := \sum_{k \in S} \pi_k^{-1}$ and $\widetilde{y}_k := R_k y_k + (1 - R_k)\, \mathbf{x}_k^\top \widehat{\boldsymbol{\beta}}_R$ with

$$\widehat{\boldsymbol{\beta}}_R = \left(\sum_{k \in S_r} \frac{\mathbf{x}_k \mathbf{x}_k^\top}{\pi_k} \right)^{-1} \sum_{k \in S_r} \frac{\mathbf{x}_k y_k}{\pi_k} = \left(\boldsymbol{X}_R \boldsymbol{\Pi}_R^{-1} \boldsymbol{X}_R^\top \right)^{-1} \boldsymbol{X}_R \boldsymbol{\Pi}_R^{-1} \boldsymbol{y}_R \tag{12}$$

denoting the weighted least squares estimator of $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}}_R := \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min} \sum_{k \in S_r} \frac{\left(y_k - \mathbf{x}_k^\top \boldsymbol{\beta}\right)^2}{\pi_k}.$$

In (12), the quantities $\boldsymbol{X}_R$, $\boldsymbol{\Pi}_R$ and $\boldsymbol{y}_R$ correspond to the counterparts of $\boldsymbol{X}_S$, $\boldsymbol{\Pi}_S$ and $\boldsymbol{y}_S$, respectively, restricted to the set of respondents S_r.

To estimate the variance of $\widehat{\mu}_{lr}$, we consider the reverse framework, originally proposed by Fay (1991) and Shao and Steel (1999); see also Kim and Rao (2009) and Haziza and Vallée (2020). Using this framework, the total variance of $\widehat{\mu}_{lr}$ can be expressed as

$$\mathbb{V}(\widehat{\mu}_{lr}) = \mathbb{E}_m \mathbb{E}_q \mathbb{V}_p(\widehat{\mu}_{lr}) + \mathbb{E}_q \mathbb{V}_m \mathbb{E}_p(\widehat{\mu}_{lr} - \mu_y), \tag{13}$$

where the subscript q denotes the nonresponse mechanism. Let us define $\widehat{h}_{k\ell}^{\pi} := \mathbf{x}_k^\top \boldsymbol{A}_{\Pi R}^{-1} d_\ell \mathbf{x}_\ell$ and $\widehat{\Gamma}_k := \sum_{\ell \in S_m} \widehat{h}_{k\ell}^{\pi}$, where $\boldsymbol{A}_{\Pi R} := \boldsymbol{X}_R^\top \boldsymbol{\Pi}_R^{-1} \boldsymbol{X}_R$. Again, we assume that the matrix $\boldsymbol{A}_{\Pi R}$ is non-singular.

Proposition 2.2. *An estimator of the variance of $\widehat{\mu}_{lr}$ based on a first-order Taylor expansion is given by*

$$\widehat{V}_{I,\text{tay}} = \frac{1}{\widehat{N}^2} \sum_{k \in S} \sum_{\ell \in S} \frac{\Delta_{k\ell}}{\pi_{k\ell}} \frac{\widehat{\xi}_k - \widehat{\mu}_{lr}}{\pi_k} \frac{\widehat{\xi}_\ell - \widehat{\mu}_{lr}}{\pi_\ell} + \frac{\widehat{\sigma}^2}{\widehat{N}^2} \sum_{k \in S_r} \frac{1}{\pi_k} \left\{ 1 - R_k(1 + \widehat{\Gamma}_k) \right\}^2, \tag{14}$$

where

$$\widehat{\sigma}^2 = \frac{1}{n_r - p} \sum_{k \in S_r} (y_k - \mathbf{x}_k^\top \widehat{\boldsymbol{\beta}}_R)^2. \tag{15}$$

and

$$\widehat{\xi}_k := \widetilde{y}_k + R_k \widehat{\Gamma}_k \, \widehat{\epsilon}_{kR} \tag{16}$$

with $\widehat{\epsilon}_{kR} = y_k - \mathbf{x}_k^\top \widehat{\boldsymbol{\beta}}_R$, $k \in S_r$.

The proof of Proposition 2.2 is straightforward and is thus omitted. We use a slight abuse of notation here: $\widehat{\sigma}^2$ in (15) is different than that of (6); the former is to be used in a nonresponse framework while the latter should be used with a model-assisted framework.

We now turn to jackknife variance estimation in the context of deterministic linear regression. Berger and Rao (2006) extended the results of Berger and Skinner (2005) to the case of mean and ratio imputation. We extend the results of Berger and Rao (2006) to the more general setting of deterministic linear regression imputation.

Result 2.1. *The estimator of the variance of $\widehat{\mu}_{lr}$ based on the generalized jackknife variance estimator of Berger and Rao (2006) has a closed-form expression given by*

$$\widehat{V}_{I,\text{jack}} = \frac{1}{\widehat{N}^2}\sum_{k\in S}\sum_{\ell\in S}\frac{\Delta_{k\ell}}{\pi_{k\ell}}\frac{\widehat{\mu}_{lr}-\widehat{\xi}_k^{(jk)}}{\pi_k}\frac{\widehat{\mu}_{lr}-\widehat{\xi}_\ell^{(jk)}}{\pi_\ell} + \frac{\widehat{\sigma}^2}{\widehat{N}^2}\sum_{k\in S_r}\frac{1}{\pi_k}\left\{1-R_k(1+\widehat{\Gamma}_k)\right\}^2, \quad (17)$$

where

$$\widehat{\xi}_k^{(jk)} := \widetilde{y}_k + R_k\widehat{\Gamma}_k\frac{\widehat{\epsilon}_{kR}}{1-\widehat{h}_{kk}^{\pi}}. \quad (18)$$

Proof. See Appendix B.2. ■

3 Behavior of some commonly used variance estimators: Empirical studies

In this section, we present the results of two simulation studies. In Section 3.1, we first examine the empirical performance of the variance estimators described in Section 2.1, whereas Section 3.2 considers the variance estimators discussed in Section 2.2. In the model-assisted setup, we denote by $\kappa = p/n$, the ratio of the number of predictors to the expected sample size. In the linear regression imputation setup, we denote by $\kappa = p/\mathbb{E}[n_r]$, the ratio of the number of predictors to the expected number of respondents.

3.1 Model-assisted estimation: the GREG estimator

We generated a finite population U of size $N = 5,000$ consisting of 203 explanatory variables $X_1,\dots,X_{203}$, and a survey variable Y. The variables $X_1,\dots,X_{203}$, were generated from a multivariate normal distribution with a mean vector equal to $5\times\mathbf{1}^\top$ and correlation matrix, whose diagonal elements were equal to 1 and off-diagonal elements equal to 0.3, where $\mathbf{1}$ denotes the vector of ones. Given the X-variables, we generated a survey variable Y according to the linear regression model

$$y_k = 14 - 4x_{1k} + 3x_{2k} + 4x_{3k} + \epsilon_k, \quad (19)$$

where the errors ϵ_k were generated from a normal distribution with mean equal to 0 and variance equal to 20. This led to a model R^2 approximately equal to 0.6. In (19), note that only the first three variables X_1, X_2, and X_3 were used for generating the Y-variable.

From the population, we selected $R = 10,000$ samples of (expected) sample size $n = 300$, according to two sampling designs: simple random sampling design without replacement and Bernoulli sampling. In each sample, we computed several GREG estimators, $\widehat{\mu}_{greg}$, given by (2), based on different sets of explanatory variables. In addition to X_1, X_2 and X_3, we included several noise variables denoted by p_{noise}. The values for p_{noise} were set to: 0, 20, 40, 60, 80, 100, 120, 140, 160, 180, and 200. This led to 12 estimators, $\widehat{\mu}_{greg}$, of μ_y. To estimate the variance of $\widehat{\mu}_{greg}$, we computed $\widehat{V}_{\text{tay}}$ given by (5), $\widehat{V}_g$ given by (7), and $\widehat{V}_{\text{jack}}$ given by (10).

As a measure of bias of a variance estimator, we computed its Monte Carlo percent relative bias (RB). Using the generic notations $\widehat{\mu}$ and $\widehat{V}$ for a point and a variance estimator, respectively, the RB of $\widehat{V}$ is defined as

$$RB(\widehat{V}) := 100 \times \frac{1}{R} \sum_{r=1}^{R} \frac{\widehat{V}^{(r)} - V_{MC}(\widehat{\mu})}{V_{MC}(\widehat{\mu})}, \tag{20}$$

where $V_{MC}(\widehat{\mu})$ denotes the Monte-Carlo variance of $\widehat{\mu}$ and $\widehat{V}^{(r)}$ denotes the estimator $\widehat{V}$ at the rth iteration, $r = 1, \ldots, 10,000$. The results of Bernoulli sampling and for simple random sampling without replacement are shown in Figures 1 and 2, respectively.

From Figures 2 and 1, we note that the three variance estimators performed well for small values of κ. For instance, for $\kappa = 3/300$, which corresponds to the case of $p_{noise} = 0$, the estimator $\widehat{V}_{\text{tay}}$ exhibited a value of RB of about -1.7% for Bernoulli sampling and -0.7% for simple random sampling without replacement. The g-weighted version $\widehat{V}_g$ showed a bias of -0.6% for Bernoulli sampling and of 0.2% for simple random sampling without replacement. The jackknife variance estimator $\widehat{V}_{\text{jack}}$ showed a bias of about 2.0% for Bernoulli sampling and 2.8% for simple random sampling without replacement. However, for $\kappa = 83/300 \approx 0.28$, the RB of $\widehat{V}_{\text{tay}}$ was equal to -46% in the case of Bernoulli sampling and -44.7% in the case of simple random sampling without replacement. The magnitude of the underestimation got worse as p/n increased. The g-weighted version $\widehat{V}_g$ did better than $\widehat{V}_{\text{tay}}$ with a value of RB equal to -28% for Bernoulli sampling, and equal to -26.3% for simple random sampling without replacement. On the other hand, the jackknife variance estimator exhibited significant overestimation with values of RB equal to 36.5% in the case of Bernoulli sampling and 38.8% in the case of simple random sampling without replacement. The magnitude of the bias increased as κ increased.

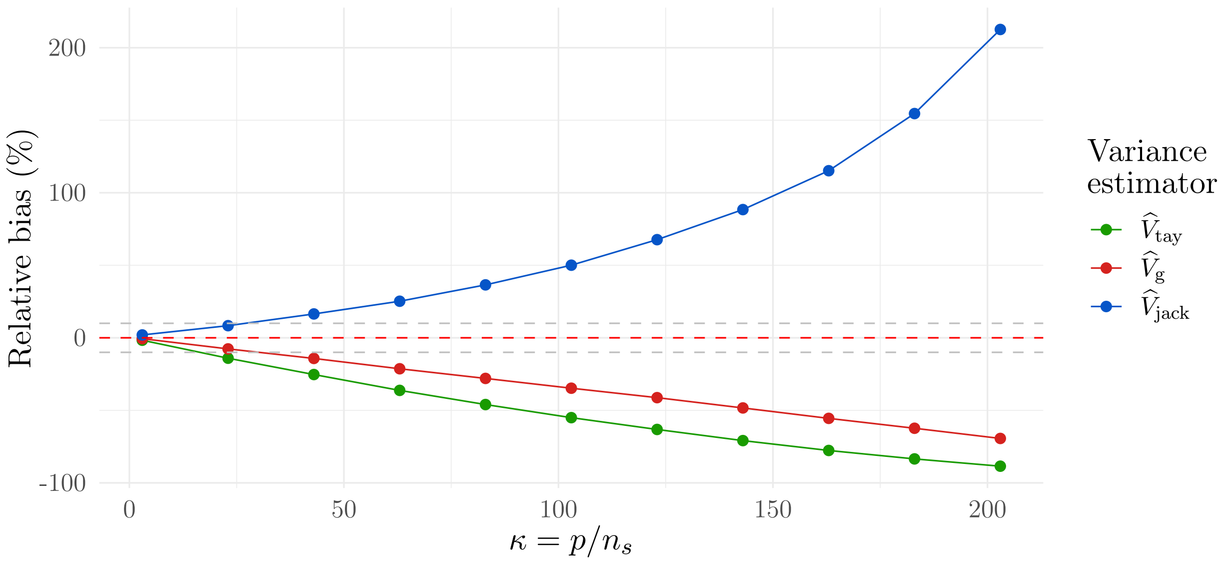


Figure 1: Behaviour of three variance estimators for $\widehat{\mu}_{greg}$ under Bernoulli sampling.

3.2 Deterministic linear regression imputation

We started by generating $5,000$ realizations of a vector of explanatory variables, of size 103, from a multivariate normal distribution with a mean vector equal to $5 \times \mathbf{1}^\top$ and correlation matrix, whose diagonal elements were equal to 1 and the off-diagonal elements were equal to 0.3. We then repeated $R = 10,000$ iterations of the following process:

(i) Given the explanatory variables, we generated the survey variable Y according to Model (19).

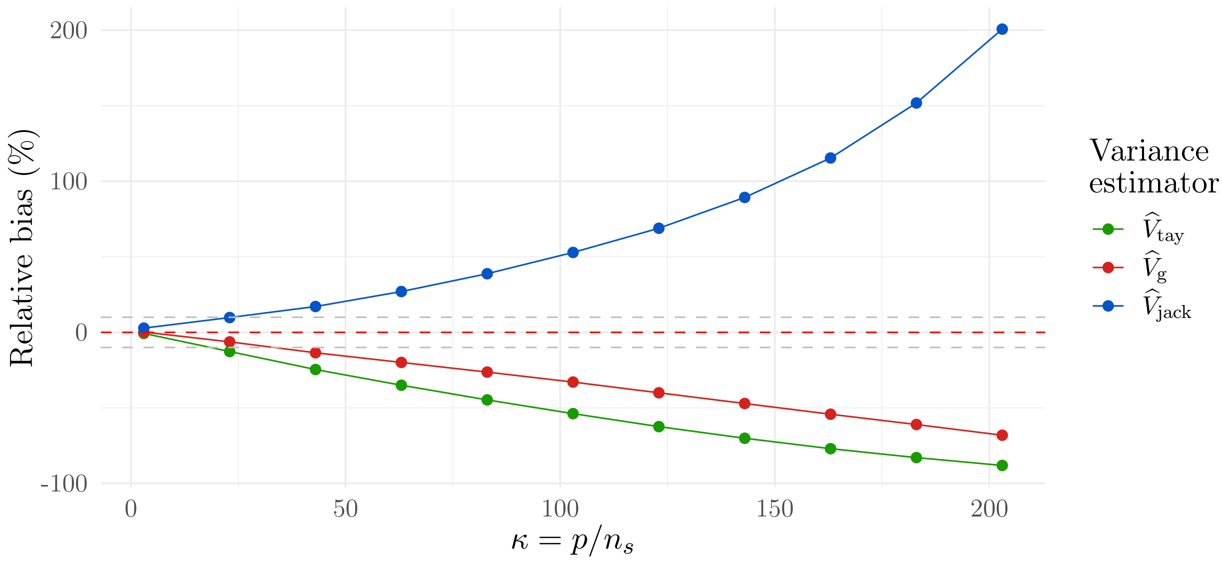


Figure 2: Behaviour of three variance estimators for $\widehat{\mu}_{greg}$ under simple random sampling without replacement.

(ii) From the finite population of size $N = 5,000$ generated in Step (i), a sample, of (expected) size $n = 300$, was selected according to (1) simple random sampling without replacement and (2) Bernoulli sampling.

(iii) In each sample, the response indicators $\{R_k\}_{k \in S}$, were independently generated according to a Bernoulli distribution with probability

$$p_k = \{1 + \exp(1 + \lambda_1 x_{1k} + \lambda_2 x_{2k} + \lambda_3 x_{3k})\}^{-1},$$

where the values of λ_1, λ_2, and λ_3 were set to obtain an overall response rate of about 50%. Thus, in each sample, the expected number of respondents, $\mathbb{E}(n_r)$, was equal to 150.

(iv) The missing values in each sample were imputed through deterministic linear regression imputation with different subsets of explanatory variables. The first subset of explanatory variables included the variables X_1, X_2, and X_3 only, corresponding to the true model. In addition to X_1, X_2 and X_3, we included several noise variables denoted by p_{noise}. This led to 12 sets of explanatory variables of size p, where $p = p_{noise} + 3$. The values for p_{noise} were set to: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, and 110. As a result the ratio $\kappa = p/\mathbb{E}(n_r)$ ranged from $3/150$ to $103/150$. Each of the 12 models was fitted on the set of responding units, which led to 12 sets of imputed values.

(v) For each of the 12 sets of imputed values, we computed the imputed estimator $\widehat{\mu}_{lr}$ given by (11), leading to a set of 12 imputed estimators.

(vi) We estimated the variance of the 12 imputed estimators using two variance estimators: (i) The variance estimator based on a first-order Taylor expansion denoted by $\widehat{V}_{I,\text{tay}}$; see Section 2.2; and (ii) The generalized jackknife variance estimator, denoted by $\widehat{V}_{I,\text{jack}}$; see Section 2.2.

As a measure of relative bias of a variance estimator $\widehat{V}$, we computed its Monte Carlo percent relative bias (RB) given by (20).

From Figures 3 and 4, we note that both $\widehat{V}_{I,\text{tay}}$ and $\widehat{V}_{I,\text{jack}}$ performed well for small values of κ. For instance, for $\kappa = 3/150$, which corresponds to the case of $p_{noise} = 0$, the estimator $\widehat{V}_{I,\text{tay}}$ exhibited a value of RB of about -5.7% for Bernoulli sampling and -4.6% simple random sampling without replacement. The jackknife variance estimator performed well with values of RB equal to -4.1% for Bernoulli sampling and 3% for simple random sampling without replacement. However, for larger values of κ both variance estimators did not perform well. For instance, for $\kappa = 33/150 \approx 0.29$, the estimator $\widehat{V}_{I,\text{tay}}$ underestimated the true variance with values of RB equal to -14.3% for Bernoulli sampling and -12.8% for simple random sampling without replacement. On the other hand, the estimator $\widehat{V}_{I,\text{jack}}$ was 10.4% too large for Bernoulli sampling and 12% too large for simple random sampling without replacement. Again, the magnitude of the bias increased significantly as κ increased.

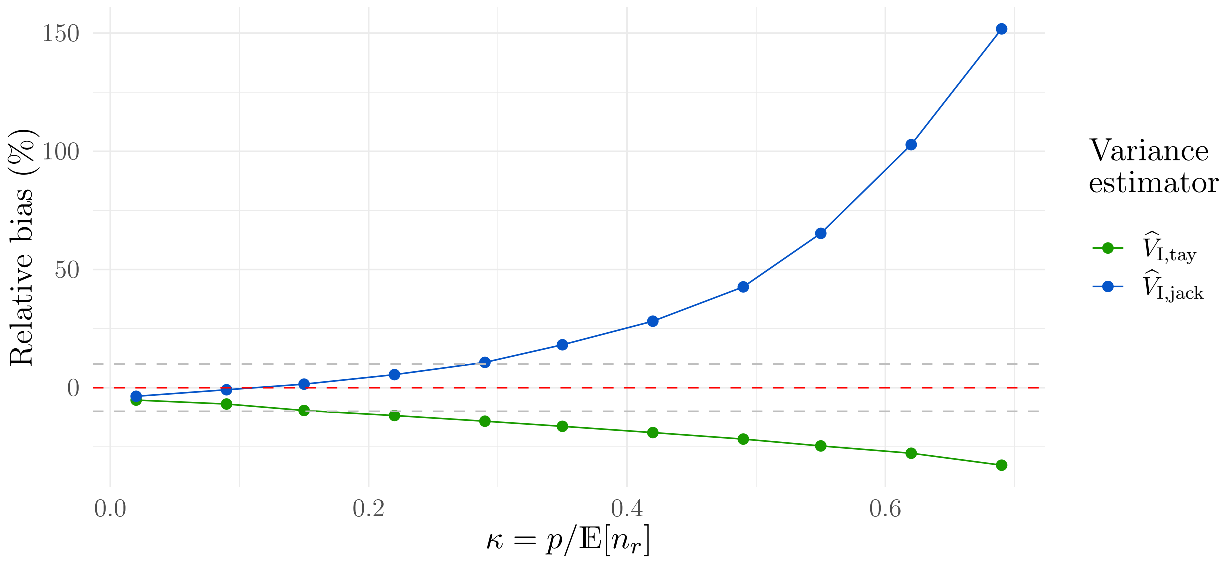


Figure 3: Behaviour of two variance estimators for $\widehat{\mu}_{lr}$ under Bernoulli sampling.

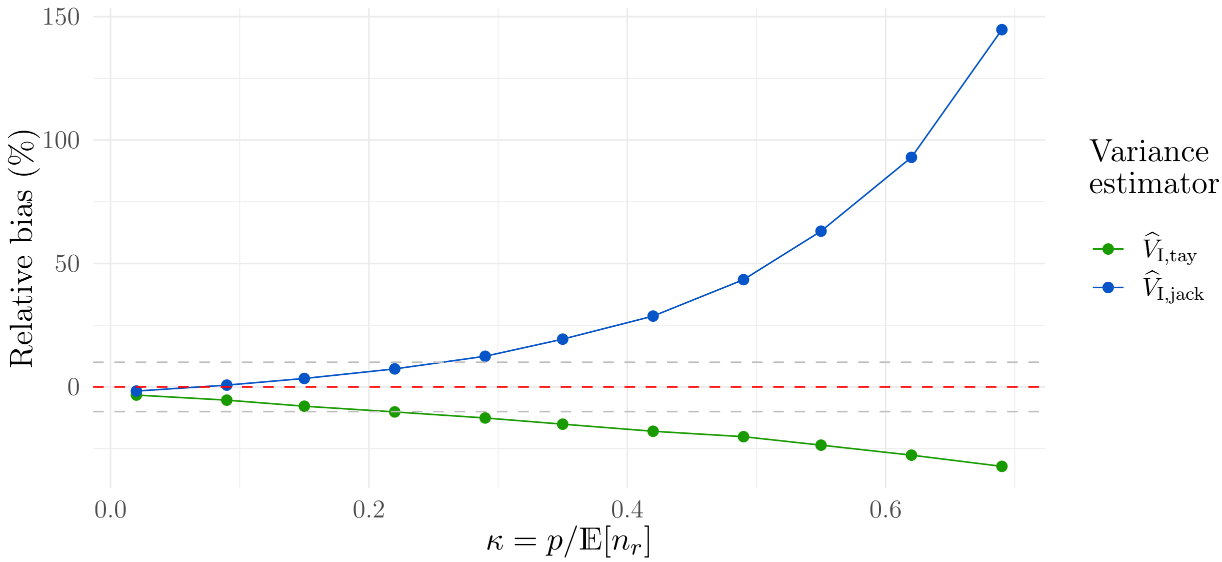


Figure 4: Behaviour of two variance estimators for $\widehat{\mu}_{lr}$ under simple random sampling without replacement.

3.3 Explaining the behavior of classical variance estimators

In the context of both model-assisted estimation and deterministic linear regression imputation, the customary linearization variance estimators and jackknife variance estimators tend to breakdown when $p/n \to \kappa \in (0, 1)$ (or $p/\mathbb{E}[n_r] \to \kappa \in (0, 1)$, respectively). In this section, we explain why this is the case. For simplicity, we confine to the case of model-assisted estimation under simple random sampling without replacement. Arguments similar to the ones below can also be used to explain the behavior of variance estimators under deterministic linear regression imputation.

The variance estimator based on a first-order Taylor expansion given by (5) involves the sample residuals $\widehat{\epsilon}_{kS}$. It turns out that, in a high-dimensional setting, the distribution of the sample residuals $\widehat{\epsilon}_{kS}$ is not a good approximation of the distribution of the errors ϵ_k in (1). In particular, we have

$$\mathbb{V}_m(\widehat{\epsilon}_{kS}) = \sigma^2(1-\widetilde{h}_{kk}), \qquad k \in S,$$

where $\widetilde{h}_{kk}$ denotes the k-th diagonal element of the hat matrix $\boldsymbol{X}_S\left(\boldsymbol{X}_S^\top \boldsymbol{X}_S\right)^{-1}\boldsymbol{X}_S^\top$. The validity of classical variance estimators relies on the assumption that $\widetilde{h}_{kk} \to 0$ as n and N go to infinity. In a high-dimensional setting, this assumption no longer holds, as it can be shown that $\widetilde{h}_{kk} \to p/n$ for a wide class of distributions for the design matrix $\boldsymbol{X}_S$ (e.g., the multivariate normal distribution); see e.g., El Karoui and Purdom (2018), Pajor and Pastur (2009) and Karoui and Koesters (2011) for a discussion and Portnoy (1987) for a proof with elliptical distributions; see Section 4 for a more rigorous treatment, and Appendix A.4 for empirical results on the behavior of the leverages for commonly encountered distributions. Therefore, the variance of the sample residuals $\widehat{\epsilon}_{kS}$ is approximately equal to $\sigma^2(1-p/n) \equiv \sigma^2(1-\kappa)$, which can be considerably smaller than σ^2 for large values of κ. This, in turn, explains why the variance estimator based on a first-order Taylor expansion tends to underestimate the true variance of $\widehat{\mu}_{greg}$ for large values of κ.

Turning to generalized jackknife variance estimators, we note from (10) that it involves the residuals $\widehat{\epsilon}_{kS}^{(k)} = \widehat{\epsilon}_{kS}/(1-\widetilde{h}_{kk})$. Since

$$\mathbb{V}_m\left(\widehat{\epsilon}_{kS}^{(k)}\right) = \frac{\sigma^2}{1-\widetilde{h}_{kk}} \simeq \frac{\sigma^2}{1-\frac{p}{n}} \equiv \frac{\sigma^2}{1-\kappa}, \qquad k \in S,$$

the variance of $\widehat{\epsilon}_{kS}^{(k)}$ may be considerably larger than σ^2 for large values of κ. As a result, the generalized jackknife variance estimator tends to overestimate the true variance of $\widehat{\mu}_{greg}$ for large values of κ.

4 Bias: Model-assisted estimation

In this section, we provide a theoretical analysis of the bias of variance estimators for the GREG estimator in a high-dimensional setting. For simplicity, we confine to the case of Bernoulli sampling.

We consider the asymptotic framework of Isaki and Fuller (1982). We consider an increasing sequence of finite populations $\{U_v\}_{v\in\mathbb{N}}$, of sizes $\{N_v\}_{v\in\mathbb{N}}$, such that $U_v \subset U_{v+1}$, for all $v \in \mathbb{N}$. From U_v, a sample S_v is selected according to the sampling design $\mathcal{P}_v\left(\,\cdot\,,\ \mathbf{Z}_v\right)$. The first and second-order inclusion probabilities of $\mathcal{P}_v\left(\,\cdot\,,\ \mathbf{Z}_v\right)$ are denoted by $\{\pi_{k,v}\}_{k\in U_v}$ and $\{\pi_{k\ell,v}\}_{k\neq\ell\in U_v}$, respectively. The asymptotic sampling fraction is given by $\lim_{v\to\infty} n_v/N_v := \pi_*$. For ease of notation, the subscript v will be omitted whenever possible. For two sequences $\{a_v\}_{v\in\mathbb{N}} \subset \mathbb{R}$ and $\{b_v\}_{v\in\mathbb{N}} \subset \mathbb{R}$, we write $a_v \simeq b_v$ to express that they have the same limit, i.e., $\lim_{v\to\infty} a_v/b_v = 1$. We write $a_v \lesssim b_v$ if $\lim_{v\to\infty} a_v \leqslant \lim_{v\to\infty} b_v$; the symbol $\gtrsim$ is defined similarly. We extend these definitions to sequences of random variables where the limit is to be understood in probability. Moreover, asymptotic order notations are to be understood in a high-dimensional framework, whereby $\lim_{v\to\infty} p_v/n_v := \kappa_* \in [0;1)$.

In this article, inference is made conditionally on the predictors. Below, we will refer to the following conditions:

(H1W) The design matrix $\boldsymbol{X}_S$ is such that, there exists a continuous function $f_1 : [0;1) \to [0;1)$ such that, for all $v \in \mathbb{N}$,

$$\min_{k\in S_v} \widetilde{h}_{kk} \geqslant f_1\left(\frac{p_v}{n_v}\right),$$

where f satisfies $f_1(0) = 0$ and $f_1(x) > 0$ for $x > 0$.

(H1S) The design matrix $\boldsymbol{X}_S$ is such that,

$$\max_{k\in S_v}\left|\widetilde{h}_{kk}-\frac{p_v}{n_v}\right|=o_{\mathbb{P}}\left(1\right).$$

Assumptions (H1W) and (H1S) pertain to the high-dimensional behavior of the leverages that generally depend on the distribution of the covariates. Assumption (H1W) is weaker than assumption (H1S); it states that the smallest leverage remains bounded away from zero whenever $\kappa_*>0$; because the intercept is included in the covariates, Assumption (H1W) is necessarily satisfied in finite samples since $\widetilde{h}_{kk}>1/n$. On the other hand, assumption (H1S) assumes that, uniformly, leverages converge to κ. Assumption (H1S) implies (H1W) with f_1 being the identity. We refer to Appendix A.4 for an empirical investigation of (H1W) and (H1S) for some common distributions.

Before stating our next result, we introduce the following notation. For an arbitrary set of elements $\{a_k\}_{k\in B}$ with index set $B\subseteq U$, we write $\mathbb{E}_{n,B}\left[a\right]:=|B|^{-1}\sum_{k\in B}a_k$ to denote its empirical mean.

Result 4.1. *Consider a Bernoulli sampling design.*

i) If the superpopulation model is linear, the model variance $\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)$ is unbiased for the unconditional variance $\mathbb{V}\left(\widehat{\mu}_{greg,v}\right)$, that is,

$$\mathbb{E}_p\left[\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)\right]=\mathbb{V}\left(\widehat{\mu}_{greg,v}\right).$$

Moreover, if $\lim_{v\to\infty}\mathbb{V}_p\left(g_{k,v}\right)=0$, for all k, then $\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)$ and $\mathbb{V}\left(\widehat{\mu}_{greg,v}\right)$ are asymptotically equivalent, that is,

$$\frac{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)}{\mathbb{V}\left(\widehat{\mu}_{greg,v}\right)}\xrightarrow[v\to\infty]{\mathbb{P}}1.$$

ii) The relative bias factor of $\widehat{V}_{\mathrm{tay,v}}$, $\widehat{V}_{\mathrm{g,v}}$ and $\widehat{V}_{\mathrm{jack,v}}$, defined respectively by (5), (9) and (10), are given by

$$\frac{\mathbb{E}_m\left[\widehat{V}_{\mathrm{tay,v}}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)}\simeq\left(1-\pi^*\right)\frac{\left(1-\kappa_*\right)}{\mathbb{E}_{n,U}\left[g\right]}+\frac{\pi^*}{\mathbb{E}_{n,U}\left[g\right]},\tag{21}$$

$$\frac{\mathbb{E}_m\left[\widehat{V}_{\mathrm{g,v}}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)} \simeq (1-\pi^*)\left(1-\frac{\mathbb{E}_{n,S}\left[g^2\widetilde{h}\right]}{\mathbb{E}_{n,U}\left[g\right]}\right)+\frac{\pi^*}{\mathbb{E}_{n,U}\left[g\right]}, \tag{22}$$

and

$$\frac{\mathbb{E}_m\left[\widehat{V}_{\mathrm{jack,v}}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)} \simeq (1-\pi^*)\frac{\mathbb{E}_{n,S}\left[g^2/\left(1-\widetilde{h}\right)\right]}{\mathbb{E}_{n,U}\left[g\right]}+\frac{\pi^*}{\mathbb{E}_{n,U}\left[g\right]}. \tag{23}$$

Proof. See Appendix B.3. ■

Part i) of Result 4.1 shows that the model variance of the GREG estimator is unbiased and consistent for the unconditional variance, provided that the linear regression model holds. For p fixed, the GREG estimator attains the Godambe-Joshi lower bound (Särndal et al., 1992, Chapter 12). Part i) also holds for general sampling designs, under appropriate assumptions on higher-order inclusion probabilities. Part ii) of Result 4.1 confirms the results of Section 3; that is, the variance estimators based on a first-order Taylor expansion lead to substantial underestimation for large values of κ, whereas the generalized jackknife variance estimators leads to substantial over-estimation of the true variance.

Remark 4.1. *The behavior of the g-weights $\{g_k\}_{k\in U}$ significantly impacts the high-dimensional behavior of the variance estimators. Indeed, if p_v is either fixed (or slowly increases with respect to n_v), it can be shown that, uniformly in k,*

$$g_k \xrightarrow[v\to\infty]{\mathbb{P}} 1.$$

We conjecture that this result may not hold for general covariates settings within a high-dimensional framework for which $\kappa > 0$. Assumption in Result 4.1 that $\lim_{v\to\infty}\mathbb{V}_p\left(g_{k,v}\right)=0$ is much weaker as it only requires that this limit to be degenerate.

Corollary 4.1. *Consider a Bernoulli sampling design with $\pi_* = 0$.*

i) Under (H1W), Expressions (21)-(23) reduce to

$$\frac{\mathbb{E}_m\left[\widehat{V}_{\text{tay,v}}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)} \simeq \frac{(1-\kappa_*)}{\mathbb{E}_{n,U}\left[g\right]}, \tag{24}$$

$$\frac{\mathbb{E}_m\left[\widehat{V}_{g,v}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)} \simeq 1 - \frac{\mathbb{E}_{n,S}\left[g^2\widetilde{h}\right]}{\mathbb{E}_{n,U}\left[g\right]} \lesssim 1 - f_1\left(\kappa_*\right),$$

$$\frac{\mathbb{E}_m\left[\widehat{V}_{jack,v}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)} \simeq \frac{\mathbb{E}_{n,S}\left[g^2/\left(1-\widetilde{h}\right)\right]}{\mathbb{E}_{n,U}\left[g\right]} \gtrsim \frac{1}{1-f_1\left(\kappa_*\right)}.$$

ii) Under (H1S), Expressions (22)-(23) reduce to

$$\frac{\mathbb{E}_m\left[\widehat{V}_{\text{g,v}}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)} \simeq 1-\kappa_*, \tag{25}$$

$$\frac{\mathbb{E}_m\left[\widehat{V}_{\text{jack,v}}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)} \simeq \frac{1}{1-\kappa_*}. \tag{26}$$

Proof. See Appendix B.4. ∎

When $\kappa_* > 0$, under (H1W), the variance estimators $\widehat{V}_g$ and $\widehat{V}_{jack}$ are biased and inconsistent. Under the stronger assumption (H1S), their expression of asymptotic bias reduces to a simpler expression that depends on κ_* only. Interestingly, under (H1S) and $\pi_* = 0$, the bias of $\widehat{V}_{\text{jack}}$ matches the bias found in El Karoui and Purdom (2018) in the context of jackknife variance estimation of a prediction for a linear regression model.

We end this section by suggesting simple bias-adjusted variance estimators in the context of a high-dimensional setting. For a small sampling fraction, Expressions (24)-(26) motivates the following bias-adjusted version of $\widehat{V}_{\text{tay}}$, $\widehat{V}_g$ and $\widehat{V}_{\text{jack}}$. They are respectively given by

$$\widehat{V}^{(adj)}_{\text{tay,v}} = \frac{\mathbb{E}_{n,U}\left[g\right]}{1-\kappa} \times \widehat{V}_{\text{tay,v}}, \tag{27}$$

$$\widehat{V}_{\mathrm{g,v}}^{(adj)} = \frac{1}{1-\kappa}\widehat{V}_{\mathrm{g,v}}, \tag{28}$$

and

$$\widehat{V}_{\mathrm{jack,v}}^{(adj)} = (1-\kappa)\,\widehat{V}_{\mathrm{jack,v}}. \tag{29}$$

Under (H1S), the bias-adjusted estimators (24)-(26) are asymptotically unbiased for all values of κ, and can be implemented using the observed data.

5 Bias: Deterministic linear regression imputation

In this section, we study the behavior of several variance estimators in the context of deterministic linear regression imputation in a high-dimensional setting. To that aim, consider the following class of variance estimators:

$$\mathcal{V} := \Bigg\{ \widehat{V}^{(\psi)} := \frac{1}{\widehat{N}^2}\sum_{k\in S}\sum_{\ell\in S}\frac{\Delta_{k\ell}}{\pi_{k\ell}}\frac{\widehat{\mu}_{lr}-\widehat{\xi}_k^{(\psi)}}{\pi_k}\frac{\widehat{\mu}_{lr}-\widehat{\xi}_\ell^{(\psi)}}{\pi_\ell} + \frac{\sigma^2}{\widehat{N}^2}\sum_{k\in S_r}\frac{1}{\pi_k}\Big\{1-R_k(1+\widehat{\Gamma}_k)\Big\}^2;$$

$$\text{with} \quad \widehat{\xi}_\ell^{(\psi)} := \widetilde{y}_k + r_k\psi(\boldsymbol{X}_R)\widehat{\Gamma}_k\,\widehat{\epsilon}_{kR}, \quad \text{for some} \quad \psi:\mathbb{R}^{n_r\times p}\to\mathbb{R}\Bigg\}. \tag{30}$$

The class of variance estimators $\mathcal{V}$ includes, as special cases, the variance estimators (14) and (17) with $\psi(\boldsymbol{X}_R)=1$ and $\psi(\boldsymbol{X}_R)=\left(1-\widehat{h}_{kk}\right)^{-1}$, respectively. It also includes the variance estimator obtained by adjusting the residuals $\widehat{\epsilon}_{kR}$ using the rescaling factor $(1-\widehat{h}_{kk})^{-1/2}$; see e.g., El Karoui and Purdom (2018). This variance estimator, which we call the corrected variance estimator, corresponds to the choice $\psi(\boldsymbol{X}_R)=(1-\widehat{h}_{kk})^{-1/2}$. The rationale behind this choice is to recover the variance of the model errors ϵ_k. Finally, we denote by $\widetilde{\mathcal{V}}$ the subset of $\mathcal{V}$ where the functions ψ are constants, independent of $\boldsymbol{X}_R$.

The following result exhibits the asymptotic bias of any variance estimator belonging to the class $\mathcal{V}$.

Result 5.1. *Consider a Bernoulli sampling design and let* $\widehat{V}^{(\psi)}$ *be an arbitrary variance* *estimator belonging to the class $\mathcal{V}$ with functions $\{\psi_v\}_{v\in\mathbb{N}}$. Then,*

$$\frac{\mathbb{E}_m\left[\widehat{V}_v^{(\psi)}\right]}{\mathbb{E}_m\left[\widehat{V}(\widehat{\mu}_{lr,v})\right]} \backsimeq \frac{[1-\pi]\sum_{k\in S} B_k + \sigma^2 A_\psi}{[1-\pi]\sum_{k\in S} B_k + \sigma^2 A_{theo}},$$

where

$$A_\psi := (1-\pi)\left(n_r + \sum_{k\in S_m} \widehat{h}_{kk} + \left(\sum_{k\in S_r}\left(1-\widehat{h}_{kk}\right)\psi(\boldsymbol{X}_R)\Gamma_k\left\{2+\psi(\boldsymbol{X}_R)\Gamma_k\right\}\right)\right) + \pi\left(n_m + \sum_{k\in S_r}\Gamma_k^2\right),$$

$$A_{theo} := (1-\pi)(n_s - 1) + n_m + \sum_{k\in S_r}\Gamma_k^2,$$

with

$$B_k := \frac{1}{n_s}\sum_{\ell\in S}\left\{(\mathbf{x}_\ell^\top\boldsymbol{\beta})^2 - \mathbf{x}_k^\top\boldsymbol{\beta}\mathbf{x}_\ell^\top\boldsymbol{\beta}\right\}. \tag{31}$$

Proof. See Appendix B.5. ■

Although Result 5.1 does not require any assumption on the covariates, it is somewhat difficult to draw a clear interpretation from it, aside from the fact that the asymptotic bias of the jackknife variance estimator is always non-negative. Unfortunately, under (H1W), Result 5.1 does not become any easier to interpret. In order to provide more insight about Result 5.1, we reformulate Assumption (H1S).

(H1S') The design matrix $\boldsymbol{X}_R$ is such that,

$$\max_{k\in S_{r,v}}\left|\widehat{h}_{kk} - \frac{p_v}{n_{r,v}}\right| = o_{\mathbb{P}}(1).$$

In the case of a negligible sampling fraction, Corollary 5.1 below exhibits the relative biases of the variance estimators (14) and (17) as well as the corrected variance estimator obtained with $\psi(\boldsymbol{X}_R) = (1-\widehat{h}_{kk})^{-1/2}$ and denoted by $\widehat{V}_{\text{cor}}$.

Corollary 5.1. *Consider a Bernoulli sampling design with* $\pi_* = 0$ *and assume (H1S').*

$$\frac{\mathbb{E}_m\left[\widehat{V}_{I,\text{tay,v}}\right]}{\mathbb{E}_m\left[\widehat{V}(\widehat{\mu}_{lr,v})\right]} - 1 \backsimeq \frac{\sum_{k\in S_m}\widehat{h}_{kk} - \kappa_*\left(2n_m + \sum_{k\in S_r}\Gamma_k^2\right)}{\sum_{k\in S}B_k/\sigma^2 + \left(n_r + 2n_m + \sum_{k\in S_r}\Gamma_k^2\right)}, \tag{32}$$

$$\frac{\mathbb{E}_m\left[\widehat{V}_{I,\text{jack,v}}\right]}{\mathbb{E}_m\left[\widehat{V}(\widehat{\mu}_{lr,v})\right]} - 1 \backsimeq \frac{\sum_{k\in S_m}\widehat{h}_{kk} + \frac{\kappa_*}{1-\kappa_*}\sum_{k\in S_r}\Gamma_k^2}{\sum_{k\in S}B_k/\sigma^2 + \left(n_r + 2n_m + \sum_{k\in S_r}\Gamma_k^2\right)}, \tag{33}$$

and

$$\frac{\mathbb{E}_m\left[\widehat{V}_{I,\text{cor,v}}\right]}{\mathbb{E}_m\left[\widehat{V}(\widehat{\mu}_{lr,v})\right]} - 1 \backsimeq \frac{\sum_{k\in S_m}\widehat{h}_{kk} - 2n_m\left(1 - \sqrt{1-\kappa_*}\right)}{\sum_{k\in S}B_k/\sigma^2 + \left(n_r + 2n_m + \sum_{k\in S_r}\Gamma_k^2\right)}. \tag{34}$$

Note that the terms on the right-hand side of (32)-(34) have the same denominator. Since $\sum_{k\in S}B_k \geq 0$, it follows that the sign of the bias in (32)-(34) depends on the sign of the numerator. In the case of $\widehat{V}_{I,\text{tay}}$, the sign of the numerator is expected to be negative, and the bias may be substantial for large values of κ. It follows from (33) that the bias of $\widehat{V}_{I,\text{jack}}$ is positive. Noting that $\kappa/(1-\kappa)$ is monotonically increasing in $\kappa \in (0,1)$, the bias of $\widehat{V}_{I,\text{jack}}$ is expected to be large for large values of κ. Finally, looking at Expression (34), we expect the bias of $\widehat{V}_{I,\text{cor}}$ to be small to moderate as the term $1-\sqrt{1-\kappa}$ lies between 0 and 1.

Expressions (32)-(34) may be used to derive bias-adjusted versions of $\widehat{V}_{I,\text{tay}}$, and $\widehat{V}_{I,\text{jack}}$. Note that all the terms but B_k and σ^2 on the right side of (32)-(34) can be computed using the data at hand. An estimator of B_k is obtained by replacing $\boldsymbol{\beta}$ in (31) by its weighted least square estimator given by (12). Also, a model-unbiased estimator of σ^2 is given by (15). In Section 6.2, the bias-adjusted versions of $\widehat{V}_{I,\text{tay}}$ and $\widehat{V}_{I,\text{jack}}$ will be denoted by $\widehat{V}_{I,\text{jack}}^{(adj)}$ and $\widehat{V}_{I,\text{tay}}^{(adj)}$, respectively.

Remark 5.1. *The numerators on the right side of (32)-(34) all include the term* $\sum_{k\in S_m}\widehat{h}_{kk}$, *which is a function of the predictors for both the respondents and the nonrespondents. Its* *magnitude heavily depends on the type of nonresponse mechanism, which is unknown. Indeed, when p is fixed, it can be shown that*

$$\sum_{k\in S_m} \widehat{h}_{kk} \simeq \lim_{v\to\infty} \frac{n_{m,v}}{n_{r,v}} \times tr\left(\mathbb{E}\left[\mathbf{x}_1\mathbf{x}_1^\top | R_1 = 1\right]^{-1} \mathbb{E}\left[\mathbf{x}_1\mathbf{x}_1^\top | R_1 = 0\right]\right).$$

As a result, unless the data are Missing Completely At Random, it is generally not possible to establish a general result about the magnitude of $\sum_{k\in S_m} \widehat{h}_{kk}$.

We end this section by determining the coefficients $\psi(\boldsymbol{X}_R)$ in (30) that produce an asymptotically unbiased variance estimator. This is given in the next result.

Result 5.2. *Recall that $\widetilde{\mathcal{V}}$ denotes the subset of $\mathcal{V}$ where the functions ψ are constants and assume a Bernoulli sampling design. Consider the following notation additional notation:*

$$A := \sum_{k\in S_r} \left(1 - \widehat{h}_{kk}\right) \widehat{\Gamma}_k^2,$$

$$B := 2\sum_{k\in S_m} \left(1 - \widehat{h}_{kk}\right) \widehat{\Gamma}_k,$$

$$C := \sum_{k\in S_m} \widehat{h}_{kk} - 2n_m - \sum_{k\in S_r} \widehat{\Gamma}_k^2.$$

Then, there exist asymptotically unbiased variance estimators in $\widetilde{\mathcal{V}}$ for all values of κ if and only if $\Delta := B^2 - 4AC \geqslant 0$. Assuming this condition is satisfied, the asymptotically unbiased estimators are identified by the following roots:

$$\psi_1(\boldsymbol{X}_R) := \frac{-B - \sqrt{B^2 - 4AC}}{2A}, \tag{35}$$

and

$$\psi_2(\boldsymbol{X}_R) := \frac{-B + \sqrt{B^2 - 4AC}}{2A}. \tag{36}$$

Proof. See Appendix B.6. ■

Note that the roots $\psi_1(\boldsymbol{X}_R)$ and $\psi_2(\boldsymbol{X}_R)$ can be readily computed using the data at hand. Result 5.2 does not require any assumption on the leverage. If one is willing to assume (H1S), then A and B reduce to

$$A = (1 - \kappa_*) \sum_{k \in S_r} \widehat{\Gamma}_k^2, \qquad B = 2\,(1 - \kappa_*)\, n_m,$$

respectively.

Remark 5.2. *If the discriminant $\Delta := B^2 - 4AC$ is negative, no estimator in $\widetilde{V}$, computable from observed data, is unbiased for this particular realization. One may then choose the midpoint $\psi := -B/2A$ which, when $\Delta < 0$, minimizes the absolute relative bias.*

6 Empirical behavior of bias-adjusted estimators

In this section, we present the results from a simulation study assessing the performance of several variance estimators and their associated bias-adjusted version in terms of relative bias. In Section 6.1, we consider the model-assisted estimation setup, while Section 6.2 discusses the linear regression imputation setup.

6.1 Bias-adjusted variance estimators: Model-assisted estimation

We used the same simulation setup as the one described in Section 3.1. To estimate the variance of $\widehat{\mu}_{greg}$, we computed several variance estimators in each sample: (i) $\widehat{V}_{\text{tay}}$ given by (5); (ii) $\widehat{V}_g$ given by (7); (iii) $\widehat{V}_{\text{jack}}$ given by (10); (iv) The bias-adjusted variance estimators $\widehat{V}_{\text{tay}}^{(adj)}$, $\widehat{V}_g^{(adj)}$ and $\widehat{V}_{\text{jack}}^{(adj)}$, given respectively by (27)-(29). In addition, we computed the variance $\mathbb{V}_m := \mathbb{V}_m(\widehat{\mu}_{greg})$, which corresponds to the model variance $\widehat{\mu}_{greg}$. For each variance estimator, we computed its Monte Carlo percent relative bias given by (20). The results for Bernoulli sampling are shown in Table 1, whereas Table 2 shows the results for simple random sampling without replacement.

From Table 1, we first note that observe that the target $\mathbb{V}_m$ exhibited a small bias for all values of κ, which is consistent with Part i) of Result 4.1. The results pertaining to $\widehat{V}_{\text{tay}}$, $\widehat{V}_g$ and $\widehat{V}_{\text{jack}}$ are virtually identical to those presented in Section 3.1. The bias-adjusted versions

					Relative bias (in %)			
p	κ	$\mathbb{V}_m$	$\widehat{V}_{\text{tay}}$	$\widehat{V}_g$	$\widehat{V}_{\text{jack}}$	$\widehat{V}_{\text{tay}}^{(adj)}$	$\widehat{V}_g^{(adj)}$	$\widehat{V}_{\text{jack}}^{(adj)}$
3	0.01	0.8	-1.7	-0.6	2.0	0.2	0.4	1.1
23	0.08	0.5	-14.1	-7.7	8.4	-0.1	0.1	0.9
43	0.14	0.7	-25.3	-14.2	16.5	0.0	0.2	1.2
63	0.21	0.2	-36.2	-21.3	25.2	-0.5	-0.3	1.0
83	0.28	0.3	-46.0	-28.0	36.5	-0.5	-0.2	1.3
103	0.34	0.3	-55.0	-34.7	50.1	-0.6	-0.3	1.5
123	0.41	0.6	-63.2	-41.2	67.7	-0.3	0.0	2.2
143	0.48	0.0	-70.9	-48.4	88.4	-1.0	-0.6	1.9
163	0.54	-0.7	-77.7	-55.5	115.2	-1.8	-1.6	1.4
183	0.61	-0.9	-83.5	-62.4	154.6	-2.2	-1.9	1.8
203	0.68	-1.5	-88.5	-69.4	212.6	-3.0	-2.7	1.9

Table 1: Monte Carlo percent relative bias of several variance estimators for Bernoulli sampling: model-assisted estimation.

					Relative bias (in %)			
p	κ	$\mathbb{V}_m$	$\widehat{V}_{\text{tay}}$	$\widehat{V}_g$	$\widehat{V}_{\text{jack}}$	$\widehat{V}_{\text{tay}}^{(adj)}$	$\widehat{V}_g^{(adj)}$	$\widehat{V}_{\text{jack}}^{(adj)}$
3	0.01	1.2	-0.7	0.2	2.8	1.2	1.2	1.9
23	0.08	1.5	-12.7	-6.3	9.8	1.5	1.5	2.3
43	0.14	1.0	-24.6	-13.5	17.1	1.0	0.9	1.9
63	0.21	1.6	-35.0	-19.9	27.0	1.5	1.4	2.6
83	0.28	2.0	-44.7	-26.3	38.8	1.9	1.8	3.3
103	0.34	2.4	-53.8	-32.9	52.9	2.2	2.1	3.9
123	0.41	1.9	-62.4	-40.0	68.9	1.7	1.6	3.7
143	0.48	1.4	-70.1	-47.1	89.3	1.2	1.0	3.5
163	0.54	0.7	-77.0	-54.2	115.4	0.4	0.2	3.2
183	0.61	0.7	-82.9	-61.0	151.8	0.2	-0.1	3.4
203	0.68	-0.7	-88.1	-68.1	200.7	-1.2	-1.6	2.7

Table 2: Monte Carlo percent relative bias of several variance estimators for simple random sampling without replacement: model-assisted estimation.

$\widehat{V}_{\text{tay}}^{(adj)}$, $\widehat{V}_g^{(adj)}$ and $\widehat{V}_{\text{jack}}^{(adj)}$ performed very well in terms of bias for all values of κ with an absolute RB less than 3%. Similar results (see Table 2) were obtained for simple random sampling without replacement, which can be explained by the fact that, in terms of variance, the strategy consisting of Bernoulli sampling and the GREG estimator is asymptotically equivalent to the strategy consisting of simple random sampling without replacement and the GREG estimator.

6.2 Bias-adjusted variance estimators: Linear regression imputation

We used the same simulation setup as the one described in Section 3.2. To estimate the variance of $\widehat{\mu}_{lr}$, we computed several variance estimators in each sample: (i) $\widehat{V}_{I,\mathrm{tay}}$ given by (5); (ii) $\widehat{V}_{I,\mathrm{jack}}$ given by (10); (iii) $\widehat{V}_{\mathrm{cor}}$ obtained from (30) with $\psi(\boldsymbol{X}_R) = (1 - \widehat{h}_{kk})^{-1/2}$; (iv) The bias-adjusted variance estimators $\widehat{V}_{I,\mathrm{tay}}^{(adj)}$, $\widehat{V}_{I,\mathrm{jack}}^{(adj)}$; (v) The variance estimators obtained from (30) with $\psi_1(\boldsymbol{X}_R)$ and $\psi_2(\boldsymbol{X}_R)$ given by (35) and (36), respectively, and denoted by $\widehat{V}_{\psi_1}^{(adj)}$ and $\widehat{V}_{\psi_2}^{(adj)}$. In addition, we computed the (unfeasible) target given by (42). Finally, we computed the (unfeasible) bias-adjusted variance estimators $\widehat{V}_{I,\mathrm{tay_T}}^{(adj)}$ and $\widehat{V}_{I,\mathrm{jk_T}}^{(adj)}$ that are identical to $\widehat{V}_{I,\mathrm{tay}}^{(adj)}$ and $\widehat{V}_{I,\mathrm{jack}}^{(adj)}$, respectively, except that they use the true value of B_k (see Equation (31)) and the true value of σ^2. For each variance estimator, we computed its Monte Carlo percent relative bias given by (20). The results for Bernoulli sampling are shown in Table 3, whereas Table 4 shows the results for simple random sampling without replacement.

From Table 3, we first note that the target $\mathbb{V}_{target}$ exhibited a small bias for all values of κ, as expected. The results pertaining to $\widehat{V}_{I,\mathrm{tay}}$ and $\widehat{V}_{I,\mathrm{jack}}$ were virtually identical to those presented in Section 3.2. The variance estimator $\widehat{V}_{\mathrm{cor}}$ performed significantly better than $\widehat{V}_{I,\mathrm{tay}}$ and $\widehat{V}_{I,\mathrm{jack}}$, particularly for small to moderate values of κ, with an RB of -2% for $\kappa = 0.42$, compared to -19.1% and 27.6% for $\widehat{V}_{I,\mathrm{tay}}$ and $\widehat{V}_{I,\mathrm{jack}}$, respectively. However, the performance of $\widehat{V}_{\mathrm{cor}}$ slightly deteriorated for larger values of κ. For $\kappa = 0.69$, the value of RB was approximately equal to 17%. Turning to the bias-adjusted variance estimators $\widehat{V}_{I,\mathrm{tay}}^{(adj)}$ and $\widehat{V}_{I,\mathrm{jack}}^{(adj)}$, they performed well for small to moderate values of κ. For instance, for $\kappa = 0.42$, the variance estimators $\widehat{V}_{I,\mathrm{tay}}^{(adj)}$ and $\widehat{V}_{I,\mathrm{jack}}^{(adj)}$ exhibited a value of RB of about -7.6% and -0.9%, respectively. However, for a large value of κ of 0.69, we note a deterioration for both $\widehat{V}_{I,\mathrm{tay}}^{(adj)}$ and $\widehat{V}_{I,\mathrm{jack}}^{(adj)}$ with a value of RB of about -16.2% and 11%, respectively. This deterioration in terms of bias seems to correspond to the price to pay for estimating B_k in (31) in a high-dimensional setting. Indeed, when comparing the relative bias of $\widehat{V}_{I,\text{tay}}^{(adj)}$ and $\widehat{V}_{I,\text{jack}}^{(adj)}$ with their unfeasible counterparts $\widehat{V}_{I,\text{tay}_\text{T}}^{(adj)}$ and $\widehat{V}_{I,\text{jk}_\text{T}}^{(adj)}$, we observe that the latter performed very well for all values of κ with values of absolute RB remaining below 10% in all scenarios. Further research is needed to develop estimators of B_k that are more robust to the dimensionality of the vector $\mathbf{x}$. Finally, the variance estimators $\widehat{V}_{\psi_1}^{(adj)}$ and $\widehat{V}_{\psi_2}^{(adj)}$ performed very well in terms of bias for all values of κ. This is consistent with Result 5.2. Again, the results for simple random sampling without replacement were similar to those obtained for Bernoulli sampling; see Table 4.

				Relative bias (in %)						
κ	$\mathbb{V}_{\text{target}}$	$\widehat{V}_{I,\text{tay}}$	$\widehat{V}_{I,\text{jack}}$	$\widehat{V}_{cor}$	$\widehat{V}_{I,\text{tay}}^{(adj)}$	$\widehat{V}_{I,\text{jack}}^{(adj)}$	$\widehat{V}_{\psi_1}^{(adj)}$	$\widehat{V}_{\psi_2}^{(adj)}$	$\widehat{V}_{I,\text{tay}_\text{T}}^{(adj)}$	$\widehat{V}_{I,\text{jk}_\text{T}}^{(adj)}$
0.02	-3.1	-5.7	-4.1	-4.9	-5.3	-4.8	-5.3	-5.3	-5.3	-4.8
0.15	-2.8	-9.6	1.5	-4.5	-5.4	-4.2	-5.5	-5.2	-5.2	-4.5
0.29	-3.0	-14.3	10.4	-4.0	-6.5	-3.5	-6.1	-5.8	-5.7	-4.5
0.42	-2.4	-19.1	27.6	-2.0	-7.6	-0.9	-6.2	-5.8	-5.7	-3.8
0.55	-4.3	-26.6	59.3	-1.2	-11.8	1.5	-8.7	-8.4	-8.2	-5.0
0.69	-4.6	-34.5	162.1	3.7	-16.2	11.0	-10.0	-10.0	-9.4	-2.8

Table 3: Monte Carlo percent relative bias of several variance estimators for Bernoulli sampling: linear regression imputation.

				Relative bias (in %)						
κ	$\mathbb{V}_{target}$	$\widehat{V}_{I,\text{tay}}$	$\widehat{V}_{I,\text{jack}}$	$\widehat{V}_{cor}$	$\widehat{V}_{I,\text{tay}}^{(adj)}$	$\widehat{V}_{I,\text{jack}}^{(adj)}$	$\widehat{V}_{\psi_1}^{(adj)}$	$\widehat{V}_{\psi_2}^{(adj)}$	$\widehat{V}_{I,\text{tay}_\text{T}}^{(adj)}$	$\widehat{V}_{I,\text{jk}_\text{T}}^{(adj)}$
0.02	-2.1	-4.6	-3.0	-3.8	-4.1	-3.7	-4.3	-4.1	-4.1	-3.7
0.15	-1.5	-8.2	3.0	-3.0	-4.0	-2.8	-4.3	-3.8	-3.8	-3.0
0.29	-1.5	-12.8	12.0	-2.4	-4.9	-1.9	-4.7	-4.2	-4.1	-3.0
0.42	-1.1	-17.7	29.0	-0.5	-6.0	0.6	-4.8	-4.3	-4.2	-2.3
0.55	-1.1	-23.6	62.6	2.4	-8.4	5.0	-5.3	-5.0	-4.7	-1.5
0.69	0.4	-30.0	151.8	9.2	-10.8	15.7	-4.6	-4.5	-3.9	2.1

Table 4: Monte Carlo percent relative bias of several variance estimators for simple random sampling without replacement: linear regression imputation.

7 Final remarks

In this article, we studied the behavior of linearization and jackknife variance estimators when p/n is not negligible. For the GREG estimator, we showed that the customary linearization variance estimator and its g-weighted version are biased negatively, whereas the jackknife is biased positively. These biases did not vanish, even asymptotically, unless $\lim_{v\to\infty} p_v/n_v = 0$. For model-assisted variance estimators, we obtained closed-form expressions for the bias of both Taylor, g-weights, and the jackknife variance estimators. These expressions can be used to define bias-adjusted variance estimators, which are unbiased regardless of the value of p/n. Results from a simulation study support these findings. We have also examined the behavior of imputed estimators under linear regression imputation. In this context, the bias of linearization and jackknife variance estimators depends on unknown quantities. Developing bias-adjusted variance estimators requires an intermediate estimation step. This is beyond the scope of this article, and is a topic currently under investigation.

Our theoretical investigations were limited to simple random sampling without replacement and Bernoulli sampling. While we empirically evaluated the performance of several variance estimators for Poisson sampling with unequal probabilities, a theoretical analysis of the properties of linearization and jackknife variance estimators for unequal probability sampling designs remains to be explored.

In this paper, we did not examine bootstrap variance estimators in high-dimensional settings. Unreported results suggest that customary finite population bootstrap procedures may exhibit substantial positive bias in high dimensions, consistent with recent findings by El Karoui and Purdom (2018); Zhao and Candes (2022) in an i.i.d. setup. This is a topic of future research.

Recently, Stefan and Hidiroglou (2024) proposed a jackknife version of the GREG estimator along with a jackknife variance estimator for its mean squared error in a low-dimensional setting. The extension of these methods to the high-dimensional setting will be addressed elsewhere.

References

Berger, Y. G. and Rao, J. (2006). Adjusted jackknife for imputation under unequal probability sampling without replacement. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):531–547.

Berger, Y. G. and Skinner, C. J. (2005). A jackknife variance estimator for unequal probability sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):79–89.

Breidt, F. J. and Opsomer, J. D. (2017). Model-assisted survey estimation with modern prediction techniques. *Statistical Science*, 32(2):190–205.

Campbell, C. (1980). A different view of finite population estimation. In Washington, D., editor, *Proc. Survey Res. Meth. Sect. Am. Statist. Assoc.*, pages 319–324.

Cardot, H., Goga, C., and Shehzad, M.-A. (2017). Calibration and partial calibration on principal components when the number of auxiliary variables is large. *Statistica Sinica*, 27(243-260).

Chauvet, G. and Goga, C. (2022). Asymptotic efficiency of the calibration estimator in a high-dimensional data setting. *Journal of Statistical Planning and Inference*, 217:177–187.

Chen, S. and Haziza, D. (2019). Recent developments in dealing with item non-response in surveys: a critical review. *International Statistical Review*, 87:S192–S218.

Dagdoug, M., Goga, C., and Haziza, D. (2022). Model-assisted estimation in high-dimensional settings for survey data. *Journal of Applied Statistics*. https://doi.org/10.1080/02664763.2022.2047905.

Duchesne, P. (2000). A note on jackknife variance estimation for the general regression estimator. *Journal of Official Statistics*, 16(2):133.

El Karoui, N. and Purdom, E. (2018). Can we trust the bootstrap in high-dimensions? the case of linear models. *The Journal of Machine Learning Research*, 19(1):170–235.

Fay, R. (1991). *A design-based perspective on missing data variance.* US Census Bureau.

Fuller, W. A. (2009). *Sampling statistics.* John Wiley & Sons.

Haziza, D. (2009). Imputation and inference in the presence of missing data. In Pfeffermann, D. and Rao, C., editors, *Handbook of statistics*, volume 29A, pages 215–246. Elsevier.

Haziza, D. and Vallée, A.-A. (2020). Variance estimation procedures in the presence of singly imputed survey data: a critical review. *Japanese Journal of Statistics and Data Science*, 3(2):583–623.

Isaki, C.-T. and Fuller, W.-A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.*, 77:49–61.

Karoui, N. E. and Koesters, H. (2011). Geometric sensitivity of random matrix results: consequences for shrinkage estimators of covariance and related statistical methods. *arXiv: Statistics Theory.*

Kim, J. K. and Rao, J. (2009). A unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika*, 96(4):917–932.

Mashreghi, Z., Haziza, D., and Léger, C. (2016). A survey of bootstrap methods in finite population sampling. *Statistics Surveys*, 10(none).

Pajor, A. and Pastur, L. (2009). On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution. *Studia Mathematica*, 1(195):11–29.

Pfeffermann, D. and Sverchkov, M. (2009). Inference under informative sampling. In *Handbook of statistics*, volume 29, pages 455–487. Elsevier.

Portnoy, S. (1987). A central limit theorem applicable to robust regression estimators. *Journal of multivariate analysis*, 22(1):24–50.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Särndal, C. E. (1980). On pi-inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika*, 67(3):639–650.

Särndal, C.-E. (1992). Methods for estimating the precision of survey estimates when imputation has been used. *Survey Methodology*, 18:241–252.

Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, 33:99–119.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model assisted survey sampling*. Springer Series in Statistics. Springer-Verlag, New York.

Särndal, C.-E., Swensson, B., and Wretman, J. H. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika*, 76(3):527–537.

Shao, J. and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association*, 94(445):254–265.

Stefan, M. and Hidiroglou, M. A. (2024). Jackknife bias-corrected generalized regression estimator in survey sampling. *Journal of Survey Statistics and Methodology*, 12(1):211–231.

Ta, T., Shao, J., Li, Q., and Wang, L. (2020). Generalized regression estimators with high-dimensional covariates. *Statistica Sinica*.

Valliant, R. (2002). Variance estimation for the general regression estimator. *Survey methodology*, 28(1):103–108.

Wang, Z., Peng, L., and Kim, J. K. (2022). Bootstrap inference for the finite population mean under complex sampling designs. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4):1150–1174.

Wolter, K. (2007). *Introduction to Variance Estimation*. New York: Springer.

Yung, W. and Rao, J. (1996). Jackknife linearization variance estimators under stratified multi-stage sampling. *Survey Methodology*, 22:23–32.

Zhao, Q. and Candes, E. J. (2022). An adaptively resized parametric bootstrap for inference in high-dimensional generalized linear models. *arXiv preprint arXiv:2208.08944*.

Appendices

A Additional simulation studies

A.1 Proportional-to-size Poisson sampling

To evaluate the performance of linearization and jackknife variance estimators, along with their bias-adjusted versions, under unequal probability sampling designs, we conducted an additional simulation study using Poisson sampling with first-order inclusion probabilities proportional to size. The data-generating processes used in Section 3 remain unchanged. More specifically, the first-order inclusion probability for unit k was defined as

$$\pi_k := \widetilde{n} \frac{x_{k2}^2}{\sum_{\ell \in U} x_{\ell 2}^2}, \qquad k \in U,$$

where $\widetilde{n}$ denote the expected sample size. The correlation between the inclusion probabilities and the survey variable was about 0.4. The results for model-assisted estimation and linear regression imputation are presented in Tables 5 and 6, respectively.

			Relative bias (in %)					
p	κ	$\mathbb{V}_m$	$\widehat{V}_{\text{tay}}$	$\widehat{V}_g$	$\widehat{V}_{\text{jack}}$	$\widehat{V}_{\text{tay}}^{(adj)}$	$\widehat{V}_g^{(adj)}$	$\widehat{V}_{\text{jack}}^{(adj)}$
3	0.01	1.7	-0.8	-0.2	6.2	1.3	0.9	5.3
23	0.08	2.4	-12.7	-6.4	22.2	3.2	1.6	13.8
43	0.14	3.1	-23.9	-12.6	41.0	5.1	2.3	22.8
63	0.21	3.0	-35.0	-19.7	61.9	6.4	2.0	30.7
83	0.28	2.4	-45.5	-27.0	85.7	7.2	1.5	38.1
103	0.34	2.2	-54.8	-34.1	114.5	8.7	1.2	45.3
123	0.41	3.3	-62.8	-40.3	152.4	11.6	2.2	54.1
143	0.48	2.2	-70.7	-47.9	193.6	12.4	1.0	59.0
163	0.54	2.5	-77.4	-54.7	251.4	14.8	1.1	65.8
183	0.61	2.2	-83.3	-61.8	325.8	16.5	0.6	70.4
203	0.68	0.9	-88.5	-69.0	432.8	17.1	-0.8	73.9

Table 5: Monte Carlo percent relative bias of several variance estimators for Poisson sampling: model-assisted estimation.

In the model-assisted estimation setting, the behavior of $\widehat{V}_{\text{tay}}$, $\widehat{V}_{\text{g}}$ and $\widehat{V}_{\text{jack}}$ remained similar to that observed in simple random sampling without replacement and Bernoulli sampling. Although the adjusted jackknife variance estimator showed substantially lower RB than its

Relative bias (in %)										
κ	$\mathbb{V}_{target}$	$\widehat{V}_{I,\text{tay}}$	$\widehat{V}_{I,\text{jack}}$	$\widehat{V}_{cor}$	$\widehat{V}^{(adj)}_{I,\text{tay}}$	$\widehat{V}^{(adj)}_{I,\text{jack}}$	$\widehat{V}^{(adj)}_{\psi_1}$	$\widehat{V}^{(adj)}_{\psi_2}$	$\widehat{V}^{(adj)}_{I,\text{tay}_\text{T}}$	$\widehat{V}^{(adj)}_{I,\text{jk}_\text{T}}$
0.02	-0.3	-2.3	0.1	-1.1	-1.6	-0.6	-11.8	-1.8	-1.6	-0.6
0.15	-0.3	-6.0	9.7	0.7	-1.0	2.9	-7.8	-1.7	-0.8	2.6
0.29	0.6	-10.1	27.3	3.6	-0.5	9.3	-6.6	-1.4	0.4	8.0
0.42	2.0	-15.6	58.4	6.8	-1.1	17.7	-7.0	-1.9	1.4	13.8
0.55	5.7	-21.6	129.9	12.7	-1.1	32.7	-6.2	-1.7	4.0	23.4
0.69	12.1	-30.4	4331.7	21.7	-2.5	257.3	-4.8	-1.8	10.0	114.2

Table 6: Monte Carlo percent relative bias of several variance estimators for Poisson sampling: linear regression imputation.

unadjusted counterpart, $\widehat{V}_{\text{jack}}$, it still exhibited significant RB for moderate to large values of κ. The adjusted estimator $\widehat{V}^{(adj)}_{\text{tay}}$ performed well for small to moderate values of κ but the bias slightly deteriorated as κ increased. For instance, for $\kappa = 0.68$, its RB was equal to 17.1%. Finally, the adjusted estimator $\widehat{V}^{(adj)}_{\text{g}}$ performed very well for all values κ with an absolute value of RB less than 2.5% in all the scenarios.

In the linear regression imputation setting, the behaviour of $\widehat{V}_{I,\text{tay}}$ and $\widehat{V}_{I,\text{jack}}$ was similar to that observed under simple random sampling without replacement and Bernoulli sampling. Again, the adjusted jackknife variance estimator $\widehat{V}^{(adj)}_{I,\text{jack}}$ outperformed its unadjusted counterpart $\widehat{V}_{I,\text{jack}}$ but but still exhibited significant positive bias for large values of κ. The estimator $\widehat{V}_{cor}$ performed well for small to moderate values of κ, whereas it exhibited a moderate positive RB for $\kappa = 0.55$ and $\kappa = 0.69$. Turning to $\widehat{V}^{(adj)}_{\psi_1}$ and $\widehat{V}^{(adj)}_{\psi_2}$, they performed surprisingly very well despite being developed under Bernoulli sampling.

A.2 Non Gaussian distributions

We conducted a simulation study assessing the performance of the variance estimators when both the explanatory variables and the model errors were generated using non Gaussian distributions. We present the results for simple random sampling without replacement, as the findings for Bernoulli sampling were similar and are therefore omitted. We replicated the simulation processes described in Sections 6.1 and 6.2, with the following modifications: i)The explanatory variables were independently generated from a gamma distribution with shape parameter equal to 8 and rate parameter equal to 1; ii) The errors in (19), were generated from a uniform distribution $\mathcal{U}[-15; 15]$. The results for model-assisted estimation and linear regression imputation are presented in Tables 7 and 8, respectively.

The results in Table 7 suggest that the behavior of all variance estimators was similar to that observed Section 6.1 for simple random sampling without replacement. Indeed, the magnitude of the RB associated to $\widehat{V}_{\text{tay}}$, $\widehat{V}_g$ and $\widehat{V}_{\text{jack}}$ got worse as κ increased. The bias-adjusted versions $\widehat{V}_{\text{tay}}^{(adj)}$, $\widehat{V}_g^{(adj)}$ and $\widehat{V}_{\text{jack}}^{(adj)}$ still performed very well with an absolute RB less than 4% for all values of κ.

From Table 8, the estimators $\widehat{V}_{I,\text{tay}}$ and $\widehat{V}_{I,\text{jack}}$ exhibited the same behavior as observed in Section 6.2 with their RB increasing significantly in magnitude as κ increased. In contrast, $\widehat{V}_{I,\text{cor}}$ performed very well, with an absolute RB less than 3% for all values of κ. Turning to the bias-adjusted estimators $\widehat{V}_{I,\text{tay}}^{(adj)}$ and $\widehat{V}_{I,\text{jack}}^{(adj)}$, both performed well, with values of absolute RB below 6% in all scenarios. Finally, the variance estimators $\widehat{V}_{\psi_1}^{(adj)}$, $\widehat{V}_{\psi_2}^{(adj)}$ continued to perform well in terms of RB across all values of κ.

		Relative bias (in %)						
p	κ	$\mathbb{V}_m$	$\widehat{V}_{\text{tay}}$	$\widehat{V}_g$	$\widehat{V}_{\text{jack}}$	$\widehat{V}_{\text{tay}}^{(adj)}$	$\widehat{V}_g^{(adj)}$	$\widehat{V}_{\text{jack}}^{(adj)}$
3	0.01	0.5	-1.4	-0.5	2.1	0.5	0.5	1.2
23	0.08	0.5	-13.6	-7.2	8.8	0.5	0.4	1.4
43	0.14	1.0	-24.6	-13.5	17.3	1.0	0.9	2.1
63	0.21	0.5	-35.6	-20.7	25.9	0.5	0.3	1.7
83	0.28	0.2	-45.6	-27.6	36.6	0.2	0.0	1.7
103	0.34	0.1	-54.7	-34.4	49.8	0.1	-0.2	1.8
123	0.41	0.6	-62.8	-40.8	67.0	0.5	0.2	2.5
143	0.48	-0.2	-70.6	-47.9	86.5	-0.3	-0.6	2.0
163	0.54	0.6	-77.1	-54.3	115.1	0.3	0.0	3.1
183	0.61	0.3	-83.0	-61.1	151.1	-0.1	-0.5	3.2
203	0.68	0.3	-88.0	-67.8	204.1	-0.1	-0.6	3.9

Table 7: Monte Carlo percent relative bias of several variance estimators for simple random sampling without replacement: model-assisted estimation.

Relative bias (in %)										
κ	$\mathbb{V}_{target}$	$\widehat{V}_{I,\text{tay}}$	$\widehat{V}_{I,\text{jack}}$	$\widehat{V}_{cor}$	$\widehat{V}_{I,\text{tay}}^{(adj)}$	$\widehat{V}_{I,jack}^{(adj)}$	$\widehat{V}_{\psi_1}^{(adj)}$	$\widehat{V}_{\psi_2}^{(adj)}$	$\widehat{V}_{I,\text{tay}_\text{T}}^{(adj)}$	$\widehat{V}_{I,\text{jk}_\text{T}}^{(adj)}$
0.02	-1.8	-3.6	-2.2	-2.9	-3.3	-2.7	-3.8	-3.3	-3.3	-2.7
0.15	-1.9	-9.8	6.8	-2.8	-4.1	-2.4	-4.2	-3.9	-3.9	-2.8
0.29	-1.3	-18.2	38.6	-0.4	-5.6	1.2	-4.4	-4.3	-4.1	-1.5
0.42	-1.8	-3.6	-2.2	-2.9	-3.3	-2.7	-3.8	-3.3	-3.3	-2.7
0.55	-1.9	-9.8	6.8	-2.8	-4.1	-2.4	-4.2	-3.9	-3.9	-2.8
0.69	-1.3	-18.2	38.6	-0.4	-5.6	1.2	-4.4	-4.3	-4.1	-1.5

Table 8: Monte Carlo percent relative bias of several variance estimators for simple random sampling without replacement: linear regression imputation.

A.3 Asymptotic experiments with p/n fixed

In this simulation, we assess the relative biases of the variance estimators discussed in the article, keeping κ fixed while allowing n_v and N_v to increase. We considered two asymptotic regimes: $\kappa_* = 0.1$ and $\kappa_* = 0.5$. We considered three population sizes $\{N_v\}$ given by $N_1 = 1000, N_2 = 5000, N_3 = 1000$. The sample sizes $\{n_v\}$ were given by $n_v = \lfloor N_v^{58/100} \rfloor$. This corresponds to a scenario where $\pi^* = 0$. The samples was selected according to simple random sampling without replacement. The results for model-assisted estimation are presented in Table 9 and Table 10, respectively.

p	n	$\mathbb{V}_m$	$\widehat{V}_{\text{tay}}$	$\widehat{V}_{\text{g}}$	$\widehat{V}_{\text{jack}}$	$\widehat{V}_{\text{tay}}^{(adj)}$	$\widehat{V}_{\text{g}}^{(adj)}$	$\widehat{V}_{\text{jack}}^{(adj)}$
$\kappa = 0.1$								
6	54	-0.63	-20.94	-12.12	15.91	-1.98	-2.30	5.45
14	139	-2.06	-20.59	-12.17	10.23	-2.33	-2.41	-0.27
21	208	1.37	-17.82	-8.93	13.76	1.24	1.19	2.79
$\kappa = 0.5$								
28	54	0.08	-76.19	-53.70	125.00	-4.88	-7.15	17.33
70	139	0.97	-74.62	-50.70	107.72	-0.52	-1.39	6.08
105	208	0.21	-75.04	-51.00	104.64	-1.23	-1.97	3.95

Table 9: Evolution of the Monte Carlo percent relative bias of several variance estimators for simple random sampling: model-assisted estimation.

We observe from Table 9 that the RB of $\widehat{V}_{\text{tay}}$, $\widehat{V}_{\text{g}}$ converged to strictly negative quantities. Similarly, the RB of $\widehat{V}_{\text{jack}}$ converged to strictly positive quantities. These results support our theoretical results, confirming that $\widehat{V}_{\text{tay}}$, $\widehat{V}_{\text{g}}$, and $\widehat{V}_{\text{jack}}$ remain asymptotically biased whenever

$\kappa > 0$. On the other hand, the bias-adjusted variance estimators exhibited only negligible bias asymptotically. We obtained similar results in the linear regression imputation setting; see Table 10.

p	$\mathbb{E}[n_r]$	$\mathbb{V}_{target}$	$\widehat{V}_{I,\text{tay}}$	$\widehat{V}_{I,\text{jack}}$	$\widehat{V}_{cor}$	$\widehat{V}^{(adj)}_{I,\text{tay}}$	$\widehat{V}^{(adj)}_{I,\text{jack}}$	$\widehat{V}^{(adj)}_{\psi_1}$	$\widehat{V}^{(adj)}_{\psi_2}$	$\widehat{V}^{(adj)}_{I,\text{tay}_\text{T}}$	$\widehat{V}^{(adj)}_{I,\text{jk}_\text{T}}$
$\kappa = 0.1$											
3	28	-0.39	-7.20	5.14	-1.78	-4.54	0.01	-5.78	-4.50	-4.44	-0.17
7	70	-0.54	-5.08	2.92	-1.35	-2.43	-0.85	-3.00	-2.37	-2.34	-0.97
11	104	-0.88	-5.11	2.73	-1.43	-2.24	-1.12	-2.57	-2.16	-2.14	-1.23
$\kappa = 0.5$											
14	28	-0.09	-26.90	666.28	4.97	-13.11	105.32	-11.89	-10.73	-9.06	88.48
35	70	0.38	-19.44	58.88	4.55	-5.50	9.31	-3.63	-2.97	-2.50	3.94
53	104	-0.02	-19.44	55.25	4.26	-5.10	7.27	-2.77	-2.38	-2.05	2.00

Table 10: Evolution of the Monte Carlo percent relative bias of several variance estimators for simple random sampling: linear regression imputation.

A.4 Empirical validity of (H1W) and (H1S)

In this section, we present the results from a simulation study, assessing the validity of Assumption (H1W) and (H1S) for some common continuous distributions of the covariates $\mathbb{P}_{\mathbf{x}}$: the normal distribution, the uniform distribution, and the exponential distribution. Under Bernoulli sampling, if $\{\mathbf{x}_k\}_{k\in U}$ are i.i.d. $\mathbb{P}_{\mathbf{x}}$, then, given S, $\{\mathbf{x}_k\}_{k\in S}$ are also i.i.d. $\mathbb{P}_{\mathbf{x}}$; see Theorem 1.3.1. in Fuller (2009). Therefore, we performed a Monte-Carlo simulation so that, at each of the $R = 200$ iterations, we selected a sample of size n_v of p_v covariates according to Bernoulli sampling, and computed $\max_{k\in[n_v]}|h_{kk,v} - p_v/n_v|$ and $\min_{k\in[n_v]} h_{kk,v}$. Here, $[n] := \{1, 2, \ldots, n\}$. In Figure 5, we plotted the Monte-Carlo distribution of these statistics for different values of v, ranging from $v = 1$ to $v = 10$. The sample sizes $\{n_v\}$ were set to $n_v := v \times 1000$ and $\{p_v\}$ was set such that $p_v = \lfloor \kappa n_v \rfloor$, with $\kappa := 0.05$. Here, the value of κ is small but not negligible, chosen specifically to reduce the computational complexity of the simulations. We also conducted simulations for larger values of κ. The results are very similar to those presented in Figure 5.

For each covariate distribution, both (H1S) and (H1W) appear to hold. The plots on the left suggest that the mean of the distributions of the maximum discrepancy converges to zero, with its variance also decreasing to zero. From the plots on the right, we observe that the minimum leverage remains bounded away from zero, as expected. A theoretical investigation establishing the validity of both (H1S) and (H1W) is beyond the scope of this article but is currently being pursued.

B Proofs

As mentioned in the main article, we assume below, without loss of generality, that σ^2 is known.

B.1 Proof of Proposition 2.1.

The generalized jackknife variance estimator of $\widehat{\mu}_{greg}$ is defined in Berger and Skinner (2005) as

$$\widehat{V}_{\text{jack}}\left(\widehat{\mu}_{greg}\right) := \sum_{k \in S} \sum_{\ell \in S} \frac{\Delta_{k\ell}}{\pi_{k\ell}} z_k z_\ell, \tag{37}$$

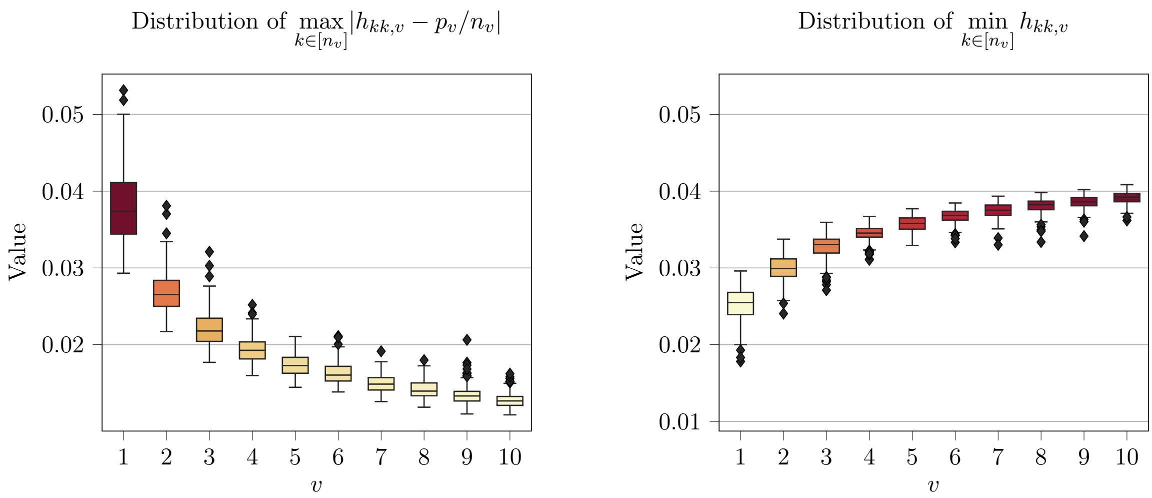
where

$$z_k := (1 - w_k)\left(\widehat{\mu}_{greg} - \widehat{\mu}_{greg}^{(k)}\right). \tag{38}$$

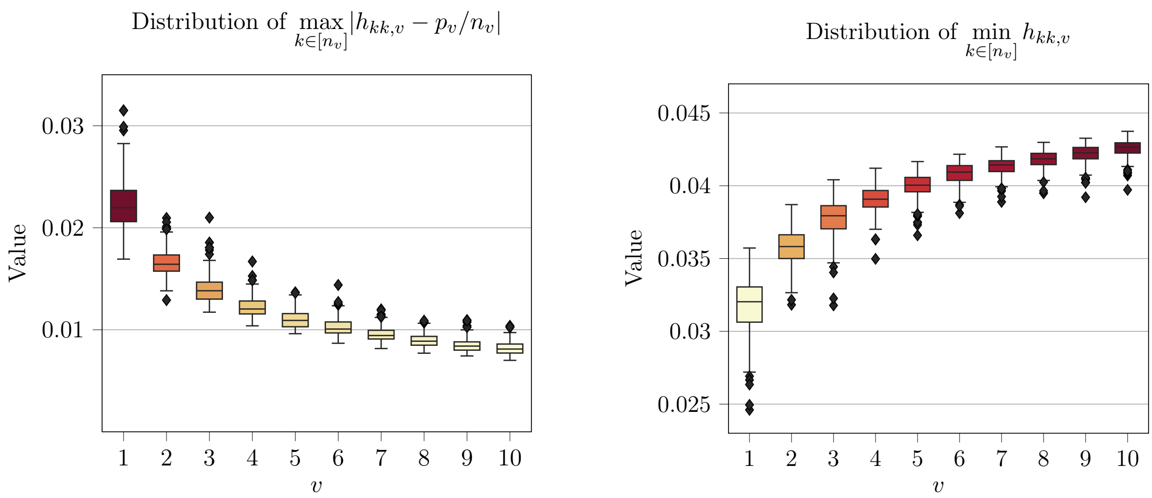
In (38), $\widehat{\mu}_{greg}^{(k)}$ denotes the GREG estimator as defined in (2), computed by deleting the element k of the sample (but not of the population). More generally, in what follows, we use the subscript (k) to denote any statistic (or, set) computed after deleting element k.

To derive a closed-form formula for (37), it suffices to find a closed-form formula for $\widehat{\mu}_{greg}^{(k)}$. In Duchesne (2000), a formula is given when the inclusion probabilities are reweighted after the deletion of an element, a practice that we do not do with the generalized jackknife variance estimator. Below, we give a simple proof relying on Lemma 1. We let $\widehat{t}_{greg} := N\widehat{\mu}_{greg}$ be the usual total GREG estimator. We begin by noting that if the full-sample GREG estimator can be written in projection form, then the leave-one-out version of the estimator can, too. It follows that

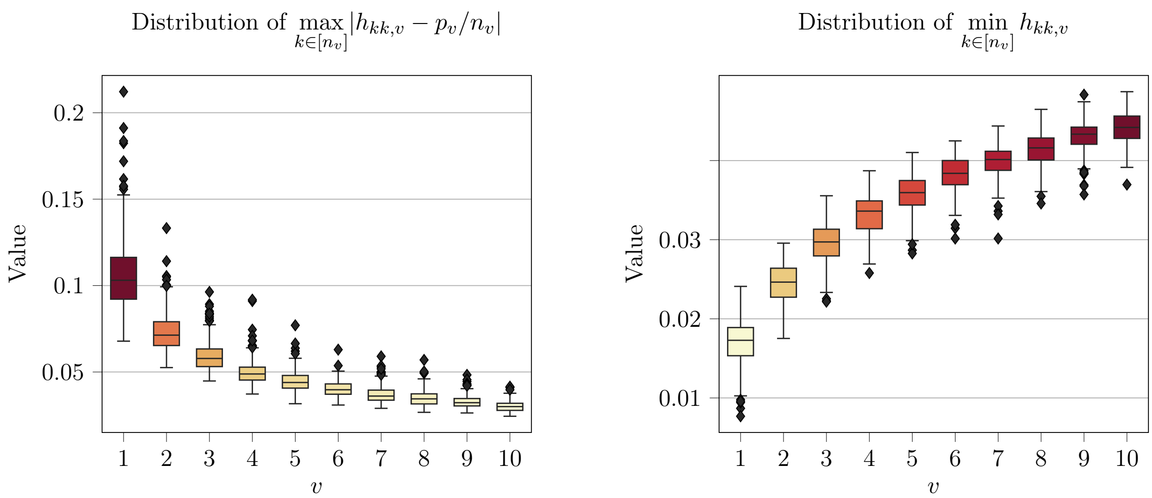
$$\widehat{t}_{greg}^{(k)} = \boldsymbol{t}_x^\top \widehat{\boldsymbol{\beta}}^{(k)}$$



Gaussian covariates: $\mathbb{P}_{\mathbf{x}} = \mathcal{N}(0,1)^{\otimes p}$



Uniform covariates: $\mathbb{P}_{\mathbf{x}} = \mathcal{U}[0,1]^{\otimes p}$



Exponential covariates: $\mathbb{P}_{\mathbf{x}} = \mathcal{E}(1)^{\otimes p}$

Figure 5: Asymptotic evolution of the distributions of statistics involved in (H1S) and (H1W). Colors represent the value of the mean of the distribution: warmer colors indicate higher values.

$$= \boldsymbol{t}_x^\top \left(\widehat{\boldsymbol{\beta}} - \frac{\boldsymbol{A}_{\Pi S}^{-1} d_k \mathbf{x}_k \widehat{\epsilon}_{kS}}{1 - \widetilde{h}_{kk}^\pi} \right)$$

$$= \widehat{t}_{greg} - \frac{d_k g_k \widehat{\epsilon}_{ks}}{1 - \widetilde{h}_{kk}^\pi}.$$

Now,

$$\widehat{\mu}_{greg}^{(k)} := \frac{\widehat{t}_{greg}^{(k)}}{N} = \widehat{\mu}_{greg} - \frac{d_k g_k \widehat{\epsilon}_{ks}}{N\left(1 - \widetilde{h}_{kk}^\pi\right)},$$

from which it follows that

$$\widehat{\mu}_{greg}^{(k)} - \widehat{\mu}_{greg} = -\frac{1}{N}\frac{d_k g_k \widehat{\epsilon}_{ks}}{1 - \widetilde{h}_{kk}^\pi}.$$

Finally,

$$z_k = \frac{1}{N}\frac{(1 - w_k)\, g_k \widehat{\epsilon}_{ks}}{\pi_k \left(1 - \widetilde{h}_{kk}^\pi\right)},$$

which concludes the proof.

B.2 Proof of Result 2.1.

The Generalized Jackknife variance estimator proposed in Berger and Rao (2006) is

$$\widehat{V}_{\text{I,jack}}\left(\widehat{\mu}_{lr}\right) := \sum_{k\in S}\sum_{\ell\in S} \frac{\Delta_{k\ell}}{\pi_{k\ell}} e_k e_\ell, \tag{39}$$

where

$$e_k := (1 - w_k)\left(\widehat{\mu}_{lr} - \widehat{\mu}_{lr}^{(k)}\right), \qquad \text{and} \qquad w_k := \frac{d_k}{\widehat{N}}.$$

The effect of deleting a respondent or nonrespondent is different; we treat these cases separately. For $k \in S_m$, write

$$\widehat{\mu}_{lr}^{(k)} := \frac{1}{\widehat{N} - d_k}\left(\sum_{\ell\in S_r} \frac{y_\ell}{\pi_\ell} + \sum_{j\in S_m^{(k)}} \frac{\mathbf{x}_\ell^\top \widehat{\boldsymbol{\beta}}_R}{\pi_\ell} \right)$$

$$= \frac{1}{\widehat{N} - d_k}\left(\sum_{\ell\in S_r} \frac{y_\ell}{\pi_\ell} + \sum_{\ell\in S_m} \frac{\mathbf{x}_\ell^\top \widehat{\boldsymbol{\beta}}_R}{\pi_\ell} - \frac{\mathbf{x}_k^\top \widehat{\boldsymbol{\beta}}_R}{\pi_k} \right)$$

$$= \frac{\widehat{N}}{\widehat{N} - d_k} \left(\widehat{\mu}_{lr} - \frac{\mathbf{x}_k^\top \widehat{\boldsymbol{\beta}}_R}{\widehat{N}\pi_k} \right).$$

Similarly, using Lemma 1, for $k \in S_r$, we have

$$\widehat{\mu}_{lr}^{(k)} := \frac{1}{\widehat{N} - d_k} \left(\sum_{\ell \in S_r^{(k)}} \frac{y_\ell}{\pi_\ell} + \sum_{\ell \in S_m} \frac{\mathbf{x}_\ell^\top \widehat{\boldsymbol{\beta}}_R^{(k)}}{\pi_\ell} \right)$$

$$= \frac{1}{\widehat{N} - d_k} \left\{ \sum_{\ell \in S_r} \frac{y_\ell}{\pi_\ell} - \frac{y_k}{\pi_k} + \sum_{\ell \in S_m} \frac{\mathbf{x}_\ell^\top}{\pi_\ell} \left(\widehat{\boldsymbol{\beta}}_r - \frac{\mathbf{A}_{\Pi R}^{-1} \mathbf{x}_k \widehat{\epsilon}_{Rk}}{\pi_k (1 - \widehat{h}_{kk}^\pi)} \right) \right\}$$

$$= \frac{\widehat{N}}{\widehat{N} - d_k} \left[\widehat{\mu}_{lr} - \frac{1}{\widehat{N}\pi_k} \left\{ y_k + \sum_{\ell \in S_m} \frac{\mathbf{x}_\ell^\top}{\pi_\ell} \frac{\mathbf{A}_{\Pi R}^{-1} \mathbf{x}_k \widehat{\epsilon}_{Rk}}{1 - \widehat{h}_{kk}^\pi} \right\} \right].$$

Thus, introducing response indicators, we obtain for an arbitrary element $k \in S$,

$$\widehat{\mu}_{lr}^{(k)} = \frac{\widehat{N}}{\widehat{N} - d_k} \left(\widehat{\mu}_{lr} - \frac{1}{\widehat{N}\pi_k} \widehat{\xi}_k^{(jk)} \right),$$

from which it follows that

$$\widehat{\mu}_{lr}^{(k)} - \widehat{\mu}_{lr} = \frac{d_k}{\widehat{N} - d_k} \left(\widehat{\mu}_{lr} - \widehat{\xi}_k^{(jk)} \right).$$

Therefore, a closed-form formula of e_k is given by

$$e_k = \frac{d_k}{\widehat{N}} \left(\widehat{\mu}_{lr} - \widehat{\xi}_k^{(jk)} \right).$$

Replacing e_k by its closed formula in (39) leads to the result.

B.3 Proof of Result 4.1.

Statement i).

Consider the following variance decomposition:

$$\mathbb{V}\left(\widehat{\mu}_{greg}\right) = \mathbb{E}_p\left[\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)\right] + \mathbb{V}_p\left(\mathbb{E}_m\left[\widehat{\mu}_{greg}\right]\right).$$

Note that,

$$\mathbb{E}_m\left[\widehat{\mu}_{greg}\right] = \mathbb{E}_m\left[\frac{1}{N} \sum_{k \in U} \mathbf{x}_k^\top \widehat{\boldsymbol{\beta}}_S\right] = \frac{1}{N} \sum_{k \in U} \mathbf{x}_k^\top \boldsymbol{\beta}.$$

In particular, observe that this quantity is independent of S, so that

$$\mathbb{V}\left(\widehat{\mu}_{greg}\right)=\mathbb{E}_p\left[\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)\right],$$

when the model is correctly specified. This establishes unbiasedness. For the asymptotic equivalence, we show that

$$\lim_{v\to\infty} n_v^2\times\mathbb{E}\left[\left\{\mathbb{V}\left(\widehat{\mu}_{greg}\right)-\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)\right\}^2\right]=0.$$

Write $\widehat{\mu}_{greg}$ as $\widehat{\mu}_{greg}=\boldsymbol{\mu}_x^\top\widehat{\boldsymbol{\beta}}_S$ with $\boldsymbol{\mu}_x:=\boldsymbol{t}_x/N$. It follows that

$$\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)=\boldsymbol{\mu}_x^\top\mathbb{V}_m\left(\widehat{\boldsymbol{\beta}}_S\right)\boldsymbol{\mu}_x=\frac{\sigma^2}{\pi}\boldsymbol{\mu}_x^\top\boldsymbol{A}_{\Pi S}^{-1}\boldsymbol{\mu}_x=\frac{\sigma^2}{\pi N^2}\sum_{k\in U}g_k=\frac{\sigma^2}{\pi N}\mathbb{E}_{n,U}\left[g\right]. \tag{40}$$

We begin with the following decomposition:

$$\mathbb{V}\left(\widehat{\mu}_{greg}\right)-\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)=\frac{\sigma^2}{\pi N_v^2}\sum_{k\in U_v}\left(g_{k,v}-\mathbb{E}_p\left[g_{k,v}\right]\right).$$

It thus follows that

$$\begin{aligned} n_v^2\times\mathbb{E}_p\left[\left\{\mathbb{V}\left(\widehat{\mu}_{greg}\right)-\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)\right\}^2\right]&=\frac{n_v^2\sigma^4}{\pi^2N_v^4}\mathbb{E}_p\left[\left\{\sum_{k\in U_v}\left(g_{k,v}-\mathbb{E}_p\left[g_{k,v}\right]\right)\right\}^2\right]\\ &\leqslant\frac{n_v^2\sigma^4}{\pi^2N_v^3}\sum_{k\in U_v}\mathbb{E}_p\left[\left\{\left(g_{k,v}-\mathbb{E}_p\left[g_{k,v}\right]\right)\right\}^2\right]. \end{aligned}$$

By symmetry, we obtain

$$n_v^2\times\mathbb{E}_p\left[\left\{\mathbb{V}\left(\widehat{\mu}_{greg}\right)-\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)\right\}^2\right]\leqslant\frac{n_v^2\sigma^4}{\pi^2N_v^2}\mathbb{V}_p\left(g_{1,v}\right)=o(1),$$

by assumption.

Statement ii).

Taylor: Under a Bernoulli sampling design, the Taylor variance estimator reduces to

$$\widehat{V}_{\text{tay}}=\frac{1}{N^2}\frac{1-\pi}{\pi^2}\sum_{k\in S}\widehat{\epsilon}_{kS}^2+\frac{\sigma^2}{N}.$$

Thus,

$$\mathbb{E}_m\left[\widehat{V}_{\text{tay}}\right] = \frac{1}{N^2}\frac{1-\pi}{\pi^2}\sum_{k\in S}\sigma^2\left(1-\widetilde{h}_{kk}\right)+\frac{\sigma^2}{N}$$

$$= \frac{\sigma^2}{N^2}\frac{1-\pi}{\pi^2}\left(n_s-p\right)+\frac{\sigma^2}{N}$$

$$= \frac{\sigma^2 n_s}{N^2}\frac{1-\pi}{\pi^2}\left(1-\kappa\right)+\frac{\sigma^2}{N}$$

$$\simeq \frac{\sigma^2}{N}\frac{1-\pi}{\pi}\left(1-\kappa\right)+\frac{\sigma^2}{N}.$$

Recalling (40), it follows that

$$\frac{\mathbb{E}_m\left[\widehat{V}_{\text{tay}}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)} = \frac{\frac{\sigma^2}{N}\frac{1-\pi}{\pi}\left(1-\kappa\right)}{\frac{\sigma^2}{\pi N}\mathbb{E}_{n,U}\left[g\right]}+\frac{\frac{\sigma^2}{N}}{\frac{\sigma^2}{\pi N}\mathbb{E}_{n,U}\left[g\right]} \simeq \frac{\left(1-\pi_*\right)\left(1-\kappa_*\right)}{\mathbb{E}_{n,U}\left[g\right]}+\frac{\pi_*}{\mathbb{E}_{n,U}\left[g\right]}.$$

G-weighted: Using Lemma 2, we may write

$$\mathbb{E}_m\left[\widehat{V}_{\text{g}}\right] = \frac{1}{N^2}\frac{1-\pi}{\pi^2}\sum_{k\in S}\sigma^2\left(1-\widetilde{h}_{kk}\right)g_k^2+\frac{\sigma^2}{N}$$

$$= \frac{\sigma^2}{N^2}\frac{1-\pi}{\pi^2}\left(\sum_{k\in S}g_k^2-\sum_{k\in S}\widetilde{h}_{kk}g_k^2\right)+\frac{\sigma^2}{N}$$

$$= \frac{\sigma^2}{N^2}\frac{1-\pi}{\pi^2}\left(\frac{N^2\pi^2}{\sigma^2}\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)-\sum_{k\in S}\widetilde{h}_{kk}g_k^2\right)+\frac{\sigma^2}{N}$$

$$= \left(1-\pi\right)\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)-\frac{\sigma^2}{N}\frac{1-\pi}{\pi}\mathbb{E}_{n,S}\left[g^2h\right]+\frac{\sigma^2}{N}.$$

Therefore,

$$\frac{\mathbb{E}_m\left[\widehat{V}_{\text{g}}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)} = \frac{\left(1-\pi\right)\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)-\frac{\sigma^2}{N}\frac{1-\pi}{\pi}\mathbb{E}_{n,S}\left[g^2h\right]+\frac{\sigma^2}{N}}{\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)}$$

$$= \left(1-\pi\right)-\frac{\left(1-\pi\right)\mathbb{E}_{n,S}\left[g^2h\right]}{\mathbb{E}_{n,U}\left[g\right]}+\frac{\pi}{\mathbb{E}_{n,U}\left[g\right]}$$

$$\simeq \left(1-\pi_*\right)\left\{1-\frac{\mathbb{E}_{n,S}\left[g^2h\right]}{\mathbb{E}_{n,U}\left[g\right]}\right\}+\frac{\pi_*}{\mathbb{E}_{n,U}\left[g\right]}.$$

Jackknife: In Bernoulli sampling, the Jackknife variance estimator reduces to

$$\widehat{V}_{\text{jack}} = \frac{(n-1)^2}{n^2N^2}\frac{1-\pi}{\pi^2}\sum_{k\in S}\frac{g_k^2\widehat{\epsilon}_{ks}^2}{\left(1-\widetilde{h}_{kk}\right)^2}+\frac{\sigma^2}{N} \simeq \frac{1}{N^2}\frac{1-\pi}{\pi^2}\sum_{k\in S}\frac{g_k^2\,\widehat{\epsilon}_{ks}^2}{\left(1-\widetilde{h}_{kk}\right)^2}+\frac{\sigma^2}{N},$$

so that

$$\mathbb{E}_m\left[\widehat{V}_{\text{jack}}\right] \simeq \frac{\sigma^2}{N^2}\frac{1-\pi}{\pi^2}\sum_{k\in S}\frac{g_k^2\left(1-\widetilde{h}_{kk}\right)}{\left(1-\widetilde{h}_{kk}\right)^2}+\frac{\sigma^2}{N} = \frac{\sigma^2}{N^2}\frac{1-\pi}{\pi^2}\sum_{k\in S}\frac{g_k^2}{\left(1-\widetilde{h}_{kk}\right)}+\frac{\sigma^2}{N}.$$

It follows that

$$\frac{\mathbb{E}_m\left[\widehat{V}_{\text{jack}}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg}\right)} \simeq \frac{\frac{\sigma^2}{N}\frac{1-\pi}{\pi}\mathbb{E}_{n,S}\left[g^2/\left(1-h\right)\right]}{\frac{\sigma^2}{\pi N}\mathbb{E}_{n,U}\left[g\right]}+\frac{\pi}{\mathbb{E}_{n,U}\left[g\right]} \simeq (1-\pi_*)\frac{\mathbb{E}_{n,S}\left[g^2/\left(1-h\right)\right]}{\mathbb{E}_{n,U}\left[g\right]}+\frac{\pi_*}{\mathbb{E}_{n,U}\left[g\right]}.$$

B.4 Proof of Corollary 4.1.

Under Bernoulli sampling with $\pi_* = 0$, the bias ratios of $\widehat{V}_{\text{tay}}$, $\widehat{V}_{\text{g}}$ and $\widehat{V}_{\text{jack}}$ are given by

$$\frac{\mathbb{E}_m\left[\widehat{V}_{\text{tay,v}}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)} \simeq \frac{(1-\kappa_*)}{\mathbb{E}_{n,U}\left[g\right]},$$

$$\frac{\mathbb{E}_m\left[\widehat{V}_{\text{g,v}}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)} \simeq \left(1-\frac{\mathbb{E}_{n,S}\left[g^2\widetilde{h}\right]}{\mathbb{E}_{n,U}\left[g\right]}\right),$$

$$\frac{\mathbb{E}_m\left[\widehat{V}_{\text{jack,v}}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)} \simeq \frac{\mathbb{E}_{n,S}\left[g^2/\left(1-\widetilde{h}\right)\right]}{\mathbb{E}_{n,U}\left[g\right]}.$$

Statement i).

Taylor: The statement of the bias ratio of $\widehat{V}_{tay}$ has been shown above.

G-weighted: It remains to show the inequality. Under (H1W), we may write

$$\mathbb{E}_{n,S}\left[g^2\widetilde{h}\right] = \frac{1}{n_S}\sum_{k\in S}g_k^2\widetilde{h}_{kk} \geqslant f_1\left(\frac{p_v}{n_v}\right)\frac{1}{n_S}\sum_{k\in S}g_k^2 = f_1\left(\frac{p_v}{n_v}\right)\mathbb{E}_{n,S}\left[g^2\right] \simeq f_1\left(\frac{p_v}{n_v}\right)\mathbb{E}_{n,U}\left[g\right],$$

where the last equality follows from the fact, easily shown from Lemma 2, that

$$\mathbb{E}_{n,S}\left[g^2\right] = \frac{n}{n_S}\mathbb{E}_{n,U}\left[g\right].$$

Hence, using the positivity of $\mathbb{E}_{n,U}[g]$, we obtain

$$\frac{\mathbb{E}_m\left[\widehat{V}_{g,v}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)} \simeq \left(1-\frac{\mathbb{E}_{n,S}\left[g^2\widetilde{h}\right]}{\mathbb{E}_{n,U}[g]}\right) \lesssim 1-f_1\left(\frac{p_v}{n_v}\right).$$

Jackknife: By assumption, uniformly in k, we have

$$f_1\left(\frac{p_v}{n_v}\right) \leqslant \widetilde{h}_{kk} \quad \Leftrightarrow \quad 1-f_1\left(\frac{p_v}{n_v}\right) \geqslant 1-\widetilde{h}_{kk} \quad \Leftrightarrow \quad \frac{g_k^2}{1-f_1\left(\frac{p_v}{n_v}\right)} \leqslant \frac{g_k^2}{1-\widetilde{h}_{kk}}.$$

It follows from the above that

$$\frac{1}{1-f_1\left(\frac{p_v}{n_v}\right)}\mathbb{E}_{n,S}\left[g^2\right] \leqslant \mathbb{E}_{n,S}\left[\frac{g^2}{1-h}\right] \quad \Leftrightarrow \quad \frac{\mathbb{E}_m\left[\widehat{V}_{jack,v}\right]}{\mathbb{V}_m\left(\widehat{\mu}_{greg,v}\right)} \geqslant \frac{1}{1-f_1\left(\frac{p_v}{n_v}\right)}\frac{\mathbb{E}_{n,S}\left[g^2\right]}{\mathbb{E}_{n,U}[g]} \simeq \frac{1}{1-f_1\left(\frac{p_v}{n_v}\right)}.$$

B.5 Proof of Result 5.1.

Step 1: Derivation of a target.

We first obtain an expression of the target variance of $\widehat{\mu}_{lr}$ using the method of Särndal (1992). This estimator, denoted $V_{target}(\widehat{\mu}_{lr})$, will remain unbiased even in cases where $p_v/n_v \to \kappa_* > 0$. Using the law of total variance on $\widehat{\mu}_{lr}$, we get the following decomposition:

$$\mathbb{V}(\widehat{\mu}_{lr}) = \mathbb{V}_{sam}(\widehat{\mu}_{lr}) + \mathbb{V}_{nr}(\widehat{\mu}_{lr}) + \mathbb{V}_{mix}(\widehat{\mu}_{lr}),$$

where

$$\mathbb{V}_{sam}(\widehat{\mu}_{lr}) := \mathbb{E}_m\mathbb{V}_p(\widehat{\mu}_H),$$

$$\mathbb{V}_{nr}(\widehat{\mu}_{lr}) = \mathbb{E}_q\mathbb{E}_p\mathbb{V}_m(\widehat{\mu}_{lr}-\widehat{\mu}_H),$$

and

$$\mathbb{V}_{\text{mix}}(\widehat{\mu}_{lr}) := 2\mathbb{E}_p\mathbb{E}_q\text{Cov}_m\left\{\widehat{\mu}_H-\mu_y, \widehat{\mu}_{lr}-\widehat{\mu}_H\right\},$$

and $\widehat{\mu}_H := \widehat{N}^{-1}\sum_{k\in S}\pi_k^{-1}y_k$. In the case of linear regression imputation and Bernoulli sampling, it can be shown that $\mathbb{V}_{mix}(\widehat{\mu}_{lr}) = 0$. Now, using a first-order Taylor expansion, a full sample estimator of $\mathbb{V}_{sam}(\widehat{\mu}_{lr})$ is given by

$$\widehat{V}_{sam}(\widehat{\mu}_{lr}) = \frac{1-\pi}{n_s^2}\sum_{k\in S}(y_k-\widehat{\mu}_H)^2.$$

We now turn to the nonresponse component. It can be shown that

$$\mathbb{V}_{nr} = \mathbb{E}_q \mathbb{E}_p \left[\frac{\sigma^2}{(\widehat{N}\pi)^2} \sum_{k \in S} \left\{ R_k (1 + \widehat{\Gamma}_k) - 1 \right\}^2 \right]. \tag{41}$$

Assuming σ^2 is known, the above quantity can be estimated by

$$\widehat{V}_{nr} = \frac{\sigma^2}{n_s^2} \widehat{A}_n,$$

where

$$\widehat{A}_n = \sum_{k \in S_r} (1 + \widehat{\Gamma}_k)^2 - n_s = n_m + \sum_{k \in S_r} \widehat{\Gamma}_k^2.$$

The target variance estimator is therefore given by

$$\widehat{V}_{target} := \widehat{V}_{sam} + \widehat{V}_{nr} = \frac{1-\pi}{n_s^2} \sum_{k \in S} (y_k - \widehat{\mu}_H)^2 + \frac{\sigma^2}{n_s^2} \widehat{A}_n, \tag{42}$$

Expanding the square and taking model expectations give

$$\mathbb{E}_m \left[\widehat{V}_{sam}(\widehat{\mu}_{lr}) \right] = \frac{1-\pi}{n_s^2} \sum_{k \in S} \mathbb{E}_m \left[(y_k - \widehat{\mu}_H)^2 \right]$$

$$= \frac{1-\pi}{n_s^2} \left\{ \sum_{k \in S} (\mathbf{x}_k^\top \boldsymbol{\beta})^2 + n_s \sigma^2 - \frac{2}{n_s} \sum_{k \in S} \sum_{\ell \in S} \mathbf{x}_k^\top \boldsymbol{\beta} \mathbf{x}_\ell^\top \boldsymbol{\beta} - 2\sigma^2 + \frac{1}{n} \sum_{k \in S} \sum_{\ell \in S} \mathbf{x}_k^\top \boldsymbol{\beta} \mathbf{x}_\ell^\top \boldsymbol{\beta} + \sigma^2 \right\}$$

$$= \frac{1-\pi}{n_s^2} \left\{ \sum_{k \in S} B_k + \sigma^2 (n_s - 1) \right\}, \tag{43}$$

where

$$B_k = \frac{1}{n_s} \sum_{\ell \in S} \left\{ (\mathbf{x}_\ell^\top \boldsymbol{\beta})^2 - \mathbf{x}_k^\top \boldsymbol{\beta} \mathbf{x}_\ell^\top \boldsymbol{\beta} \right\}.$$

Adding (43) and (41) and ignoring the negligible terms, we obtain an unbiased target variance estimator of $\mathbb{E}_m[V_{target}(\widehat{\mu}_{lr})]$

$$\mathbb{E}_m \left[V_{target}(\widehat{\mu}_{lr}) \right] = \mathbb{E}_m \left(\widehat{V}_{sam} + \widehat{V}_{nr} \right) = \frac{1}{n_s^2} \left\{ (1-\pi) \sum_{k \in S} B_k + \sigma^2 \left((1-\pi) n_s + n_m + \sum_{k \in S_r} \widehat{\Gamma}_k^2 \right) \right\}.$$

Step 2: Computing the model expectation of an arbitrary estimator in $\mathcal{V}$.

Let $\widehat{V}^{(\psi)}$ be an arbitrary estimator in $\mathcal{V}$. Since $\widehat{V}^{(\psi)} \in \mathcal{V}$, there exists some $\{\psi_v\}_{v\in\mathbb{N}}$ satisfying

$$\widehat{V}^{(\psi)} = \frac{1}{\widehat{N}^2} \sum_{k\in S}\sum_{\ell\in S} \frac{\Delta_{k\ell}}{\pi_{k\ell}} \frac{\widehat{\mu}_{lr} - \widehat{\xi}_k^{(\psi_v)}}{\pi_k} \frac{\widehat{\mu}_{lr} - \widehat{\xi}_\ell^{(\psi_v)}}{\pi_\ell} + \frac{\sigma^2}{\widehat{N}^2} \sum_{k\in S_r} \frac{1}{\pi_k} \left\{1 - R_k(1+\widehat{\Gamma}_k)\right\}^2, \qquad (44)$$

with

$$\widehat{\xi}_k^{(\psi_v)} := \widetilde{y}_k + r_k \psi_v(\boldsymbol{X}_R)\widehat{\Gamma}_k\, \widehat{\epsilon}_{kR}, \qquad k \in S.$$

In case of Bernoulli sampling, $\widehat{V}^{(\psi)}$ in (44) reduces to

$$\widehat{V}^{(\psi)} = \frac{1-\pi}{n_s^2} \sum_{k\in S} \left(\widehat{\mu}_{lr} - \widehat{\xi}_k^{(\psi_v)}\right)^2 + \frac{\pi\sigma^2}{n_s^2} \left(n_m + \sum_{k\in S_r} \widehat{\Gamma}_k^2\right).$$

We start by expanding the square as follows:

$$\mathbb{E}_m\left[\left(\widehat{\mu}_{lr} - \widehat{\xi}_k^{(\psi_v)}\right)^2\right] = \mathbb{E}_m\left[\left(\widehat{\xi}_k^{(\psi_v)}\right)^2\right] - 2\mathbb{E}_m\left[\widehat{\mu}_{lr}\widehat{\xi}_k^{(\psi_v)}\right] + \mathbb{E}_m\left[\widehat{\mu}_{lr}^2\right] := A_1(k) - 2A_2(k) + A_3.$$

After some tedious though relatively straightforward algebra, it can be shown that

$$A_1(k) = \left(\mathbf{x}_k^\top \boldsymbol{\beta}\right)^2 + \sigma^2 \left(R_k + [1-R_k]\,\widehat{h}_{kk} + 2R_k\widehat{\Gamma}_k\psi_v(\boldsymbol{X}_R)\left[1-\widehat{h}_{kk}\right] + R_k\widehat{\Gamma}_k^2\psi_v(\boldsymbol{X}_R)^2\left[1-\widehat{h}_{kk}\right]\right)$$

$$A_2(k) = \frac{1}{n_s}\left(\mathbf{x}_k^\top\boldsymbol{\beta}\sum_{\ell\in S}\mathbf{x}_l^\top\boldsymbol{\beta} + \sigma^2\left(1+\widehat{\Gamma}_k\right)\right)$$

$$A_3 = \frac{1}{n_s^2}\left(\left[\sum_{k\in S}\mathbf{x}_k^\top\boldsymbol{\beta}\right]^2 + \sigma^2\left[n_r + 2n_m + \sum_{k\in S_m}\widehat{\Gamma}_k\right]\right).$$

Summing over k and ignoring the negligible terms, we obtain the following asymptotic equivalence

$$\mathbb{E}_m\left[\sum_{k\in S}\left(\widehat{\mu}_{lr} - \widehat{\xi}_k^{(\psi_v)}\right)^2\right] \simeq \sum_{k\in S} B_k$$

$$+\sigma^2\left\{\left(n_r + \sum_{k\in S_m}\widehat{h}_{kk} + \left(\sum_{k\in S_r}\left(1-\widehat{h}_{kk}\right)\psi(\boldsymbol{X}_R)\Gamma_k\left\{2+\psi(\boldsymbol{X}_R)\Gamma_k\right\}\right)\right)\right..$$

It follows that

$$\mathbb{E}_m\left[\widehat{V}^{(\psi)}\right] = \frac{1-\pi}{n_s^2}\sum_{k\in S} B_k$$

$$+\sigma^2\left\{\frac{1-\pi}{n_s^2}\left(n_r + \sum_{k\in S_m}\widehat{h}_{kk} + \sum_{k\in S_r}\left(1-\widehat{h}_{kk}\right)\psi(\boldsymbol{X}_R)\Gamma_k\left\{2+\psi(\boldsymbol{X}_R)\Gamma_k\right\}\right) + \frac{\pi}{n_s^2}\left(n_m + \sum_{k\in S_r}\widehat{\Gamma}_k^2\right)\right\}.$$

Taking the ratios of $\mathbb{E}_m\left[\widehat{V}^{(\psi)}\right]$ and $\mathbb{E}_m\left[V_{target}(\widehat{\mu}_{lr})\right]$ yields the result.

B.6 Proof of Result 5.2.

We let $\widehat{V}^{(\psi)}$ be an arbitrary element of $\widetilde{\mathcal{V}}$ with constant $\psi \in \mathbb{R}$. Next, using Result 5.1 and Corollary 5.1, we write the absolute asymptotic relative bias of $\widehat{V}$ to get

$$\text{ARB}\left(\widehat{V}^{(\psi)}\right) := \frac{\left|\mathbb{E}_m\left[\widehat{V}^{(\psi)}\right] - \mathbb{E}_m\left[\widehat{V}_{target}\right]\right|}{\mathbb{E}_m\left[\widehat{V}_{target}\right]} = \frac{\sigma^2\left(1-\pi\right)\left|A_\psi - A_{theo}\right|}{\mathbb{E}_m\left[\widehat{V}_{target}\right]}, \tag{45}$$

where it can be shown that, in our setting,

$$A_\psi := n_r + \sum_{k\in S_m}\widehat{h}_{kk} + 2\psi\sum_{k\in S_r}\left(1-\widehat{h}_{kk}\right)\widehat{\Gamma}_k + \psi^2\sum_{k\in S_r}\left(1-\widehat{h}_{kk}\right)\widehat{\Gamma}_k^2,$$

$$A_{theo} := n_r + 2n_m + \sum_{k\in S_r}\Gamma_k^2.$$

The absolute asymptotic relative bias takes only positive values; we view this quantity as a function of ψ over the real line. Hence,

$$\min_{\psi\in\mathbb{R}}\frac{\left|\mathbb{E}_m\left[\widehat{V}^{(\psi)}\right] - \mathbb{E}_m\left[\widehat{V}_{target}\right]\right|}{\mathbb{E}_m\left[\widehat{V}_{target}\right]} \geqslant 0.$$

Proving the existence of zeros of the numerator is, therefore, sufficient to find minimizers of the (asymptotic) absolute relative bias. The key is to notice that

$$A_\psi - A_{theo} = \psi^2\sum_{k\in S_r}\left(1-\widehat{h}_{kk}\right)\widehat{\Gamma}_k^2 + \psi\cdot 2\sum_{k\in S_r}\left(1-\widehat{h}_{kk}\right)\widehat{\Gamma}_k + \sum_{k\in S_m}\widehat{h}_{kk} - 2n_m - \sum_{k\in S_r}\Gamma_k^2 := A\psi^2 + B\psi + C,$$

with

$$A := \sum_{k \in S_r} \left(1 - \widehat{h}_{kk}\right) \widehat{\Gamma}_k^2,$$

$$B := 2 \sum_{k \in S_r} \left(1 - \widehat{h}_{kk}\right) \widehat{\Gamma}_k,$$

$$C := \sum_{k \in S_m} \widehat{h}_{kk} - 2n_m - \sum_{k \in S_r} \widehat{\Gamma}_k^2,$$

is a degree two polynomial. If its discriminant $\Delta := B^2 - 4AC$ is positive, the following two roots

$$\psi_1 := \frac{-B - \sqrt{\Delta}}{2A}, \qquad \text{and,} \qquad \psi_2 := \frac{-B + \sqrt{\Delta}}{2A},$$

minimize the asymptotic absolute bias.

C Technical lemmas

Lemma 1. *The following relation holds:*

$$\widehat{\boldsymbol{\beta}}^{(k)} = \widehat{\boldsymbol{\beta}} - \frac{\boldsymbol{A}_{\Pi S}^{-1} d_k \mathbf{x}_k \widehat{\epsilon}_{kS}}{1 - \widetilde{h}_{kk}^{\pi}}.$$

Proof. At the sample level, any weighted least squares solution may be written

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{k \in S} w_k \mathbf{x}_k \mathbf{x}_k^\top \right)^{-1} \sum_{k \in S} w_k \mathbf{x}_k y_k := \boldsymbol{A}_{\Pi S}^{-1} \sum_{k \in S} w_k \mathbf{x}_k y_k$$

for some set of weights $\{w_k\}_{k \in S}$. Recall that for a full rank matrix $\boldsymbol{A}$ and two vectors $\mathbf{u}, \mathbf{v}$, the following identity holds

$$\left(\boldsymbol{A} - \mathbf{u}\mathbf{v}^\top\right)^{-1} = \boldsymbol{A}^{-1} + \frac{\boldsymbol{A}^{-1}\mathbf{u}\mathbf{v}^\top \boldsymbol{A}^{-1}}{1 - \mathbf{v}^\top \boldsymbol{A}^{-1}\mathbf{u}}$$

Hence, we get

$$\boldsymbol{A}_{\Pi S(k)}^{-1} = \left\{ \boldsymbol{A}_{\Pi S} - w_k \mathbf{x}_k \mathbf{x}_k^\top \right\}^{-1}$$

$$= \boldsymbol{A}_{\Pi S}^{-1} + \frac{\boldsymbol{A}_{\Pi S}^{-1} w_k \mathbf{x}_k \mathbf{x}_k^\top \boldsymbol{A}_{\Pi S}^{-1}}{1 - \mathbf{x}_k^\top \boldsymbol{A}_{\Pi S}^{-1} w_k \mathbf{x}_k}$$

$$= \boldsymbol{A}_{\Pi S}^{-1} + \frac{\boldsymbol{A}_{\Pi S}^{-1} w_k \mathbf{x}_k \mathbf{x}_k^\top \boldsymbol{A}_{\Pi S}^{-1}}{1 - h_{kk}(w)},$$

where $h_{kk}(w) := \mathbf{x}_k^\top \boldsymbol{A}_{\Pi S}^{-1} w_k \mathbf{x}_k$ denotes the weighted leverage of element k. Therefore,

$$\widehat{\boldsymbol{\beta}}^{(k)} = \boldsymbol{A}_{\Pi S(k)}^{-1} \sum_{\substack{l \in S \\ l \neq k}} w_l \mathbf{x}_l y_l$$

$$= \left\{ \boldsymbol{A}_{\Pi S}^{-1} + \frac{\boldsymbol{A}_{\Pi S}^{-1} w_k \mathbf{x}_k \mathbf{x}_k^\top \boldsymbol{A}_{\Pi S}^{-1}}{1 - h_{kk}(w)} \right\} \left\{ \sum_{l \in S} w_l \mathbf{x}_l y_l - w_k \mathbf{x}_k y_k \right\}$$

$$= \widehat{\boldsymbol{\beta}} - \frac{\boldsymbol{A}_{\Pi S}^{-1} w_k \mathbf{x}_k}{1 - h_{kk}(w)} \left\{ y_k \left(1 - h_{kk}(w)\right) + \mathbf{x}_k^\top \boldsymbol{\beta} + h_{kk}(w) y_k \right\}$$

$$= \widehat{\boldsymbol{\beta}} - \frac{\boldsymbol{A}_{\Pi S}^{-1} w_k \mathbf{x}_k \widehat{\epsilon}_{kS}}{1 - h_{kk}(w)}.$$

This establishes the result. ■

Lemma 2. *The following relation holds:*

$$\sum_{k \in S} g_k^2 = \pi \sum_{k \in U} g_k = \frac{N^2 \pi^2}{\sigma^2} \mathbb{V}_m \left(\widehat{\mu}_{greg}\right).$$

Proof. By definition,

$$\sum_{k \in S} g_k^2 = \sum_{k \in S} \boldsymbol{t}_{\mathbf{x}}^\top \boldsymbol{A}_{\Pi S}^{-1} \mathbf{x}_k \mathbf{x}_k^\top \boldsymbol{A}_{\Pi S}^{-1} \boldsymbol{t}_{\mathbf{x}}$$

$$= \pi \times \boldsymbol{t}_{\mathbf{x}}^\top \boldsymbol{A}_{\Pi S}^{-1} \sum_{k \in S} \frac{\mathbf{x}_k \mathbf{x}_k}{\pi} \boldsymbol{A}_{\Pi S}^{-1} \boldsymbol{t}_{\mathbf{x}}$$

$$= \pi \times \boldsymbol{t}_{\mathbf{x}}^\top \boldsymbol{A}_{\Pi S}^{-1} \boldsymbol{t}_{\mathbf{x}}.$$

Moreover,

$$\sum_{k \in U} g_k = \boldsymbol{t}_{\mathbf{x}}^\top \boldsymbol{A}_{\Pi S}^{-1} \mathbf{x}_k = \boldsymbol{t}_{\mathbf{x}}^\top \boldsymbol{A}_{\Pi S}^{-1} \boldsymbol{t}_{\mathbf{x}},$$

from which the first equality follows. The second equality follows from (40). ■