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

Survey Methodology

Tests for evaluating nonresponse bias in surveys

by Sharon L. Lohr, Minsun K. Riddles and David Morganstein

Release date: December 20, 2016

Statistics Canada Statistique Canada

Canada

How to obtain more information

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

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-877-287-4369

Depository Services Program

• Inquiries line	1-800-635-7943
• Fax line	1-800-565-7757

Standards of service to the public

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

Note of appreciation

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

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
p preliminary
r revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
E use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2016
Vol. 42, No. 2, pp. 195-218
Statistics Canada, Catalogue No. 12-001-X

Tests for evaluating nonresponse bias in surveys

Sharon L. Lohr, Minsun K. Riddles and David Morganstein[1]

Abstract

How do we tell whether weighting adjustments reduce nonresponse bias? If a variable is measured for everyone in the selected sample, then the design weights can be used to calculate an approximately unbiased estimate of the population mean or total for that variable. A second estimate of the population mean or total can be calculated using the survey respondents only, with weights that have been adjusted for nonresponse. If the two estimates disagree, then there is evidence that the weight adjustments may not have removed the nonresponse bias for that variable. In this paper we develop the theoretical properties of linearization and jackknife variance estimators for evaluating the bias of an estimated population mean or total by comparing estimates calculated from overlapping subsets of the same data with different sets of weights, when poststratification or inverse propensity weighting is used for the nonresponse adjustments to the weights. We provide sufficient conditions on the population, sample, and response mechanism for the variance estimators to be consistent, and demonstrate their small-sample properties through a simulation study.

Key Words: Inverse propensity weighting; Poststratification; Replication variance estimation; Responsive design.

1 Introduction

Nonresponse rates in probability samples are increasing worldwide. The U.S. Office of Management and Budget requires a nonresponse bias analysis when response rates are low or there are other indications that bias may be a problem (United States Office of Management and Budget 2006). Groves (2006) recommended using multiple approaches to assess potential nonresponse bias on key survey estimates.

Assessing potential nonresponse bias typically requires an external “gold standard” data source or rich sampling frame information. Common approaches for assessing nonresponse bias include: (1) comparing frame variables for respondents and nonrespondents, (2) comparing early and late respondents on frame variables and key survey variables, and (3) comparing estimates from the survey respondents (using nonresponse-adjusted weights) with estimates from an independent gold standard source. Differences in (1) and (2), however, do not necessarily imply that nonresponse bias remains after the weights are adjusted through calibration or propensity methods. If weight adjustments such as those described in Brick (2013) are successful in adjusting for nonresponse bias, the estimates from the survey using the nonresponse-adjusted weights may be approximately unbiased even if assessments (1) and (2) show differences.

In this paper we compare an estimate calculated using base weights from the selected sample with an estimate of the same quantity calculated using nonresponse-adjusted weights from the respondents only. An example might be comparing the estimated proportion of persons living in census tracts with more than 50% of housing units being owner occupied from (1) the selected sample, using the base weights, (2) the respondents, using the base weights, and (3) the respondents, using nonresponse-adjusted and/or poststratified weights. All three estimates of the proportion use the same characteristic, y, which is assumed to be known for everyone in the selected sample.

1. Sharon Lohr and David Morganstein are Vice Presidents and Minsun Riddles is Statistician at Westat, 1600 Research Blvd., Rockville MD 20850. E-mail: sharonlohr@westat.com.

The requirement that y be known for the selected sample restricts the set of variables that can be used to test for nonresponse bias. Typically, many of the key variables of interest are available only for the respondents, not for the entire selected sample. Other variables that are available for the entire selected sample may be used for poststratification or other nonresponse weighting adjustments. Poststratification forces the estimates of population totals for poststratification variables to equal the independent population counts for these variables, so these variables would not be expected to exhibit nonresponse bias after weight adjustments are performed. Variables that are available for the entire selected sample but are not used in the nonresponse weighting adjustments, and variables that are correlated with key survey variables, are the best choices for testing nonresponse bias. Examples of such variables include sample frame variables that are not used in poststratification (for example, an e-mail survey of university students may have information on academic performance that is not used in the nonresponse weighting), characteristics from a census (such as percent poverty in the block containing the sampled address), or information gathered by the interviewer (such as indications of children in the household that are visible from the street).

Eltinge (2002) and Harris-Kojetin (2012) recommended comparing estimates using different sets of weights to assess nonresponse bias and to choose among competing sets of nonresponse-adjusted weights. Such comparisons are common in nonresponse bias analyses: for example, Hamrick (2012) compared respondents with the full sample in the Eating and Health Module of the American Time Use Survey. To date, however, there has been no comprehensive examination of the statistical properties underlying these comparisons. In this paper, we derive the theoretical properties of variance estimators and hypothesis tests for the differences among estimated means that are calculated using the same outcome variable but with different weights and subsets of the data, and give conditions that will ensure consistency of the variance estimators.

Poststratification or inverse propensity weighting are commonly used to compensate for nonresponse bias. Yung and Rao (2000) derived linearization and jackknife estimators for the variance of a population mean estimated using poststratification, with and without nonresponse. They considered a uniform response mechanism in which each poststratum has the same response propensity, and considered the response indicator to be a fixed characteristic of the finite population. Kim and Kim (2007) studied asymptotic properties for inverse propensity weight adjustments, assuming that the response indicators of different units are independent. The previous work studied the variance of the poststratified or inverse-propensity-weighted statistic of interest. The problem we consider differs from the previous work because the estimated population total from the selected sample is often highly correlated with the estimate calculated using the respondents only, particularly when the response rate approaches one. The linearization and replication variance estimators in this paper account for that high correlation between the two sets of estimates, and thus can be used for testing the hypothesis that the poststratification or inverse propensity weighting removes the bias for the variables studied. We also extend previous research by allowing the response indicators to be correlated within primary sampling units, reflecting possible within-cluster homogeneity for responding to the survey.

Section 2 defines the parameter to be tested in the poststratification setting, derives the linearization and jackknife variance estimators, and gives sufficient conditions for the variance estimators to be consistent. In some circumstances the linearized variance of the test statistic may be zero under the null hypothesis, in which case higher-order terms of the variance are needed. The higher-order terms are derived for the special case of simple random sampling in Theorem 3. Section 3 provides the linearization and jackknife variance estimators for testing the hypothesis that the propensity weights remove the nonresponse bias. Section 4 presents simulation studies and Section 5 contains concluding remarks and discusses future work.

2 Poststratification

2.1 Parameter and linearization variance

Suppose the finite population U has H strata, with N_h primary sampling units (PSUs) in stratum h, M_{hi} units in PSU i of stratum h, and $M = \sum_{hi} M_{hi}$ units in total. Let y_{hik} denote the quantity of interest for unit k in PSU (hi). A probability sample S is taken from the population, with n_h PSUs selected from stratum h and $n = \sum_{h=1}^{H} n_h$. The sample of PSUs from stratum h is denoted by S_h, and the sample of units from PSU (hi) is denoted by S_{hi}. Each unit has a design weight $w_{hik} = 1/P(\text{unit } hik \in S)$, and the PSU-level design weight is $w_{hi} = 1/P(\text{PSU } hi \in S_h)$.

Two frameworks are commonly used for the nonresponse mechanism. In a two-phase "forward" framework, the sample is selected at phase 1 and the nonresponse mechanism is a second phase of selection (Oh and Scheuren 1987; Särndal and Lundström 2005). Fay (1991) proposed a "reverse framework" which was studied further by Shao and Steel (1999) and Haziza, Thompson, and Yung (2010). In this framework, the nonresponse mechanism is applied to the finite population first, and then the sample is selected. The reverse framework, which we follow in this paper, specifies a nonresponse mechanism for nonsampled as well as sampled units. We assume that every unit in the population has a value of the response indicator r_{hik}. Let $R_{hik} = E[r_{hik}]$ under the response mechanism in the finite population, so that R_{hik} is the value of the true response propensity for unit (hik) in the population.

Suppose the characteristic y is known for all units in the selected sample. We compare the estimated population total using everyone in the sample with the estimated total using the poststratification-weighted respondents. There are C poststrata and poststratum c has M_c population units with $M = \sum_{c=1}^{C} M_c$. The poststratum counts M_c may be obtained from the sampling frame if the poststratification variables are known for every unit in the frame. Often, however, the poststratum counts come from an external source such as a census. Let $\delta_{chik} = 1$ if unit (hik) is in poststratum c and 0 otherwise. The population response rate in poststratum c is $p_c = \sum_{hik \in U} R_{hik} \delta_{chik} \big/ M_c$. Yung and Rao (2000) assumed that the response rate p_c was the same for each poststratum. In many applications, however, the poststrata are formed so that response propensities within each poststratum are homogeneous, but the poststrata themselves have different mean response propensities. We therefore allow p_c to differ among the poststrata.

If y is known for all members of the selected sample, then the estimator of the population total using the sample is

$$\hat{Y}_{SS} = \sum_{hik \in S} w_{hik} y_{hik} = \sum_{hik \in U} Z_{hik} w_{hik} y_{hik}, \tag{2.1}$$

where w_{hik} is the design weight for unit k of PSU i in stratum h and Z_{hik} is the indicator variable for sample inclusion. Using the respondents only, the poststratified estimator of the population total is

$$\hat{Y}_{PS} = \sum_{c=1}^{C} M_c \frac{\sum_{hik \in S} w_{hik} r_{hik} \delta_{chik} y_{hik}}{\sum_{hik \in S} w_{hik} r_{hik} \delta_{chik}} = \sum_{c=1}^{C} M_c \frac{\hat{Y}_c^R}{\hat{M}_c^R}. \tag{2.2}$$

We define the finite population parameter of interest to be the difference between the expected value of $\hat{Y}_{PS}$ and the expected value of $\hat{Y}_{SS}$, which will be 0 if there is no nonresponse bias after poststratification. Define

$$\begin{aligned} M_c^R &= \sum_{hik \in U} \delta_{chik} R_{hik} = p_c M_c, \\ Y_c^R &= \sum_{hik \in U} \delta_{chik} R_{hik} y_{hik}, \end{aligned}$$

and

$$\theta = \sum_{c=1}^{C} M_c \frac{Y_c^R}{M_c^R} - Y = \sum_{c=1}^{C} \frac{Y_c^R}{p_c} - Y. \tag{2.3}$$

Using the relation $\sum_{hik \in U} \delta_{chik} \left(R_{hik} - p_c\right) = 0$,

$$\begin{aligned} \theta &= \sum_{c=1}^{C} \sum_{hik \in U} y_{hik} \delta_{chik} \left(\frac{R_{hik}}{p_c} - 1\right) \\ &= \sum_{c=1}^{C} \sum_{hik \in U} \delta_{chik} \left(\frac{R_{hik}}{p_c} - 1\right)\left(y_{hik} - \frac{Y_c^R}{M_c^R}\right). \end{aligned}$$

We are interested in testing the hypothesis $H_0: \theta = 0$ vs. $H_A: \theta \neq 0$, or alternatively in obtaining a confidence interval for θ. If the response propensity in each poststratum c is uniform with $R_{hik} = p_c$ for all units having $\delta_{chik} = 1$, then θ will be zero. Alternatively, $\theta = 0$ if there is no variability in the response variable y_{hik} within each poststratum. If either of these conditions holds, poststratification corrects for bias from nonresponse. Note that if each of the poststrata has uniform response propensity – that is, the poststratification variables completely explain the variability in underlying response propensities – then the poststratification will in fact remove bias for every possible y variable. If the variance of y_{hik} is 0 within each poststratum, poststratification removes bias for y but it does not necessarily remove bias for other variables.

We estimate θ by $\hat{\theta} = \hat{Y}_{PS} - \hat{Y}_{SS}$, which may be rewritten as

$$\hat{\theta} = \hat{Y}_{PS} - \hat{Y}_{SS} = \sum_{c=1}^{C} \frac{1}{p_c}\left(\hat{Y}_c^R - \overline{Y}_c^R\left(\hat{M}_c^R - M_c^R\right) + \hat{T}_c\right) - \hat{Y}_{SS}, \tag{2.4}$$

where $\overline{Y}_c^R = Y_c^R / M_c^R$, $\overline{y}_c^R = \hat{Y}_c^R / \hat{M}_c^R$, and

$$\hat{T}_c = -\left(\bar{y}_c^R - \bar{Y}_c^R\right)\left(\hat{M}_c^R - M_c^R\right). \tag{2.5}$$

Theorem 1 gives the variance of $\hat{\theta}$. Define

$$e_{Rhik} = \sum_{c=1}^{C} \delta_{chik} \left\{ \frac{R_{hik}}{p_c} \left(y_{hik} - \bar{Y}_c^R \right) - y_{hik} \right\}.$$

We assume the following regularity conditions.

(A1) The number of poststrata, C, is fixed and $M_c / M \to \lambda_c \in (0,1)$.

(A2) There exists a constant K such that $|y_{hik}| < K$ for all (hik).

(A3) $\max_{hik} w_{hik} = O(M/n)$ and $\max_{hik} w_{hik} / w_{hi}$ is bounded.

(A4) $R_{hik} > \varepsilon$ for all (hik), for a fixed $\varepsilon > 0$. This guarantees that every unit has a positive response propensity that is bounded away from 0.

(A5) The vector of response indicators $\mathbf{r} = [r_{hik}]$ is independent of the vector of sample inclusion indicators $\mathbf{Z} = [Z_{hik}]$. In addition, r_{hik} and r_{ljp} are independent when $(hi) \neq (lj)$, so that the response indicators in different PSUs are uncorrelated.

Assumptions (A1) and (A4) ensure that the denominator in (2.3) is nonzero almost surely. Assumption (A2) could be replaced by weaker Liapunov-type conditions such as those in Theorem 1.3.2 of Fuller (2009) or Yung and Rao (2000) if more restrictive assumptions are placed on the covariance structure of the response indicators r_{hik}; however, in practice it can be assumed that almost any characteristic measured in a finite population is bounded. Assumption (A5) is weaker than the assumption used in Kim and Kim (2007) that the response indicators are independent across units. With assumption (A5), individuals in the same PSU (for example, persons in the same household or same city) may exhibit dependence when choosing whether to respond to a survey, but the response indicators of individuals in different PSUs are independent.

Theorem 1. *Under conditions (A1) – (A5), the variance of* $\hat{\theta}$ *is*

$$V\left(\hat{\theta}\right) = V_1\left(\hat{\theta}\right) + V_2\left(\hat{\theta}\right),$$

where

$$V_1\left(\hat{\theta}\right) = V\left(\sum_{hik \in U} Z_{hik} w_{hik} e_{Rhik} \right) + E\left[V\left[\sum_{hik \in U} Z_{hik} w_{hik} \sum_{c=1}^{C} \delta_{chik} \frac{r_{hik}}{p_c} \left(y_{hik} - \bar{Y}_c^R \right) \middle| \mathbf{Z} \right] \right] \tag{2.6}$$

and

$$V_2\left(\hat{\theta}\right) = V\left[\sum_{c=1}^{C} \frac{\hat{T}_c}{p_c} \right] + 2 \ \mathrm{Cov}\left[\sum_{c=1}^{C} \frac{\hat{T}_c}{p_c}, \sum_{c=1}^{C} \frac{\left(\bar{y}_c^R - \bar{Y}_c^R\right) \hat{M}_c^R}{p_c} - \hat{Y}_{SS} \right] = o\left(M^2/n\right).$$

The proof is given in the appendix. Usually, only $V_1\left(\hat{\theta}\right)$ would be considered because for most applications it has higher order than $V_2\left(\hat{\theta}\right)$. Unlike situations typically studied in survey sampling, however, the first-order term of the linearization variance can be zero for some situations, and in those cases

$V(\hat{\theta}) = V_2(\hat{\theta})$. If the first-order term is not exactly zero but has order $o(M^2/n)$, both terms of the variance are needed.

The second term in (2.6) equals 0 if $p_c = 1$ for all poststrata c (that is, there is full response), or if there is no variability among the y values within poststratum c for each poststratum with $p_c < 1$. If the response indicators r_{hik} are all independent, then

$$E\left[V\left(\sum_{hik\in U} Z_{hik} w_{hik} \sum_{c=1}^{C} \delta_{chik} \frac{r_{hik}}{p_c}\left(y_{hik} - \bar{Y}_c^R\right) \middle| \mathbf{Z}\right)\right] = \sum_{hik\in U} w_{hik} \sum_{c=1}^{C} \delta_{chik} \frac{1-p_c}{p_c}\left(y_{hik} - \bar{Y}_c^R\right)^2.$$

Under the hypothesized uniform response propensity mechanism that $R_{hik} = p_c$ for all population units in poststratum c, the first term in (2.6) is

$$V\left(\sum_{hik\in U} Z_{hik} w_{hik} e_{Rhik}\right) = V\left\{\sum_{hik\in U} Z_{hik} w_{hik} \sum_{c=1}^{C} \delta_{chik}\left(-\bar{Y}_c^R\right)\right\} = V\left(\sum_{c=1}^{C} \hat{M}_c \bar{Y}_c^R\right).$$

If response propensities are uniform, this term equals zero if the population mean of $\bar{Y}_c^R$ is the same for all poststrata and the estimated poststratum sizes sum to M.

If $(n/M^2)V_1(\hat{\theta})$ converges to a positive constant, a linearization variance estimator for $V(\hat{\theta})$ is

$$\hat{V}_L(\hat{\theta}) = \sum_{h=1}^{H} \frac{n_h}{n_h - 1} \sum_{i\in S_h} \left(b_{hi} - b_h\right)^2 \tag{2.7}$$

where

$$b_{hi} = \sum_{k\in S_{hi}} w_{hik} \left\{\sum_{c=1}^{C} \frac{M_c}{\hat{M}_c^R} r_{hik} \delta_{chik}\left(y_{hik} - \bar{y}_c^R\right) - y_{hik}\right\}$$

and

$$b_h = \frac{1}{n_h} \sum_{i\in S_h} b_{hi}.$$

Theorem 2. *Suppose conditions (A1) – (A5) hold and that $(n/M^2)V_1(\hat{\theta})$ converges to a positive constant. Then $(n/M^2)\left[\hat{V}_L(\hat{\theta}) - V_1(\hat{\theta})\right] \to 0$ in probability.*

Theorem 2 is proven in the Appendix.

2.2 Higher-order terms of the variance

When $V_1(\hat{\theta}) = o(M^2/n)$, the higher-order terms of the variance are needed. Theorem 3 gives these higher-order terms for the special case of simple random sampling. For simple random sampling, each unit is denoted by the subscript i instead of hik.

Theorem 3. *Suppose conditions (A1) – (A5) are met, and that a simple random sample of n units is selected from the population of M units, where $n/M \to 0$. Let $\hat{Y}_c^{NR} = \sum_{i \in S} w_i \delta_{ci} y_i \left(1 - r_i\right)$ be the estimated total for the nonrespondents in poststratum c. Assume that $\bar{y}_c^R$ is independent of $\hat{M}_c^R$ and $\hat{Y}_c^{NR}$, and that all r_i are independent and are independent of Z_i. Then*

$$V_2\left(\hat{\theta}\right) = \sum_{c=1}^{C} \frac{2p_c - 1}{p_c^2} V\left[\bar{y}_c^R - \bar{Y}_c^R\right] V\left[\hat{M}_c^R - M_c^R\right] + o\left(M^2/n^2\right).$$

We can estimate $V_2\left(\hat{\theta}\right)$ in a simple random sample by

$$\sum_{c=1}^{C} \frac{2\hat{p}_c - 1}{\hat{p}_c^2} \frac{s_c^2}{n_c^R} \frac{M_c \hat{p}_c \left(M - M_c \hat{p}_c\right)}{n},$$

where $\hat{p}_c$ is the empirical response rate in poststratum c, n_c^R is the number of respondents in poststratum c, and s_c^2 is the sample variance of y for the respondents in poststratum c.

In practice, the estimated first-order term of the variance using (2.7) will in general be nonzero even when $V_1\left(\hat{\theta}\right) = 0$. Thus, the estimated first-order term cannot be used to diagnose whether the higher-order terms are needed. However, the variance expression in (2.6) implies that $V_1\left(\hat{\theta}\right)$ is sufficiently large for the first-order approximation to be valid when all poststrata have response rates bounded away from one and non-negligible within-poststratum variance.

2.3 Jackknife

The jackknife estimator of the variance is defined as follows:

$$\hat{V}_J\left(\hat{\theta}\right) = \sum_{g=1}^{H} \frac{n_g - 1}{n_g} \sum_{j \in S_g} \left(\hat{\theta}^{(gj)} - \hat{\theta}\right)^2, \tag{2.8}$$

where

$$\begin{aligned} \hat{\theta}^{(gj)} &= \hat{Y}_{PS}^{(gj)} - \hat{Y}_{SS}^{(gj)}, \\ \hat{Y}_{PS}^{(gj)} &= \sum_{c=1}^{C} M_c \frac{\sum_{hik \in S} w_{hik}^{(gj)} r_{hik} \delta_{chik} y_{hik}}{\sum_{hik \in S} w_{hik}^{(gj)} r_{hik} \delta_{chik}}, \\ \hat{Y}_{SS}^{(gj)} &= \sum_{hik \in S} w_{hik}^{(gj)} y_{hik}, \end{aligned}$$

and the jackknife weights are:

$$w_{hik}^{(gj)} = \begin{cases} 0 & \text{if} \quad (hi) = (gj) \\ \dfrac{n_h}{n_h - 1} w_{hik} & \text{if} \quad h = g, i \neq j. \\ w_{hik} & \text{if} \quad h \neq g \end{cases} \tag{2.9}$$

If $(n/M^2)V_1(\hat{\theta})$ converges to a positive constant and assumptions (A1) – (A5) hold, then $\hat{V}_J(\hat{\theta})/V_1(\hat{\theta})$ converges to 1 in probability. This follows by standard jackknife arguments (Theorem 6.1 of Shao and Tu 1995) because the population parameter is a continuously differentiable function of population totals. Under the conditions of Theorem 2, either $\hat{\theta}/\sqrt{\hat{V}_L(\hat{\theta})}$ or $\hat{\theta}/\sqrt{\hat{V}_J(\hat{\theta})}$ may be used as a test statistic. Each approximately follows a standard normal distribution when the null hypothesis $H_0: \theta = 0$ is true.

2.4 Remarks and extensions

In this section we derived the linearization variance estimator for comparing the estimated population total of a quantity known for everyone in the selected sample with the poststratified estimate calculated using the respondents only. Theorems 1 and 2 also give the variance and variance estimator for comparing the estimator calculated using the selected sample with that from the base-weighted respondents. In that case, $\hat{Y}_{PS}$ reduces to an estimator with one poststratum, $\hat{Y}_{PS} = (M/\hat{M}^R)\hat{Y}^R$, where $\hat{M}^R = \sum_{(hik)\in S} w_{hik} r_{hik}$.

What happens if y is one of the poststratification variables? In the framework used in this section, the population counts for the poststratification variables are obtained from the sampling frame or an external source. If y is a linear combination of poststratification class indicators, then $\hat{Y}_{PS}$ is the same for all possible samples and thus has zero variance. Then $V(\hat{\theta}) = V(\hat{Y}_{SS})$, which is the first-order term of the variance in Theorem 1. If y is also a stratification variable in the design, then $V(\hat{\theta})$ will be zero. If y is not a stratification variable, then typically $\hat{Y}_{SS}$ will vary from sample to sample and will have variance of order $O(M^2/n)$ so that the test of nonresponse bias can be performed. We would expect the rejection rate for the test to be the significance level α in this case.

The parameter θ in (2.3) was defined as the difference between the poststratified population total, calculated using the population response propensities under the poststratification scheme adopted, and the unadjusted population total. In (2.4), the unadjusted population total Y was estimated by the Horvitz-Thompson estimator. The parameter θ could alternatively be estimated by

$$\hat{\theta}_2 = \hat{Y}_{PS} - \sum_{c=1}^{C} M_c \frac{\hat{Y}_c}{\hat{M}_c},$$

in which a poststratified estimator is used instead of $\hat{Y}_{SS}$. The variance of $\hat{\theta}_2$ is expected to be less than the variance of $\hat{\theta}$ under the poststratification assumptions, resulting in a more powerful test. However, when y is a linear combination of the poststratum indicators, the statistic $\hat{\theta}_2$ cannot be used to test $H_0: \theta = 0$ because $V(\hat{\theta}_2) = 0$. A similar problem can occur when y is highly correlated with the poststratification variables. The estimator $\hat{\theta}$, by contrast, typically has positive variance even when y is one of the poststratification variables.

Sometimes poststratification is performed using less-than-perfect poststratification totals – for example, the totals may come from a large survey such as the American Community Survey which has its own sampling and nonsampling errors, or they may be from a census of a slightly different population. In some cases, poststratification variables such as race or ethnicity may be measured differently in the survey than in the source of the external population totals. Using $\hat{\theta}$ rather than $\hat{\theta}_2$ may detect differences that might be caused by a flawed poststratification.

If desired, the tests may be performed using means rather than totals. In this case, the population parameter is

$$\theta_M = \sum_{c=1}^{C} \frac{M_c}{M} \frac{Y_c^R}{M_c^R} - \bar{Y}$$

where $\bar{Y} = Y/M$, and may be estimated by

$$\hat{\theta}_M = \sum_{c=1}^{C} \frac{M_c}{M} \frac{\hat{Y}_c^R}{\hat{M}_c^R} - \frac{\hat{Y}_{SS}}{\sum_{hik \in S} w_{hik}}. \quad (2.10)$$

3 Propensity weighting

An alternative to poststratification is to use inverse propensity weighting of the respondents (see, for example, Folsom 1991; Kim and Kim 2007).

In this framework, the true response propensity of unit (hik) is R_{hik} and a model is used to predict the propensity from characteristics known for everyone in the selected sample. Logistic regression is often used to estimate propensities. Suppose that the p- vector $\mathbf{x}_{hik}$ is known for each unit in S. The modeled response propensity, if $\mathbf{x}_{hik}$ and R_{hik} were known for each unit in the population, is

$$R_{hik}^M = \left[1 + \exp\left(-\mathbf{x}_{hik}' \boldsymbol{\beta}\right)\right]^{-1},$$

where $\boldsymbol{\beta}$ is the solution to the expected population score equations

$$\sum_{hik \in U} \left[R_{hik} - R_{hik}^M\right] \mathbf{x}_{hik} = 0.$$

The model removes the bias for the estimated population total of y if

$$\theta = \sum_{hik \in U} \left[R_{hik} \frac{y_{hik}}{R_{hik}^M} - y_{hik}\right] \quad (3.1)$$

equals 0. If $R_{hik} = R_{hik}^M$, that is, the response propensity model is correctly specified, then the weighting adjustments remove the bias for every possible response variable y. The population parameter θ is estimated by

$$\hat{\theta} = \sum_{hik \in S} w_{hik} \left[r_{hik} y_{hik} \left[1 + \exp\left(-\mathbf{x}_{hik}' \hat{\boldsymbol{\beta}}\right)\right] - y_{hik}\right],$$

where $\hat{\boldsymbol{\beta}}$ is the solution to the pseudolikelihood score equations

$$\sum_{hik \in S} w_{hik} \left[r_{hik} - \left[1 + \exp\left(-\mathbf{x}_{hik}' \hat{\boldsymbol{\beta}}\right)\right]^{-1}\right] \mathbf{x}_{hik} = 0.$$

Unlike the poststratification situation, the population parameter θ in (3.1) is not an explicit function of population totals. Similarly to Kim and Kim (2007), we can obtain the linearization variance and a linearization variance estimator of $\hat{\theta}$ by using the estimating equation for $(\boldsymbol{\beta}, \theta)$, as derived in Binder (1983): $(\hat{\boldsymbol{\beta}}, \hat{\theta})$ is the solution to

$$\hat{\mathbf{A}}(\boldsymbol{\beta},\theta,\mathbf{r}) = \sum_{hik\in S} w_{hik}\mathbf{u}\left(y_{hik},\mathbf{x}_{hik},r_{hik},\boldsymbol{\beta}\right) - [0,0,\ldots,0,\theta]' = 0, \tag{3.2}$$

where

$$\mathbf{u}\left(y_{hik},\mathbf{x}_{hik},r_{hik},\boldsymbol{\beta}\right) = \begin{bmatrix}\mathbf{u}_1\left(y_{hik},\mathbf{x}_{hik},r_{hik},\boldsymbol{\beta}\right)\\ u_2\left(y_{hik},\mathbf{x}_{hik},r_{hik},\boldsymbol{\beta}\right)\end{bmatrix} = \begin{bmatrix}\left[r_{hik} - \left[1+\exp\left(-\mathbf{x}'_{hik}\boldsymbol{\beta}\right)\right]^{-1}\right]\mathbf{x}_{hik}\\ r_{hik}y_{hik}\left[1+\exp\left(-\mathbf{x}'_{hik}\boldsymbol{\beta}\right)\right] - y_{hik}\end{bmatrix}.$$

The population parameter θ solves the population estimating equation

$$\mathbf{A}(\boldsymbol{\beta},\theta,\mathbf{R}) = \sum_{hik\in U} \mathbf{u}\left(y_{hik},\mathbf{x}_{hik},R_{hik},\boldsymbol{\beta}\right) - [0,0,\ldots,0,\theta]' = 0.$$

Theorem 4. *Let* $\hat{\mathbf{U}}(\boldsymbol{\beta},\theta) = \sum_{hik\in S} w_{hik}\mathbf{u}\left(y_{hik},\mathbf{x}_{hik},r_{hik},\boldsymbol{\beta}\right) = \left[\hat{\mathbf{U}}_1(\boldsymbol{\beta})',\hat{U}_2(\boldsymbol{\beta})\right]'$. *Suppose conditions (A2) – (A5) are met and there exists a value* B *such that* $\left|\mathbf{x}_{hik,j}\right| < B$ *for all units* (hik) *and components* j. *Then* $V\left(\hat{\theta}\right) = V_L\left(\hat{\theta}\right) + o(M^2/n)$, *where*

$$V_L\left(\hat{\theta}\right) = \mathbf{T}'\mathbf{QXC}V\left[\hat{\mathbf{U}}_1(\boldsymbol{\beta})\right]\mathbf{CX}'\mathbf{QT} - 2\mathbf{T}'\mathbf{QXC}\,Cov\left[\hat{\mathbf{U}}_1(\boldsymbol{\beta}),\hat{U}_2(\boldsymbol{\beta})\right] + V\left[\hat{U}_2(\boldsymbol{\beta})\right], \tag{3.3}$$

$\mathbf{X}$ *is the* $M\times p$ *matrix with rows* $\mathbf{x}'_{hik}$, $\mathbf{T}$ *is the* M-*vector with elements* $R_{hik}y_{hik}$, $\mathbf{Q}$ *is the* $M\times M$ *diagonal matrix with entries* $\exp\left(-\mathbf{x}'_{hik}\boldsymbol{\beta}\right)$, *and* $\mathbf{C} = \left(\mathbf{X}'[\mathbf{I}+\mathbf{Q}]^{-2}\,\mathbf{QX}\right)^{-1}$.

A linearization variance estimator for $\hat{\theta}$ may be obtained by substituting estimators for the population quantities in (3.3) to obtain

$$\begin{aligned}\hat{V}_L\left(\hat{\boldsymbol{\beta}},\hat{\theta}\right) &= \mathbf{t}'_S\mathbf{W}_S\mathbf{Q}_S\mathbf{X}_S\hat{\mathbf{C}}\hat{V}\left[\hat{\mathbf{U}}_1\left(\hat{\boldsymbol{\beta}}\right)\right]\hat{\mathbf{C}}\mathbf{X}'_S\mathbf{Q}_S\mathbf{W}_S\mathbf{t}_S\\ &\quad - 2\mathbf{t}'_S\mathbf{W}_S\mathbf{Q}_S\mathbf{X}_S\hat{\mathbf{C}}\,\widehat{\mathrm{Cov}}\left[\hat{\mathbf{U}}_1\left(\hat{\boldsymbol{\beta}}\right),\hat{U}_2\left(\hat{\boldsymbol{\beta}}\right)\right] + \hat{V}\left[\hat{U}_2\left(\hat{\boldsymbol{\beta}}\right)\right],\end{aligned}$$

where $\mathbf{X}_S$ is the $m\times p$ matrix with rows $\mathbf{x}'_{hik}$ for the sampled units with m the size of the selected sample, $\mathbf{W}_S$ is the $m\times m$ diagonal matrix of weights w_{hik} for sampled units, $\mathbf{t}_S$ is the m- vector with elements $r_{hik}y_{hik}$ for sampled units, $\mathbf{Q}_S$ is the $m\times m$ diagonal matrix with entries $\exp\left(-\mathbf{x}'_{hik}\hat{\boldsymbol{\beta}}\right)$ for values of $\mathbf{x}_{hik}$ in the sample, and $\hat{\mathbf{C}} = \left(\mathbf{X}'_S\mathbf{W}_S\,[\mathbf{I}+\mathbf{Q}_S]^{-2}\,\mathbf{Q}_S\mathbf{X}_S\right)^{-1}$.

The jackknife variance estimator for inverse propensity weighting is defined using the formula in (2.8) with jackknife weights in (2.9). For the propensity setting,

$$\hat{\theta}^{(gj)} = \sum_{hik\in S} w_{hik}^{(gj)}\left[r_{hik}y_{hik}\left[1+\exp\left(-\mathbf{x}'_{hik}\hat{\boldsymbol{\beta}}^{(gj)}\right)\right] - y_{hik}\right],$$

where $\hat{\boldsymbol{\beta}}^{(gj)}$ solves

$$\sum_{hik\in S} w_{hik}^{(gj)}\left[r_{hik} - \left[1+\exp\left(-\mathbf{x}'_{hik}\boldsymbol{\beta}\right)\right]^{-1}\right]\mathbf{x}_{hik} = 0.$$

Theorem 5. *Assume that the conditions of Theorem 4 hold. If* $(n/M^2)V_L\left(\hat{\theta}\right)$ *converges to a positive constant, then* $(n/M^2)\left[\hat{V}_L\left(\hat{\theta}\right) - V_L\left(\hat{\theta}\right)\right]$ *and* $(n/M^2)\left[\hat{V}_J\left(\hat{\theta}\right) - V_L\left(\hat{\theta}\right)\right]$ *both converge in probability to 0.*

The proof of Theorem 5 follows by standard arguments in Fuller (2009) and Shao and Tu (1995) and is hence omitted.

The parameter θ for examining bias with inverse propensity weighting was defined for population totals. As with poststratification, it may be desired to compare means instead of totals, particularly if weight trimming is used to truncate large and influential values of the propensity weight $\left[1+\exp\left(-\mathbf{x}'_{hik}\hat{\boldsymbol{\beta}}\right)\right]$. In this case, the parameter to be evaluated is

$$\theta_M = \frac{\sum_{hik\in U}\left[R_{hik}\,y_{hik}/R_{hik}^{M}\right]}{\sum_{hik\in U} R_{hik}/R_{hik}^{M}} - \frac{\sum_{hik\in U} y_{hik}}{M}$$

with estimator

$$\hat{\theta}_M = \frac{\sum_{hik\in S} r_{hik}\,w_{hik}\,y_{hik}\left[1+\exp\left(-\mathbf{x}'_{hik}\hat{\boldsymbol{\beta}}\right)\right]}{\sum_{hik\in S} r_{hik}\,w_{hik}\left[1+\exp\left(-\mathbf{x}'_{hik}\hat{\boldsymbol{\beta}}\right)\right]} - \frac{\sum_{hik\in S} w_{hik}\,y_{hik}}{\sum_{hik\in S} w_{hik}}.$$

Special adjustments are needed to account for weight trimming with the linearization variance estimator; in general, we recommend using the jackknife or another replication method for finding the variance of $\hat{\theta}$ or $\hat{\theta}_M$.

4 Simulation results

We examine the performance of the variance estimators in two simulation studies. The first study generates finite populations with response indicators r_{hik} and then draws simple random samples from the population. The second simulation uses data from the 2009-2013 5-year American Community Survey Public Use Microdata Samples (ACS PUMS) as a population and then draws repeated cluster samples from this population under different nonresponse mechanisms.

For the simulation involving simple random sampling, we generated finite populations of 1,000,000 units. To study the poststratification estimator we used $C = 6$ poststrata to generate nonresponse. The experimental factors were:

- sample size, n: 300 or 1,000.
- population proportion $\left(M_c/M\right)$ in each poststratum: (P1) (1/6, 1/6, 1/6, 1/6, 1/6, 1/6), (P2) (1/21, 2/21, 3/21, 4/21, 5/21, 6/21), and (P3) (6/21, 5/21, 4/21, 3/21, 2/21, 1/21).
- response rates in poststrata: (R1) (0.2, 0.3, 0.5, 0.6, 0.8, 0.9), (R2) (0.3, 0.7, 0.3, 0.7, 0.3, 0.7), and (R3) (1, 1, 1, 1, 1, 1). Level (R3), with full response, is included to explore the accuracy of the higher-order approximation to the variance when $V_1\left(\hat{\theta}\right) = 0$.
- poststratum means: (M1) (0, 0, 0, 0, 0, 0), (M2) (-2, -1, 0, 1, 2, 3) and (M3) (0, 1, 0, 1, 0, 1).
- number of poststrata used in nonresponse adjustment: 1, 3 (collapse adjacent pairs of poststrata), or 6. Only the settings with 6 poststrata are guaranteed to correct for the nonresponse bias.

Within each poststratum, population values y_i were generated from a normal distribution with the specified poststratum mean and variance 1. The response indicators r_i were generated as independent

Bernoulli random variables with mean R_i. The simple random sampling simulations were done in version 3.2.2 of R (R Core Team 2015), and 2,000 iterations were performed for each of the 162 simulation settings, which results in a standard error less than 0.005 for the Monte Carlo estimate of the rejection proportion when the null hypothesis of $\theta = 0$ is true. Some of the generated samples had fewer than two respondents in one or more poststrata, which would result in some jackknife resamples having no respondents in those poststrata. For such samples, the two poststrata with the smallest number of respondents were combined iteratively until all poststrata had at least two respondents.

For each simulation setting, the Monte Carlo (MC) variance of $\hat{\theta}$, $\hat{V}_{MC}\left(\hat{\theta}\right)$, was calculated as the sample variance of $\hat{\theta}_b$ for $b = 1, \ldots, 2{,}000$. The linearization and jackknife variance estimates were calculated for each simulated sample, and the means of those estimates over the 2,000 samples are denoted as $\hat{V}_L\left(\hat{\theta}\right)$ and $\hat{V}_J\left(\hat{\theta}\right)$, respectively.

Figures 4.1 and 4.2 display results for the simulation settings in which $V_1\left(\hat{\theta}\right) > 0$. Figure 4.1 displays histograms of the ratios of the mean linearization and jackknife variance estimates to $\hat{V}_{MC}\left(\hat{\theta}\right)$. The scatterplot in Figure 4.2 displays the percentage of the 2,000 iterations in which the null hypothesis $H_0 : \theta = 0$ is rejected at the 5% significance level. Most of the variance estimates are close to the MC variance and the rejection rate for $H_0 : \theta = 0$ is approximately 5% when $\theta = 0$, with higher power for larger values of $|\theta|$. Four of the simulation runs with $\theta = 0$, however, have linearization and jackknife variances that are approximately twice the MC variance, and rejection rates that are between 0 and 1%. These results are from the simulations with poststratum means (M3), response rates (R3), population proportions (P2) or (P3), and three collapsed poststrata. Although the population means for the collapsed poststrata differ, they do not differ greatly and a sample size of 1,000 is too small for the first-order asymptotic approximation to be accurate. For these settings, a sample size of approximately 15,000 was needed to reduce the variance ratios $\hat{V}_L\left(\hat{\theta}\right) / \hat{V}_{MC}\left(\hat{\theta}\right)$ and $\hat{V}_J\left(\hat{\theta}\right) / \hat{V}_{MC}\left(\hat{\theta}\right)$ to 1.2.

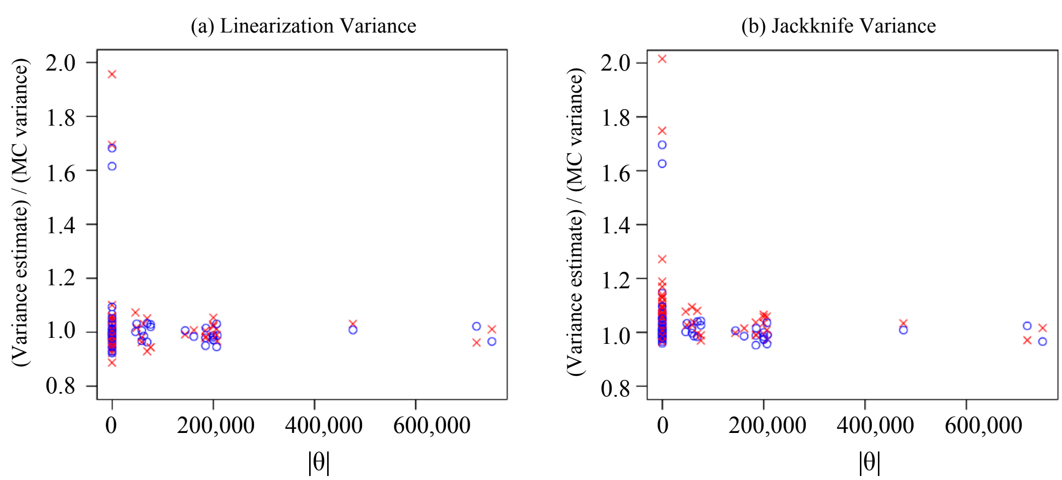

Figure 4.1 Ratios of (a) $\hat{V}_L\left(\hat{\theta}\right)$ and (b) $\hat{V}_J\left(\hat{\theta}\right)$ to $\hat{V}_{MC}\left(\hat{\theta}\right)$, for the simple random sample poststratification simulation settings in which $V_1\left(\hat{\theta}\right) > 0$. The blue circles represent simulations with $n = 1{,}000$ and the red Xs represent simulations with $n = 300$.

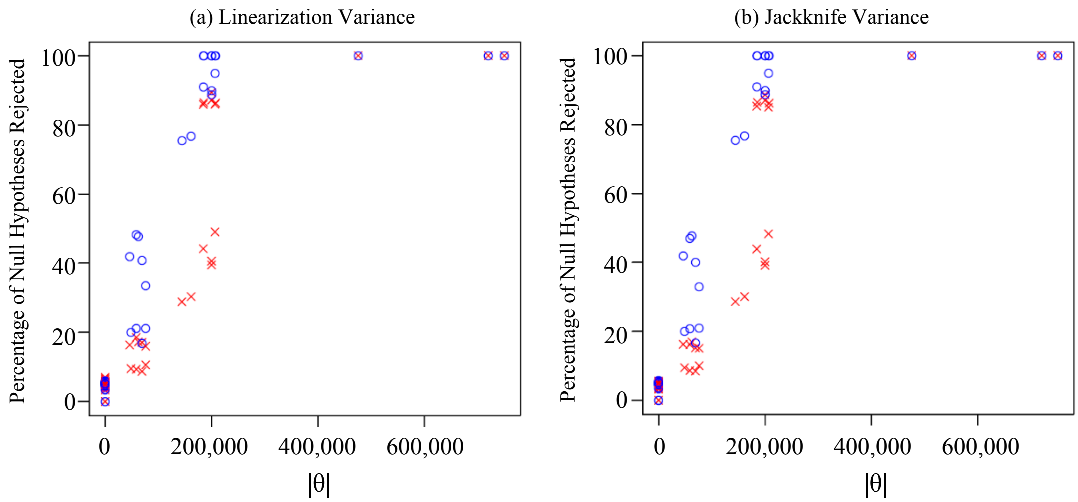

Figure 4.2 Empirical power for tests using linearization and jackknife variance, for the simple random sample poststratification simulation settings in which $V_1(\hat{\theta}) > 0$. The blue circles represent simulations with $n = 1{,}000$ and the red Xs represent simulations with $n = 300$.

Figure 4.3 shows the behavior of $\hat{V}_L(\hat{\theta})$, $\hat{V}_J(\hat{\theta})$, and $\hat{V}_2(\hat{\theta})$ when the first-order term of the variance is $V_1(\hat{\theta}) = 0$ but $V_2(\hat{\theta}) > 0$. For all of those simulations, the true value of θ was 0 and the second-order term $\hat{V}_2(\hat{\theta})$ was calculated using the SRS approximation in Theorem 3. Even though the true first-order variance $V_1(\hat{\theta})$ is zero for these settings, the estimated first-order variances from linearization and jackknife are nonzero. For the simulations with poststratum means (M1) and response rates (R3), for example, all poststrata have the same population mean. The sample means for the poststrata differ, however, and this causes the linearization and jackknife variance estimators to be positive and, on average, about twice as large as the MC variance. The same thing happens with poststratum means (M3), population proportions (P1), and response rates (R3) when three poststrata are used: the three collapsed poststrata each have population mean 1/2 but the sample means vary.

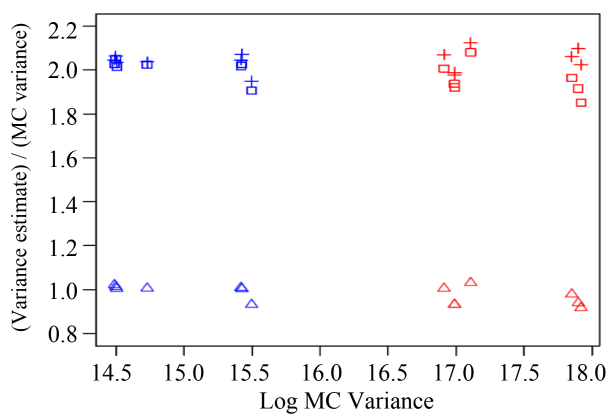

Figure 4.3 Ratios of $\hat{V}_L(\hat{\theta})$ (squares), $\hat{V}_J(\hat{\theta})$ (plus signs), and $\hat{V}_2(\hat{\theta})$ (triangles) to $\hat{V}_{MC}(\hat{\theta})$, plotted against $\ln \hat{V}_{MC}(\hat{\theta})$, for the simple random sample poststratification simulation settings in which $V_1(\hat{\theta}) = 0$. For all of these settings, $\theta = 0$. The blue symbols (with log MC variance < 16) represent simulations with $n = 1{,}000$ and the red symbols (with log MC variance > 16) represent simulations with $n = 300$.

Only simulation settings with response rates (R3) required the use of higher-order terms or large sample sizes for the linearization and jackknife variance estimators to be accurate. It would be easy to identify these situations in practice from the absence of nonresponse.

To study the properties of the estimators in Section 3, we used a subset of the populations generated for the poststratification simulation as well as populations generated with continuous covariate x, giving factors:

- Sample size, n: 300 or 1,000.
- Population values and nonresponse generation.
  1. Nonresponse is generated in 6 poststrata with population proportions (P1) or (P2), and response rates (R1) or (R2). The variable of interest y is generated with poststratum means (M1) and (M2) plus a $N(0,1)$ error term.
  2. Covariate x is generated from a $N(0,1)$ distribution. Then y is generated as (Y1) $0 + N(0,1)$ (independent of x), (Y2) $x + N(0,1)$, or (Y3) $x^2 + N(0,1)$. The response propensities are generated as (R1P) $R = 0.8$ for all units, (R2P) $\text{logit}(R) = 1/(1+\exp(-x))$, and (R3P) $\text{logit}(R) = 1/(1+\exp(-x^2/3))$.
- Response propensity model used.
  1. For poststratified populations, treat x as a continuous variable with values 1–6.
  2. For populations with generated covariate x, use linear logistic regression with covariate x. This model is correctly specified for response-generating mechanisms (R1P) and (R2P) but incorrectly specified for mechanism (R3P).

To reduce the instability of the estimators, estimated response propensities less than 0.05 were replaced by 0.05, corresponding to trimming weight adjustments larger than 20. Figures 4.4 and 4.5 display the variance ratios and empirical power for the propensity model simulations. All settings in this simulation had $V_1(\hat{\theta}) > 0$. As in the poststratification simulation, the linearization and jackknife variance estimators both perform well in general. There are a few settings, however, in which the linearization variance is substantially larger than the jackknife. This occurs because of the weight trimming: the jackknife automatically accounts for the effect of weight trimming on the variance because the jackknife replicates also trim the weights. The linearization variance used in this simulation was from Theorem 5, and the formula would need to be modified to include the effects of trimming. We also ran simulations using the jackknife in which the mean was estimated instead of the population total, and the jackknife performed well for that parameter as well.

The second simulation study used a population of 6,019,599 household-level records from the ACS PUMS studied in Lohr, Hsu and Montaquila (2015). There are 3,344 PSUs in the population defined by the public use microdata areas. Eight poststrata were formed based on the cross-classification of households by tenure (rent or own), presence of children in the household (yes or no), and number of income earners (0-1 or 2+). The primary outcome variable y was household income. Additionally, a less skewed outcome variable $\log(y)$ was studied, where $\log(y)$ was set to 0 if $y < 1$.

A $2 \times 2 \times 3$ factorial design was used for this study with factors

- overall response rate: 50% or 80%.
- number of PSUs for each sample: 25 or 100.
- nonresponse generating mechanism: (N1) missing completely at random (MCAR), with response propensity for all records equal to the response rate for all households; (N2) missing at random (MAR), where a linear logistic model with main effect terms for tenure, presence of children, and number of income earners generates the response propensities; and (N3) missing not at random (MNAR), where a linear logistic model with main effect terms for tenure, presence of children, and household income generates the response propensities.

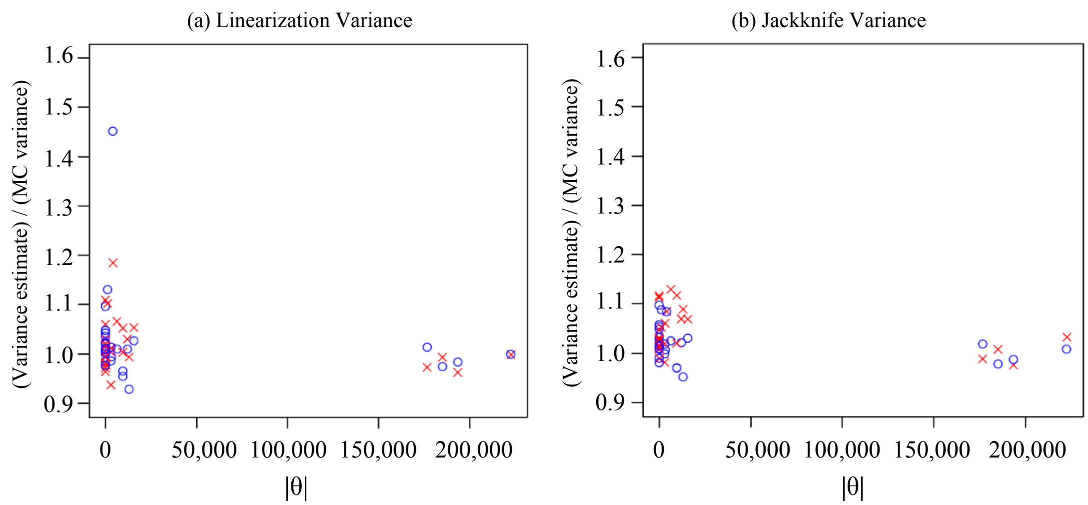

Figure 4.4 Ratios of (a) $\hat{V}_L(\hat{\theta})$ and (b) $\hat{V}_J(\hat{\theta})$ to $\hat{V}_{MC}(\hat{\theta})$, for the propensity model simulation. The blue circles represent simulations with $n = 1{,}000$ and the red Xs represent simulations with $n = 300$.

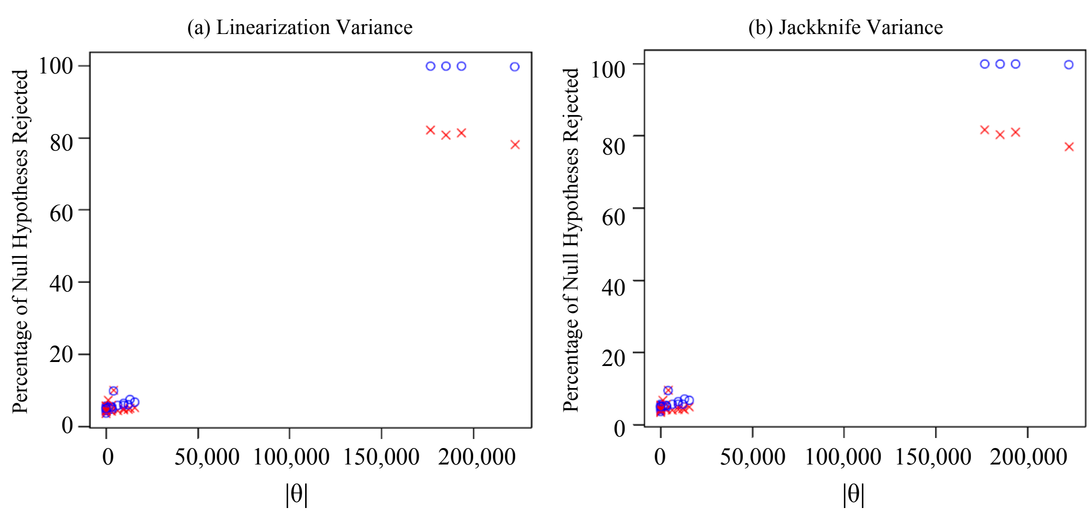

Figure 4.5 Empirical power for tests in the propensity model simulation using linearization and jackknife variance. The blue circles represent simulations with $n = 1{,}000$ and the red Xs represent simulations with $n = 300$.

For the first two nonresponse generating mechanisms, $\theta = 0$. For the first mechanism, there is no nonresponse bias. Poststratification corrects for the bias in the second mechanism because $R_{hik} = p_c$ for units in poststratum c. Poststratification does not correct for the bias in the third mechanism because the nonresponse depends on the y variable, household income.

For each simulation setting, response indicators were generated independently for the population units using the calculated response propensities. One thousand samples were drawn for each setting, in which PSUs were selected with probability proportional to size and a simple random sample of 100 households was selected from each sampled PSU. The standard error for the rejection proportion when $\theta = 0$ is less than 0.007.

Calculations for the ACS simulation were done in SAS® software (SAS Institute, Inc. 2011). We first calculated the weights and jackknife weights for the selected sample, and then calculated the poststratified and jackknife poststratified weights for the respondents. The two sets of jackknife weights used the same replication structure, so that replicate weight k for the respondents deleted the same PSU as replicate weight k for the selected sample. To simplify computation of $\hat{\theta}_M$ in (2.10), we concatenated the selected sample and respondents, with their respective weights, into one data set and set $u_i = 1$ for records in the respondent data set and $u_i = 0$ for records in the selected sample data set. The linear model $y_i = \beta_0 + \beta_1 u_i$ was fit to the concatenated data using the SURVEYREG procedure, and $\hat{\theta}_M = \hat{\beta}_1$ from the regression model.

Table 4.1 gives the results from the simulation. For all but one of the simulation settings, the mean of the jackknife variance estimates is larger than the Monte Carlo variance of $\hat{\theta}_M$, but the bias of the jackknife variance is reduced when more PSUs are sampled or the response rate is higher. The outcome variable y, household income, is highly skewed, and the rejection rate when $\theta_M = 0$ is closer to the nominal α of 0.05 when the log-transformed variable is used.

Table 4.1
Simulation results from ACS population

Nonresponse Mechanism	Response Rate (%)	Number of PSUs	Outcome variable y			Outcome variable $\log(y)$		
			θ_M	% Reject	$\frac{\hat{V}_J(\hat{\theta}_M)}{\hat{V}_{MC}(\hat{\theta}_M)}$	θ_M	% Reject	$\frac{\hat{V}_J(\hat{\theta}_M)}{\hat{V}_{MC}(\hat{\theta}_M)}$
MCAR	50	25	0	3.3	1.21	0	4.5	1.20
MCAR	50	100	0	3.0	1.09	0	4.4	1.08
MCAR	80	25	0	3.8	1.14	0	4.0	1.19
MCAR	80	100	0	3.9	1.07	0	5.2	1.05
MAR	50	25	0	4.5	1.16	0	4.2	1.11
MAR	50	100	0	4.9	1.04	0	4.4	1.05
MAR	80	25	0	3.5	1.16	0	4.7	1.20
MAR	80	100	0	3.5	1.12	0	4.6	1.11
NMAR	50	25	8,882	70.8	1.41	0.118	6.3	1.60
NMAR	50	100	8,882	99.5	1.11	0.118	37.7	1.11
NMAR	80	25	3,706	45.6	1.18	0.047	14.5	1.20
NMAR	80	100	3,706	99.4	1.09	0.047	61.0	0.99

5 Discussion

In this paper, we considered tests for nonresponse bias after poststratification or inverse propensity weighting has been used. The arguments in the theorems could be extended to similar methods that are used to adjust for nonresponse bias such as raking, which iteratively poststratifies to marginal population totals, or calibration, which adjusts the weights so that estimated population totals agree with control totals for a set of auxiliary variables. Haziza and Lesage (2016) argued that using a two-step procedure of propensity weighting followed by calibration provides more protection against nonresponse bias than using calibration alone in a single step, because single-step calibration implies a model relating the response propensities and the calibration variables and that model may be misspecified. The tests proposed in this paper could be extended to situations in which both propensity weighting and poststratification are used, or could be used separately to assess the bias removed in each step of a two-step process.

We employed the jackknife for the replication variance estimation. However, all of the estimators are smooth functions of population totals, so other replication variance estimators such as balanced repeated replication or bootstrap could be used as well.

A challenge for evaluating nonresponse bias is the limited amount of information available for the selected sample. For some surveys all available auxiliary information is used or considered for forming poststrata, raking classes, or inverse propensity weights. The poststratified estimator for characteristics used in the poststratification has no variance or bias, so testing these or closely related characteristics will not uncover nonresponse bias in other survey variables. Auxiliary variables that are not used for nonresponse adjustments are often omitted only because they were not selected in model selection method used to form the poststrata or select variables for the logistic regression, and that typically occurs because they have low explanatory power for predicting the response indicator after the other variables are included in the model. For surveys with less frame information, it may be possible to obtain auxiliary information from other sources, such as administrative records associated with the respondents' addresses or paradata. It is important to make sure that the variables used to test nonresponse bias are recorded consistently for respondents and nonrespondents. If, for example, y is the interviewer's curbside assessment about whether children are present in the household, that initial assessment should be used for both respondents and nonrespondents: the assessment used in the nonresponse bias analysis should not be updated after the interviewer ascertains the actual number of children in a responding household.

After testing available variables for nonresponse bias, we still do not know whether the adjustments have removed the bias for outcome variables that are available only for the respondents. Abraham, Helms and Presser (2009) and Kohut, Keeter, Doherty, Dimock and Christian (2012) found that estimates of volunteering and civic participation are higher from surveys with low response rates than from the Current Population Survey, indicating that weighting adjustments do not remove bias for civic engagement variables although they appear to remove bias for demographic variables and home ownership. But testing a wide range of auxiliary variables for residual bias may give more confidence in the results of a survey on the untested variables, or may indicate concerns about inferences from the survey for variables of interest. We recommend that survey designers plan the survey with nonresponse bias assessment in mind, and collect additional information for the selected sample whenever possible. In general, the more information that can be collected about the selected sample, the better.

The comparison of estimates using different sets of weights may be of special interest when studying responsive or adaptive design strategies such as those described in Groves and Heeringa (2006) and summarized in Tourangeau, Brick, Lohr and Li (2016). In these, later phases of the design are modified using information gleaned in the early returns. One responsive design strategy may be to estimate response rates after the first phase of the survey, and then to allocate resources in the second phase to equalize rates across subgroups of interest. In an experimental comparison of different responsive design strategies, it may be of interest to evaluate the estimated nonresponse bias from the strategies. Riddles, Marker, Rizzo, Wiley and Zukerberg (2015) compared nonresponse-weighted estimates from different data cutoff points in the U.S. Schools and Staffing Survey, to see if estimates changed with earlier truncation of data collection.

The results in Theorems 1 through 5 are expressed for probability samples. There is increased interest in using nonprobability samples to study populations (Baker, Brick, Bates, Battaglia, Couper, Dever, Gile and Tourangeau 2013). Proponents of nonprobability samples argue that with response rates sometimes below 10%, an inexpensive large nonprobability sample can have smaller mean squared error than a small probability sample. The same methods of poststratification and inverse propensity weighting are typically used with nonprobability samples. The tests proposed in this paper can be adapted for use with nonprobability samples, provided that auxiliary information is known for a collection of individuals that can serve as a stand-in for a sampling frame. For a web survey, it might be possible to compare characteristics of persons visiting the web page with those of persons completing the survey. Further research is needed in this area.

Acknowledgements

The authors thank the reviewers for their helpful suggestions that led to improvements in the article.

Appendix

The following lemma shows that the additional variability due to the stochastic response mechanism is $O(M^2/n)$.

Lemma 1. *Suppose assumptions (A3) and (A5) are met, and that* $|q_{hik}| \le Q$ *for all* $(hik) \in U$. *Then*

$$E\left[V\left(\sum_{hik \in U} Z_{hik} w_{hik} q_{hik} r_{hik} \middle| \mathbf{Z}\right)\right] = O(M^2/n).$$

Proof. By assumption (A5),

$$\begin{aligned}\left|E\left[V\left(\sum_{hik\in U} Z_{hik} w_{hik} q_{hik} r_{hik} \middle| \mathbf{Z}\right)\right]\right| &= \left|E\left[\sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{k=1}^{M_{hi}}\sum_{p=1}^{M_{hi}} Z_{hik} Z_{hip} w_{hik} w_{hip} \operatorname{Cov}\left(r_{hik}, r_{hip}\right) q_{hik} q_{hip}\right]\right| \\ &\le Q^2 E\left[\sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{k=1}^{M_{hi}}\sum_{p=1}^{M_{hi}} Z_{hik} Z_{hip} w_{hik} w_{hip}\right] \\ &= Q^2 \sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{k=1}^{M_{hi}}\sum_{p=1}^{M_{hi}} P[(hi)\in S]\, P[k\in S_{hi}, p\in S_{hi}]\, w_{hik} w_{hip} \\ &= O(M^2/n).\end{aligned}$$

The last line is implied by (A3).

Proof of Theorem 1. From (2.4),

$$V_1\left(\hat{\theta}\right) = V\left[\sum_{c=1}^{C} \frac{1}{p_c}\left(\hat{Y}_c^R - \bar{Y}_c^R\left(\hat{M}_c^R - M_c^R\right)\right) - \hat{Y}_{SS}\right]$$

and

$$V_2\left(\hat{\theta}\right) = V\left[\sum_{c=1}^{C} \frac{\hat{T}_c}{p_c}\right] + 2\ \operatorname{Cov}\left[\sum_{c=1}^{C} \frac{\hat{T}_c}{p_c}, \sum_{c=1}^{C} \frac{\left(\bar{y}_c^R - \bar{Y}_c^R\right)\hat{M}_c^R}{p_c} - \hat{Y}_{SS}\right].$$

The leading term simplifies to

$$\begin{aligned}V_1\left(\hat{\theta}\right) &= V\left[\sum_{hik\in U} Z_{hik} w_{hik} \sum_{c=1}^{C} \delta_{chik}\left\{\frac{r_{hik}}{p_c}\left(y_{hik} - \bar{Y}_c^R\right) - y_{hik}\right\}\right] \\ &= V\left[E\left[\sum_{hik\in U} Z_{hik} w_{hik} \sum_{c=1}^{C} \delta_{chik}\left\{\frac{r_{hik}}{p_c}\left(y_{hik} - \bar{Y}_c^R\right) - y_{hik}\right\}\middle|\mathbf{Z}\right]\right] \\ &\quad + E\left[V\left[\sum_{hik\in U} Z_{hik} w_{hik} \sum_{c=1}^{C} \delta_{chik}\left\{\frac{r_{hik}}{p_c}\left(y_{hik} - \bar{Y}_c^R\right) - y_{hik}\right\}\middle|\mathbf{Z}\right]\right] \\ &= V\left(\sum_{hik\in U} Z_{hik} w_{hik} e_{Rhik}\right) + E\left[V\left[\sum_{hik\in U} Z_{hik} w_{hik} \sum_{c=1}^{C} \delta_{chik}\, \frac{r_{hik}}{p_c}\left(y_{hik} - \bar{Y}_c^R\right)\middle|\mathbf{Z}\right]\right].\end{aligned}$$

Lemma 1 and Assumption (A4), which guarantees that $1/p_c$ is bounded, imply that the second term is $O(M^2/n)$.

To show that $V_2\left(\hat{\theta}\right) = o(M^2/n)$, note that by (A4) and the Cauchy-Schwarz inequality,

$$V\left[\frac{\hat{T}_c}{p_c}\right] \le \frac{1}{\varepsilon^2}\sum_{c=1}^{C}\sum_{d=1}^{C}\sqrt{V\left[\left(\bar{y}_c^R - \bar{Y}_c^R\right)\left(\hat{M}_c^R - M_c^R\right)\right] V\left[\left(\bar{y}_d^R - \bar{Y}_d^R\right)\left(\hat{M}_d^R - M_d^R\right)\right]}.$$

Assumption (A2) implies (Fuller 2009, Theorem 1.3.2) that

$$\sqrt{n}\begin{bmatrix} \bar{y}_c^R - \bar{Y}_c^R \\ \hat{M}_c^R / M_c^R - 1 \end{bmatrix} \rightarrow N\left(\mathbf{0}, \mathbf{\Sigma}_c\right)$$

as $n \rightarrow \infty$, where $\mathbf{\Sigma}_c$ is a non-negative definite matrix. Consequently,

$$\left(\frac{n}{M_c^R}\right)^2 V\left[\left(\bar{y}_c^R - \bar{Y}_c^R\right)\left(\hat{M}_c^R - M_c^R\right)\right] \rightarrow \mathbf{\Sigma}_c[1,1]\,\mathbf{\Sigma}_c[2,2] + 2\left(\mathbf{\Sigma}_c[1,2]\right)^2;$$

applying the Cauchy-Schwarz inequality to the covariance term implies that $V_2\left(\hat{\theta}\right) = o\left(M^2/n\right)$.

Proof of Theorem 2. We show that

$$\tilde{V}(\theta) = \sum_{h=1}^{H} \frac{n_h}{n_h - 1} \sum_{i \in S_h} \left(\tilde{b}_{hi} - \tilde{b}_h\right)^2$$

is consistent, where

$$\tilde{b}_{hi} = \sum_{k \in S_{hi}} w_{hik} \left\{ \sum_{c=1}^{C} \frac{1}{p_c} r_{hik} \delta_{chik} \left(y_{hik} - \bar{Y}_c^R\right) - y_{hik} \right\} = \sum_{k \in S_{hi}} w_{hik} \tilde{e}_{rhik}$$

and $\tilde{b}_h = \sum_{i \in S_h} \tilde{b}_{hi} / n_h$. Arguments in Yung and Rao (2000) then imply that $\left(n/M^2\right)\left[\tilde{V}(\theta) - \hat{V}(\theta)\right]$ converges to zero in probability.

Note that

$$E\left[\tilde{b}_{hi} \mid \mathbf{Z}\right] = \sum_{k \in S_{hi}} w_{hik} e_{Rhik},$$

$$\begin{aligned} E\left[\tilde{b}_{hi}^2 \mid \mathbf{Z}\right] &= E\left[\left(\sum_{k \in S_{hi}} w_{hik} \left\{ \sum_{c=1}^{C} \frac{1}{p_c} \left[R_{hik} + r_{hik} - R_{hik}\right] \delta_{chik} \left(y_{hik} - \bar{Y}_c^R\right) - y_{hik} \right\}\right)^2 \middle| \mathbf{Z}\right] \\ &= \left(\sum_{k \in S_{hi}} w_{hik} e_{Rhik}\right)^2 + V\left(\sum_{k \in S_{hi}} w_{hik} \tilde{e}_{rhik} \middle| \mathbf{Z}\right), \end{aligned}$$

and

$$\begin{aligned} E\left[\tilde{b}_h^2 \middle| \mathbf{Z}\right] &= \frac{1}{n_h^2} E\left[\sum_{i \in S_h} b_{hi}^2 + \sum_{i \in S_h} \sum_{j \neq i} b_{hi} b_{hj} \middle| \mathbf{Z}\right] \\ &= \frac{1}{n_h^2} \sum_{i \in S_h} V\left(\sum_{k \in S_{hi}} w_{hik} \tilde{e}_{rhik} \middle| \mathbf{Z}\right) + \left(\frac{1}{n_h} \sum_{i \in S_h} \sum_{k \in S_{hi}} w_{hik} e_{Rhik}\right)^2. \end{aligned}$$

This implies that

$$\begin{aligned} E\left[\sum_{i \in S_h} \left[\tilde{b}_{hi} - \tilde{b}_h\right]^2\right] &= E\left[\sum_{i \in S_h} \left(\sum_{k \in S_{hi}} w_{hik} e_{Rhik}\right)^2 - \frac{1}{n_h} \left(\sum_{i \in S_h} \sum_{k \in S_{hi}} w_{hik} e_{Rhik}\right)^2\right] \\ &\quad + \left(1 - \frac{1}{n_h}\right) E\left[\sum_{i \in S_h} V\left(\sum_{k \in S_{hi}} w_{hik} \tilde{e}_{rhik} - y_{hik} \middle| \mathbf{Z}\right)\right], \end{aligned}$$

so that $\hat{V}_L(\hat{\theta})$ is an approximately unbiased estimator of $V_1(\hat{\theta})$. The consistency follows by (A2), which implies asymptotic normality, and the law of large numbers.

Proof of Theorem 3. For $c \neq d$,

$$\operatorname{Cov}\left[\left(\bar{y}_c^R - \bar{Y}_c^R\right)\left(\hat{M}_c^R - M_c^R\right), \left(\bar{y}_d^R - \bar{Y}_d^R\right)\left(\hat{M}_d^R - M_d^R\right)\right] = o\left(M^2/n^2\right)$$

because $E\left[\left(\bar{y}_c^R - \bar{Y}_c^R\right)\left(\bar{y}_d^R - \bar{Y}_d^R\right)\right] = o\left(n^{-1}\right)$ for simple random sampling (equation (4.26) of Lohr 2010). Consequently,

$$\begin{aligned} V\left(\sum_{c=1}^{C} \frac{\hat{T}_c}{p_c}\right) &= \sum_{c=1}^{C}\sum_{d=1}^{C} \frac{1}{p_c}\frac{1}{p_d} \operatorname{Cov}\left[\left(\bar{y}_c^R - \bar{Y}_c^R\right)\left(\hat{M}_c^R - M_c^R\right), \left(\bar{y}_d^R - \bar{Y}_d^R\right)\left(\hat{M}_d^R - M_d^R\right)\right] \\ &= \sum_{c=1}^{C} \frac{1}{p_c^2} V\left[\bar{y}_c^R - \bar{Y}_c^R\right] V\left[\hat{M}_c^R - M_c^R\right] + o\left(M^2/n^2\right). \end{aligned}$$

The second term of $V_2(\hat{\theta})$ is:

$$\begin{aligned} &2\ \operatorname{Cov}\left[\sum_{c=1}^{C} \frac{\hat{T}_c}{p_c}, \sum_{c=1}^{C} \frac{\left(\bar{y}_c^R - \bar{Y}_c^R\right)\hat{M}_c^R}{p_c} - \hat{Y}_{SS}\right] \\ &= 2\sum_{c=1}^{C}\sum_{d=1}^{C} \frac{1}{p_c p_d} \operatorname{Cov}\left[\hat{T}_c, \left(\bar{y}_d^R - \bar{Y}_d^R\right)\hat{M}_d^R - p_d \hat{M}_d^R \bar{y}_d^R - p_d \hat{Y}_d^{NR}\right] \\ &= 2\sum_{c=1}^{C} \frac{1}{p_c^2} \operatorname{Cov}\left[-\left(\bar{y}_c^R - \bar{Y}_c^R\right)\left(\hat{M}_c^R - M_c^R\right), \left(1 - p_c\right)\bar{y}_c^R \hat{M}_c^R - \bar{Y}_c^R \hat{M}_c^R - p_c \hat{Y}_c^{NR}\right] + o\left(\frac{M^2}{n^2}\right) \\ &= 2\sum_{c=1}^{C} \frac{p_c - 1}{p_c^2} V\left[\bar{y}_c^R - \bar{Y}_c^R\right] V\left[\hat{M}_c^R - M_c^R\right] + o\left(\frac{M^2}{n^2}\right). \end{aligned}$$

Combining the terms,

$$V_2\left(\hat{\theta}\right) = \sum_{c} \frac{2p_c - 1}{p_c^2} V\left[\bar{y}_c^R - \bar{Y}_c^R\right] V\left[\hat{M}_c^R - M_c^R\right] + o\left(\frac{M^2}{n^2}\right).$$

We can estimate p_c by the empirical response rate in poststratum c, $V\left[\bar{y}_c^R - \bar{Y}_c^R\right]$ by s_c^2/n_c^R, and, under simple random sampling, $V\left[\hat{M}_c^R - M_c^R\right] = M_c p_c \left(M - M_c p_c\right)/n$. The term $V_2(\hat{\theta})$ can be negative when $p_c < 1/2$ for some poststrata; however, when $p_c < 1/2$ and $V\left[\bar{y}_c^R\right] > 0$, then the first-order term of the variance, $V_1(\hat{\theta})$, is positive and the second-order term has lower order.

Proof of Theorem 4. Condition (A4) guarantees that, asymptotically, complete separation will not occur and R_{hik}^M is bounded away from 0.

The derivative of $\hat{\mathbf{A}}$ with respect to the parameters is

$$\begin{aligned} \hat{\mathbf{D}}(\mathbf{r}, \boldsymbol{\beta}, \theta) &= \frac{\partial \hat{\mathbf{A}}}{\partial (\boldsymbol{\beta}, \theta)'} \\ &= \begin{bmatrix} -\sum_{hik \in S} w_{hik} \left[1 + \exp\left(-\mathbf{x}_{hik}'\boldsymbol{\beta}\right)\right]^{-2} \exp\left(-\mathbf{x}_{hik}'\boldsymbol{\beta}\right)\mathbf{x}_{hik}\mathbf{x}_{hik}' & \mathbf{0} \\ -\sum_{hik \in S} w_{hik} r_{hik} y_{hik} \exp\left(-\mathbf{x}_{hik}'\boldsymbol{\beta}\right)\mathbf{x}_{hik}' & -1 \end{bmatrix}. \end{aligned}$$

Using successive conditioning and the independence of $\mathbf{r}$ and $\mathbf{Z}$, the expected value of $\hat{\mathbf{D}}(\mathbf{r},\boldsymbol{\beta},\theta)$ is

$$\mathbf{D}(\mathbf{R},\boldsymbol{\beta},\theta) = \begin{bmatrix} -\sum_{hik\in U}\left[1+\exp\left(-\mathbf{x}'_{hik}\boldsymbol{\beta}\right)\right]^{-2}\exp\left(-\mathbf{x}'_{hik}\boldsymbol{\beta}\right)\mathbf{x}_{hik}\mathbf{x}'_{hik} & \mathbf{0} \\ -\sum_{hik\in U} R_{hik}y_{hik}\exp\left(-\mathbf{x}'_{hik}\boldsymbol{\beta}\right)\mathbf{x}'_{hik} & -1 \end{bmatrix}$$

$$= \begin{bmatrix} -\mathbf{X}'[\mathbf{I}+\mathbf{Q}]^{-2}\,\mathbf{QX} & \mathbf{0} \\ -\mathbf{T}'\mathbf{QX} & -1 \end{bmatrix}.$$

Also, $\operatorname{Cov}\left[\operatorname{vec}\hat{\mathbf{D}}(\mathbf{r},\boldsymbol{\beta},\theta)\right] = O(M^2/n)$ because

$$V\left[\sum_{hik\in S} w_{hik}r_{hik}y_{hik}\exp\left(-\mathbf{x}'_{hik}\boldsymbol{\beta}\right)\mathbf{x}'_{hik}\right] = V\left[\sum_{hik\in S} w_{hik}R_{hik}y_{hik}\exp\left(-\mathbf{x}'_{hik}\boldsymbol{\beta}\right)\mathbf{x}'_{hik}\right] + E\left\{V\left[\sum_{hik\in S} w_{hik}r_{hik}y_{hik}\exp\left(-\mathbf{x}'_{hik}\boldsymbol{\beta}\right)\mathbf{x}'_{hik}\middle|\mathbf{Z}\right]\right\}.$$

The first term is $O(M^2/n)$ by standard arguments and the second term is $O(M^2/n)$ by Lemma 1, noting that the boundedness of R_{hik} and $\mathbf{x}_{hik}$ also bound $\exp\left(-\mathbf{x}'_{hik}\boldsymbol{\beta}\right)$. Consequently,

$$V\begin{bmatrix}\hat{\boldsymbol{\beta}}-\boldsymbol{\beta} \\ \hat{\theta}-\theta\end{bmatrix} = \mathbf{D}(\mathbf{R},\boldsymbol{\beta},\theta)^{-1}V\left[\sum_{hik\in S} w_{hik}\mathbf{u}\left(y_{hik},\mathbf{x}_{hik},r_{hik},\boldsymbol{\beta}\right)\right]\mathbf{D}(\mathbf{R},\boldsymbol{\beta},\theta)^{-T} + o(M^2/n).$$

The result in (3.3) follows because

$$[\mathbf{D}(\mathbf{R},\boldsymbol{\beta},\theta)]^{-1} = \begin{bmatrix} -\mathbf{C} & \mathbf{0} \\ \mathbf{T}'\mathbf{QXC} & -1 \end{bmatrix}.$$

References

Abraham, K.G., Helms, S. and Presser, S. (2009). How social processes distort measurement: The impact of survey nonresponse on estimates of volunteer work in the United States. *American Journal of Sociology*, 114(4), 1129-1165.

Baker, R., Brick, J.M., Bates, N.A., Battaglia, M., Couper, M.P., Dever, J.A., Gile, K.J. and Tourangeau, R. (2013). Summary report of the AAPOR task force on nonprobability sampling. *Journal of Survey Statistics and Methodology*, 1, 90-143.

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review/Revue Internationale de Statistique*, 51, 279-292.

Brick, J.M. (2013). Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics*, 29(3), 329-353.

Eltinge, J.L. (2002). Diagnostics for the practical effects of nonresponse adjustment methods. In *Survey Nonresponse*, (Eds., R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little), New York: John Wiley & Sons, Inc., 431-443.

Fay, R.E. (1991). A design-based perspective on missing data variance. In *Proceedings of the 1991 Annual Research Conference*, U.S. Bureau of the Census, Washington, DC, 429-440.

Fuller, W.A. (2009). *Sampling Statistics*. Hoboken, NJ: Wiley.

Groves, R.M., and Heeringa, S.G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 169(3), 439-457.

Groves, R.M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70(5), 646-675.

Hamrick, K.S. (2012). *Nonresponse Bias Analysis of Body Mass Index Data in the Eating and Health Module*. Technical Report 1934, United States Department of Agriculture Economic Research Service, Washington, DC.

Harris-Kojetin, B.A. (2012). *Nonresponse Bias Analysis for Establishment Surveys: Guidance from the U.S. Office of Management and Budget*. Paper presented at DC-AAPOR, http://www.dc-aapor.org/documents/Harris-Kojetin2012.pdf.

Haziza, D., and Lesage, É. (2016). A discussion of weighting procedures for unit nonresponse. *Journal of Official Statistics*, 32(1), 129-145.

Haziza, D., Thompson, K.J. and Yung, W. (2010). The effect of nonresponse adjustments on variance estimation. *Survey Methodology*, 36, 1, 35-43. Paper available at http://www.statcan.gc.ca/pub/12-001-x/2010001/article/11246-eng.pdf.

Kim, J.K., and Kim, J.J. (2007). Nonresponse weighting adjustment using estimated response probability. *Canadian Journal of Statistics*, 35(4), 501-514.

Kohut, A., Keeter, S., Doherty, C., Dimock, M. and Christian, L. (2012). *Assessing the Representativeness of Public Opinion Surveys*. Pew Research Center, Washington, DC.

Lohr, S.L., Hsu, V. and Montaquila, J. (2015). Using classification and regression trees to model survey nonresponse. *Proceedings of the Survey Research Methods Section,* American Statistical Association, 2071-2085.

Oh, H.L., and Scheuren, F.J. (1987). Weighting adjustments for unit nonresponse. In *Incomplete Data in Sample Surveys,* (Eds., W.G. Madow, I. Olkin and D.B. Rubin), New York: Academic Press, 2, 143-184.

Riddles, M., Marker, D.A., Rizzo, L., Wiley, E. and Zukerberg, A. (2015). *Adaptive Design for the National Teacher Principal Survey*. Paper presented at the AAPOR 70th Annual Conference, Hollywood, FL.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Särndal, C.-E., and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. Hoboken, NJ: Wiley.

SAS Institute, Inc. (2011). *SAS/STAT 9.3 User's Guide*. Cary, NC: SAS Institute, Inc.

Shao, J., and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association*, 93, 254-265.

Shao, J., and Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer.

Tourangeau, R., Brick, J.M., Lohr, S. and Li, J. (2016). Adaptive and responsive survey designs: A review and assessment. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, published online, http://onlinelibrary.wiley.com/doi/10.1111/rssa.12186/pdf.

United States Office of Management and Budget (2006). Standards and guidelines for statistical surveys. https://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf.

Yung, W., and Rao, J.N.K. (2000). Jackknife variance estimation under imputation for estimators using poststratification information. *Journal of the American Statistical Association*, 95(451), 903-915.